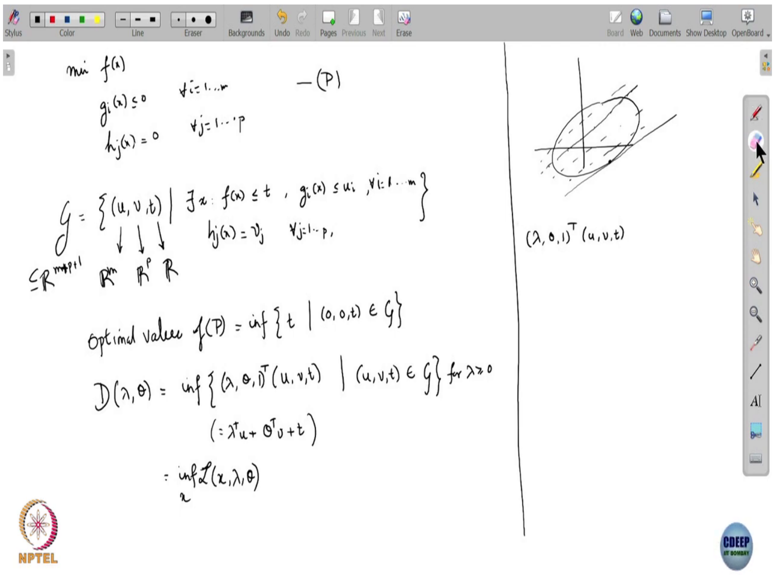
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 18A
Strong Duality in Convex Optimization – I

Ok. Welcome everyone. So, we were talking about linear under approximations to Lagrange to the Lagrangian function, you can scroll back to the previous lecture to know what that is. So, what I will talk about today is linear program is duality in the Convex Optimization world.

And the cornerstone of that is going to be the a constraint qualification type constraint qualification that we have mentioned before which is the which is Slater's which is Slater's constraint qualification ok. So, to begin with this let me recall let us state our optimization problem first.

(Refer Slide Time: 01:04)



So, we want to we are going to we are going to be minimizing this function f subject to g i of x less than equal to 0 for all i from 1 to m and j of x equal to 0 for all j from 1 to p ok. I still have not brought in convex the convex optimization problem because I wanted to first discuss a little bit of the geometry involved in this problem right.

So, let us now instead of, so the our approach is going to be to that we will not; we will not look at this problem primarily only in space of xs. So, xs are what we will call primal variables. Primal variables are the variables in which your the primitive variables in which your optimization problem has been defined. The dual variables are the variables that we introduce corresponding to the constraints in the optimization problem.

Now, the instead of looking at the problem only in the primal variable and studying the geometry of the problem only in the primal variable, what we will do is we will look at the problem jointly in the primal and dual space. Now, when you look at the problem jointly in the primal and dual space, a new type of geometry emerges. And that geometry is essentially at the heart of duality.

And one of the main lessons you should take back from this sort of analysis is that the correct space of viewing and optimization problem is in the joint primal-dual space, neither in the primal alone nor in the dual alone, but in the joint primal-dual space. You will realize all this in a much deeper way once we look at algorithms that involve both primal and dual variables ok.

So, now let me let us do this let us go into some more specifics on this ok. So, define this new quantity ok which is define this set let us call this set g, this set g is written in terms of three variables u, v and t ok. It is u, v, t, such that there exists an x for which such that f x is less than equal to t, g i of x is less than equal to u i, and h j of x; h j of x is equal to v j. So, this is for all j from 1 to p, and this is for all i from 1 to m ok. So, this is my set g.

Now, you can see what I have done here. Now, one thing you would notice here immediately is that see our original optimization problem has been posed has been written with right hand sides as 0. So, if you notice here, it is g i of x less than equal to 0 and j of x less equal to 0 for our constraints. Now, on the other hand, if you look at this set g here what we are allowing for is g i of x less than equal to some u i, and h j of x equal to some v j.

Now, the reason we are doing this is because we want to allow for constraints also to be in a to be elastic in a certain sense because and we want to see what sort of values can the objective function take as your constraints vary. This gives us then an object that lies that spans the value the basically the joint values of the objective and constraints that can possibly be attained alright.

And effectively in this space right in this space, we can then look at a new kind of object what has happened. Because we move to this sort of space we do not really need to look at we do not need to bother ourselves with the actual the x anymore. So, you will see you will see what I mean by that ok.

So, we the right hand sides here are although they are 0 here, we are defining a new object a new notion here g which is which looks something like this ok where. So, it is those xs such that it is it sorry it is u, v, t, ok, it is a three tuple of u, v and t. Here u can you can someone point out what are the dimensions of this, u must lie in R m, v must lie in R p, and t is a scalar ok. So, this entire g is a subset now of R m plus p plus 1 ok.

So, what is this set g? It is those it is those values of u, v, and t, that can be simultaneously attained by; that can be simultaneously attained by some x means its u, v and t such that there exists an x for which f x is less than equal to t, g i of x is less than equal to u i, and h j of x is equal to v j.

Now, when I mean simultaneously attained actually it is not exactly simultaneously attained. Because its it can be just simultaneously lower bounded by some x, so that means, you just need an x for which f x is less than equal to t. It does not have to be exactly equal. And similarly g i of x has to be just less than equal to u i right. But for the equality constraint we want that equality must hold h j of x should be exactly equal to v j right.

Now, suppose the optimal value of the primal, so this is my primal problem. Yeah, suppose I write what is the optimal value of the primal. What is the optimal value of the primal in terms of this set g?

Well, in terms of so what using g what we can now try to express is the optimal value of not just this particular problem P, but a family of problems, but ok not just this particular problem P. But a family of problems which in where the right hand sides are now not just are not necessarily 0, but any u i and v j right. And we will study how this varies as u i and v j vary ok.

So, the optimal value of P is this is actually equal to the infimum of the infimum, it is the least value of t such that 0, 0 comma t belongs to G. So, optimal value of P of the primal is actually the least value of t such that 0, 0 0, 0, t belongs to G. Now, how do I get that? Well, if 0's if when I put 0, u and v equal to 0 effectively I am putting g i, I am looking for those x's here for which g i of x is less than equal to 0 and h j of x is equal to 0. And then I am seeking the least value of t that is possible in this.

What would be that least value of t? Well, it would be; it would be attained at the optimal value of P itself. Because then I will get t equal to f of x star where x star is the optimal solution of P. So, the optimal value of t is not optimal value of P can be expressed in terms of this set. It is this least value of t such that 0, 0, t belongs to g.

Now, can one express the dual function now in terms of can one express the dual function in terms of; in terms of the set g? So, that also can be done. So, let us look at; let us look at this d of lambda comma theta, this is equal to the infimum of lambda theta comma 1 transpose u, v, t such that u, v, t belongs tog.

Now, let us see how this comes about. So, the what is this when I take; when I take u, v, t in g what I am taking? I am just taking any point u, v, t that such that these relations are satisfied and then what I am doing is taking saying I am taking lambda, theta, 1 transpose that.

So, effectively what is this expression? This expression is basically equal to lambda transpose u plus theta transpose v plus t right. So, this over all u, v, t that belong to G. Now, if u, v, t belong to G if as i vary u, v, t over G, ok how what would this become? So, ok one sorry one thing I forgot to mention here that this holds for lambda greater than equal to 0 ok.

So, as I vary; as I vary this, so as I vary u, v, t in G, so let us look at t first. So, t here is free. I can I need to get the least value of t what value of t would it pick out as I vary u, v, t well it would pick out the value, where t becomes equal to f x for some x right.

So, it as I vary u, v, t in G, what I am effectively varying is x, and then and t is for a fixed x t would take the value f x the least value that I would get is the value t equal to f x, v is always set equal to h of x. So, v each v j is equal to h j of x. So, v, so theta transpose v is going to be just theta transpose h j of x.

Now, because lambda is greater than equal to 0, the least value of u would also end up becoming the value of g i of x u. So, the least value of u i would end up becoming the value g i of x. So, as a result of this for lambda greater than equal to 0, we get that this actually is nothing but the infimum over x of the Lagrangian right ok.

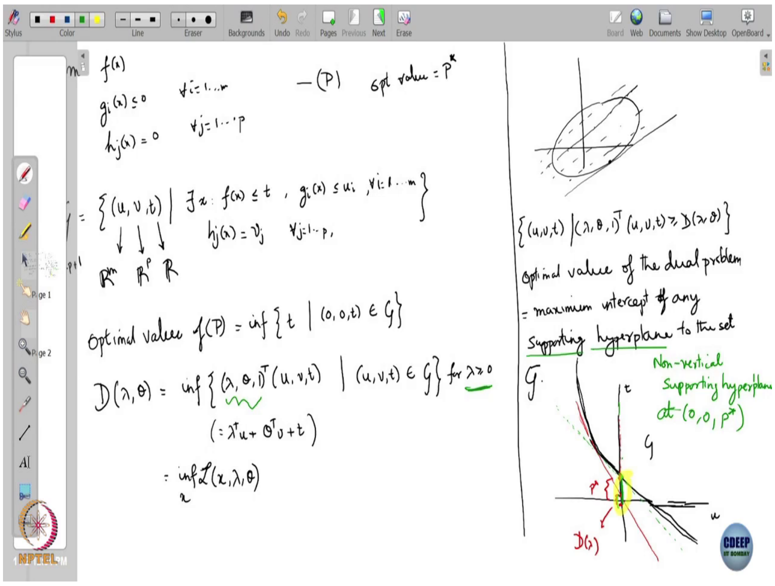
So, now, so what does this mean? This means that your so the what have we got as a result? We have got that the optimal value of P can be expressed in terms of this set G, the optimal value of the dual problem sorry or and sorry the dual function itself can be expressed in terms of this. And what kind of what does it mean for it to be expressed in terms of this? It means that the dual for the dual function what is the form what is this dual, how does this dual how is the dual function expressed in terms of G?

Well, it is you look at these kind of linear functions on the entire space and you ask what is the least value of this of linear functions with a certain type of slope over the set g that that value is actually the value of that linear function the that least value as u, v, t ranges over G is actually your dual function, and you get it as a function of the slope alright.

So, now, if you have any set G, if you have any set ok what when would if you have any set and you look for say let me draw a set here. So, some set like this. Suppose, you have some set like this. And if you are looking for linear for the if you look at let us look for linear functions with a certain slope and the look for the least value of this of this linear function over this entire set G, where would this you can plot these kind of contours of this linear function where would you get the least possible value?

Well, you would get the least possible value when you when this linear function actually forms a supporting hyper plane to that set right. So, the so somewhere embedded here is also ok a definition of a supporting hyper plane, so that means, that the dual function is effectively forming the intercept of a certain supporting hyper plane right. Because after all it you can think of it this way that lambda, theta, 1 transpose u, v.

(Refer Slide Time: 15:57)



You look at this you look at this hyper plane look for u, v, t such that lambda theta, 1 transpose, u, v, t is greater than equal to D of lambda comma theta. That what is what kind of a hyper plane this is clearly a hyper plane with a certain slope lambda theta 1 alright. And this hyper plane in this the way the intercept D of lambda comma theta has been defined, this it is a hyper plane that that contains d that contains G on one side of it. That means, the entire set g lies on one side of it.

And in fact, if the infimum is attained in the set g the infimum here is attained in the set g then it would actually touch the set g as well right. So, this is a; this is a hyper plane that you whenever the infimum is attained it is a hyper plane that supports it is a supporting hyper plane to the set g alright.

So, the dual function is basic is essentially in terms of the set g is that you have a what it is doing is well it stay you is that it is taking it is asking for the support it is asking for the intercept of a supporting hyper plane to that set alright. So, what is the maximum value of a dual of the dual, what is the optimal value of the dual problem then?

Well, the optimal value of the dual problem is the maximum value of the support of the intercept of any supporting hyperplane to the set g right. So, the optimal value of dual equal to maximum intercept of any supporting hyper plane to the set g alright ok. So, now what does this what does this effectively mean? Now, the these strong dual if you want strong duality to hold what are we effectively saying we are looking for, so let me draw a figure for you that should make things a little clearer ok.

So, here is my here is my, so I am going to draw this. So, I let me draw if a set like this. So, suppose this is my set g ok. And for simplicity I am going to skip the v variable and let us focus on only the u and t variable ok. So, this is going to be my u variable, and this is my t variable alright.

Now, if this is my set g where is in this figure where is the where is the optimal value of P ok. So, let us call that optimal value you introduce a notation for this optimal value let us call it equal as P star. So, where is P star?

P star is the least value of t such that the point 0, 0, t, or in now since we are skipping the v-axis altogether we are just taking the u-axis the point 0 comma t lies in G. And what is that point? Well, that value is here this point here this is. So, all of these; all of these points here along the t-axis are points of the form 0 comma t that lie in G right.

All of these lie in G. And what is the least value of the least value of t such that 0 comma such that 0 comma t lies in G? Well, it is equal to this. So, this is your P star alright.

Now, where is what is a supporting hyper plane here? So, let us look at let us look for now what I want to look for is something that has a I am now looking for hyper planes of the form lambda comma 1 ok. So, I am looking for hyper planes of the form lambda comma 1. So, I am looking for a hyper plane like this. And then we are what we are asking is what is the intercept, what is the intercept of this particular hyper plane? So, how do I find this particular, how do I find this particular intercept?

Well, what I need to do is put again put u equal to 0 and look for the intercept on the t-axis because my t coefficient is just one that gives me that gives me the value of the dual. And in this case the, so D of lambda, D of d of lambda is actually equal to this particular height, this height of this intercept that is my D of lambda. So I am skipping the theta because there is no for simplicity here ok.

So, now, so let us do a few more a few a little bit another example, so another more detailed example. See, so let us think of what sort of set what sort of set is G is the set g. See actually the set g that I have drawn here is not really representative of this particular set the kind of set that I that we have considered here.

If you look at the way that this set g has been defined here u can if i if any u is feasible here, then any larger u is also feasible right. So, it cannot be that the set sort of ends here it if I get; if I get a point u here, then any larger u in this direction is also feasible right. So, actually the set should look something like this.

So, if you look at this say in this along these x along this axis, it should look something like this. So, it should not really end here in this sort of way; it should actually carry on in this kind in this direction. And likewise it should carry on in the t direction as well because if there is if it if any t is feasible to the set, then a larger t is also feasible for the same values of u and v.

So, in short the set looks something like this ok. Now, I have not the way I have drawn this is it actually makes it look like it is just a very simple convex set, but their convexity needs to needs a proof and it holds only when the problem it holds when the problem is convex ok. So, just hold on until.

Then now what is then the what is then the claim of strong duality. So, if I if you look at this particular figure here this height the height that I have drawn here, this green height here this is the optimal value of my primal problem. The intercept that I have marked here is the optimal value of my dual problem.

And what you are seeing here is obvious is weak duality in play in the sense that my intercept is always going to be less than equal to this one. Because I cannot possibly have it cannot possibly you know using a supporting hyperplane I cannot possibly enter into this region. So, my intercept is always going to be less than equal to this green height.

Now, when would strong duality hold well strong duality would hold if I can do the following. If instead of taking this red intercept supporting hyperplane I can look for a supporting hyper plane like this. A supporting hyperplane that passes through the point 0, 0, P star I look for a supporting hyper plane that actually passes through the point 0, 0, P star, then its intercept would be equal to P star and then strong duality would hold right.

So, the claim of strong duality then basically says that for the set g, I should be able to find an a supporting hyperplane at the point 0, 0, P star. Now, on the phase of it this looks like a simple claim why cannot there be so then take any point you can draw a supporting hyper plane at that point why is that you know so, but it is not that simple. Firstly, you need that the set itself is it has to be a convex set that is one, but we will that is alright. We will take that.

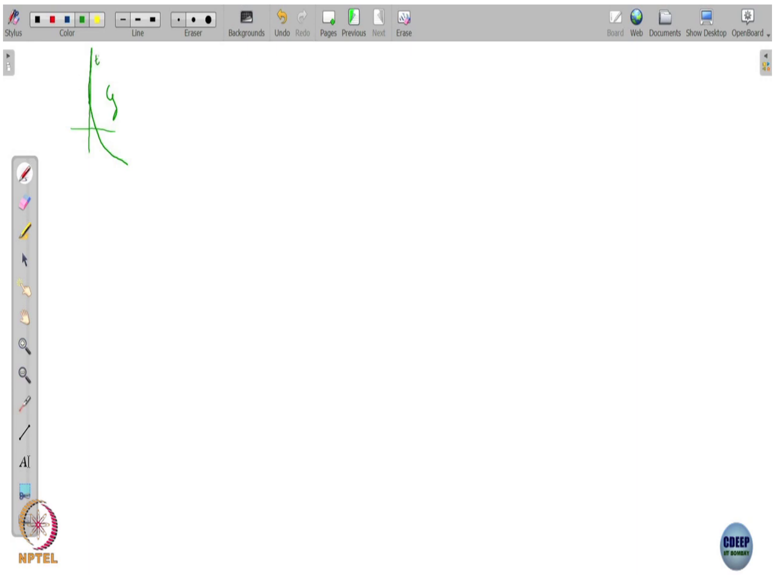
But more importantly remember that the supporting hyper plane is not just any supporting hyper plane it the all of this logic worked for supporting hyper planes that had slopes of this form. Lambda was greater than equal to 0, and the coefficient for in the t-axis of that or the slope in the t-axis was unity. So, this sort of supporting hyperplane is what is called this it is not just any supporting hyper plane it is what is called a non-vertical supporting hyperplane. Why is it non-vertical?

Now, it is non-vertical because it has a; it has a non 0 slope along t-axis. So, a vertical supporting hyper plane would be one which is parallel to the t-axis ok, that would be a vertical supporting plane hyperplane this sort of it would be bang parallel to this t-axis. We are looking for a supporting hyper plane that is non-vertical means that it should it should not be parallel to the t-axis it should have some slope along the here.

So, the t coefficient should not be 0 alright. So, that is what we mean by a non-vertical supporting hyperplane. So, that is what makes the problem then a little bit more subtle than just simply drawing a supporting hyperplane.

We are now looking for that the set g must have a non-vertical supporting hyperplane at the point 0, 0, P star. If that is the case, then strong duality would hold ok. So, what that means is for instance you cannot have is that the set g cannot take this sort of shape for instance.

(Refer Slide Time: 27:54)



So, the set g cannot you know at the point P star should not just shoot off like this. This sort of set g ok or cannot it is supporting hyper plane at P star would be exactly parallel to the t-axis. And this sort of set will not you they will you will not have yeah you cannot get that sort of supporting hyperplane for this set ok.

So, that is basically the that is basically what is at the heart of this particular theorem the heart. So, to repeat what we are looking for is a is something is this particular is this we are focusing on this particular region. If you look here we have the green segment here is the height of; the height of that green segment is or the length of that green segment is the optimal value of the primal, and the little intercept here is the optimal is the value of the dual function.

And what we want to do is we want to search over all possible support and the dual function is formed by the support is the is basically the intercept of a supporting hyper plane to this set. Now, what we are when we are maximizing the dual function, what we are doing is searching over all sorts of supporting hyper planes and looking for the largest value of the intercept.

Now, the intercept obviously by the geometry of this problem by the virtue of being a supporting hyper plane, it has to be that the intercept will always be less than this green segment. But one way by which you can get equality, that means, we can get that you would that it is that the intercept would be exactly equal to P star would be when you are able to draw is in non-vertical sub when you are able to draw a supporting hyper plane of this form at 0, 0, P star.

And now when what is this sort of form for this supporting hyperplane? Well, it must have lambda the slope in the lambda in the u direction should be greater than equal to 0 that is lambda should be greater than equal to 0 and the slope on the t in the t direction should be positive. And it should be that this sort of hyperplane is what is called a non-vertical supporting hyperplane and we. So, what we are looking for therefore, for strong duality is a non-vertical supporting hyperplane at the point 0, 0, P star all right ok.

So, now let us actually get on to the proof of this. Now, we need for it is for the existence of a non-vertical supporting hyperplane is where we will need actually the properties of convexity. Now, you can see this sort of argument where you had to just you had this we had this what segment and you are looking at the intercept we wanted to compare the segment and the intercept that kind of argument does not need convexity at all.

It will always be the case once you have a supporting hyper plane it will always be the case that its intercept will lie below ok, so that is that does not need any convexity. Convexity comes up will come up subsequently now.