Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-3

Lecture - 9: Slaters condition

So, duality gap is defined to be that we know that d star at best can be equal to p star So, g is greater than equal to 0 and this quantity g is called duality gap right and let us look at one example where we see this kind of duality gap. So, yeah before that let me just summarize this. So, we have the primal we had the primal problem to start with which was this particular problem right and the equivalent dual problem is. So, the Lagrangian dual problem would be. So, d star which is defined to be the optimal value of this is the Lagrangian dual problem ok. And we know that it is always concave this particular optimization like g lambda nu is always going to be concave even if the original functions are non-convex ok.

* Duality Gap:

$$\underline{\underline{G}} := p^* - d^* \geq 0$$

* Lagrangian dual problem:

$$d^* := \max_{\substack{(\lambda,\nu) \\ \lambda \geq 0}} g(\lambda,\nu)$$

So, let us look at an example where you actually see the duality here. So, example of duality where even if like I mean every constraint and everything those are convex, but they still see that we already get. No, no just lambda for the inequality constraint. So, as I as we saw here right like if we look at the lower bound property I mean this is anyway equal to 0, we just needed lambda to be greater than equal to 0, so as to and nu cannot be 0 right because let us say I mean as I said the vectors are going to be collinear.

So, if you make them 0, then you are saying that the vectors are in fact 0, then it becomes an unconstrained minimization, right? It can be negative, yes. I mean if the, let us say this constraint set over here, right? If it actually, I mean it does not grow out, it basically decreases out, I mean as you go outwards, then the vector gradient of b of x would be pointing inwards, right? So, then in that case it will be negative. So, it really depends on the function that you are working with I mean if I if I represent the equality constraint as b minus b x instead of b x minus b then I mean it becomes it flips the sign right because it is equality. So, I can write it both ways. So, let us say your x is in R 2.

So, you want to minimize e to the x 2 subject to norm of x less than equal to x 1 ok. So, what is the constraint set look like here? We say that a 2 norm of less is 2 x is less than equal to 1, what is the 2 norm of x? So, let us. So, the constraints are of this form x 1 x 2 such that x 1 square plus x 2 square is less than equal to x 1 right. So, when can this be true? Yeah. So, this is essentially when x 1 is greater than equal to 0 and x 2 is equal to 0, ok.

That is the only way this can be true, right, ok. So, when x 2 equal to 0, what is the value of this function? e to the 0 is 1, right. and no matter what your x I mean what like what your x 1 is this will always be 1. So, the primal objective value is ok. So, for I mean in fact yeah.

* Example of duality gap:

$$\min_{x \in \mathbb{R}^2} \; e^{x_2} = 1$$

$$\text{st} \quad \|x\|_2 \leq x_1$$

$$\mathcal{X} = \left\{ (x_1, x_2) : \sqrt{x_1^2 + x_2^2} \leq x_1 \right\}$$

$$= \left\{ (x_1, x_2) : x_1 \geq 0 \,, x_2 = 0 \right\}$$

So, this is the primal objective value is 1 ok. Is this clear? So, let us look at the dual optimal value. So, primal optimal is p star we know it is 1. So, in this case we have g lambda, lambda is going to be scalar because it is just one inequality constraint and there is no equality constraint. So, we just have g lambda and that is defined to be lambda times h of x less than equal to 0 is the constraint.

So, what is h of x? Is this clear? And what is f of x here? e to the x 2. We know that e to the x 2 no matter what your x 2 is always greater than equal to 0. What about this? is this quantity always greater than equal to 0? Yes, right. So, this is always greater than equal to 0 and the dual optimization problem is you maximize g lambda subject to lambda greater than equal to 0. So, one thing that we know is d star is greater than equal to 0.

Why? Because d star by definition is maximize over lambda greater than equal to 0 g lambda and from here we know that g lambda is greater than equal to 0. So, d star is always greater than equal to 0. That is one condition that we have. Is this clear? I mean we still have not proved or like one way or the other whether there exists a duality gap or not right. Maybe d star is greater than equal to 0 if maybe d star is equal to 1 right.

$$\boxed{p^* = 1} \quad \text{since } e^{x_2} = e^0 = 1$$

$$g(\lambda) = \min_{x \in \mathbb{R}^2} \Big\{ \underbrace{f(x)}_{\substack{e^{x_2} \\ \geq 0}} + \lambda \big(\underbrace{\sqrt{x_1^2 + x_2^2} - x_1}_{\geq 0} \big) \Big\}$$

$$\boxed{d^* \geq 0} \qquad d^* := \max_{\lambda \geq 0} g(\lambda) \geq 0$$

So, in that case there is no duality gap. As of now we have not said that there is a I mean we have not shown that there is a duality gap in this problem. So, we still need to find what d star is ok. So, let us define this function which basically is this particular function over here. So, this is nothing, but which is going to be x 2 square.

So, you just multiply with the conjugate of this term in the numerator and the denominator and that is what you get right ok. And if you choose then you can show that this particular term is upper bounded by this particular thing. I mean just simple algebra just substitute x 1 to be x 2 to the power 4, you can show that this is this eta term is upper bounded by 1 over x 2 square and as x 2 goes to negative infinity basically this particular term goes to 0 right. as x 2 goes to negative infinity.

$$\eta(x_1, x_2) = \sqrt{x_1^2 + x_2^2} - x_1 = \frac{x_2^2}{\sqrt{x_1^2 + x_2^2} + x_1}$$

$$\text{If } x_1 = x_2^4 \qquad \eta(x_1, x_2) \leq \frac{1}{x_2^2} \xrightarrow[x_2 \to -\infty]{} 0$$

So, what is the definition of g lambda? So, g lambda is defined to be minimum of x in R 2. e to the x 2 plus lambda times this particular term eta x 1 x 2 right. That is the definition of g lambda. Now this is going to be like, so I am doing this minimization over the entire of R 2 right. If I restrict this over a smaller like a subset of R 2, then that particular minimum is going to be greater than this minimum. Let us say if I try to minimize x square, a function x square over the interval 1 to the value of x square is the minimum value that of x square that I can get is 1.

But if I try to minimize the same function s square over the entire r, the minimum value is 0 right. So, if you the moment you minimize it is over a smaller set that minimum value is go always going to exceed the one that we. So, essentially if I try to minimize it over this set ok. and as x2 goes to negative infinity this particular term goes to 0 because e to the negative infinity is 0 we have already shown that this goes to 0 along this particular line along this particular constraint set. So that means g lambda is less than equal to 0 along this set and the maximum of this also is going to be less than so d star is going to be less than equal to 0 ok.

$$g(\lambda) = \min_{x \in \mathbb{R}^2} \left(e^{x_2} + \lambda \eta(x_1, x_2) \right)$$

$$\leq \min_{x_1 = x_2^4} \left(\underline{e^{x_2}} + \lambda \underline{\eta(x_1, x_2)} \right) \to 0$$

$$A \quad x_2 \to -\infty$$

$$g(\lambda) \leq 0 \qquad \boxed{d^* \leq 0}$$

So the maximum of g lambda is also going to be less than equal to 0. So d star is going to be less than equal to 0. And if I look at the previous constraint that I have obtained on d star, it was that d star is greater than equal to 0. So the only way this is possible is that d star is equal to 0. So from these two constraints, I obtain that d star is equal to 0, meaning the duality gap G which is P star minus D star that is equal to 1 which is greater than 0.

So, this is I mean there exists. So, this is a I mean basically this formulation is I mean you have a weak duality here it is not strong duality because the duality gap is strictly greater than 0. So, this is an example of a problem where you see the duality gap and in that case I mean when you solve the dual problem you only get an approximate solution not a not the exact solution right of the original problem that you wanted to solve. In cases when the duality gap is 0 that is when you can hope to find like that is when you can actually find the exact solution to the same problem primal problem in a much simpler manner and we are now going to look at constraints under which the strong duality holds. Is this clear? This one? So, we know that g of lambda is less than equal to this term and as x2 goes to negative infinity this goes to 0, e to the negative infinity also goes to 0.

So, this is 0, like the minimum of this is also going to be 0. So, g of lambda is less than equal to 0 and then now if you maximize g of lambda because this is true for any lambda, right? So, if you maximize this with respect to lambda that is also going to be less than equal to and therefore d star is less than equal to 0 ok. So, we have obtained that I mean this is basically there exists duality gap right in this problem. For the remainder of the lecture we will focus on basically looking at conditions under which strong duality amounts. Because here from here we know that d star is less than equal to 0 right.

$$\boxed{d^* = 0}$$

$$\boxed{G = p^* - d^* = 1 > 0} \rightarrow \exists \text{ duality gap}$$

and earlier when we looked at the definition of g lambda we obtained that d star is greater than equal to 0. So, the only way both can be true is when d star is exactly equal to 0 ok. So, now we are going to look at conditions under which strong duality holds. So, that means the duality gap g is equal to 0 or p star is equal to d star. So we look at something called Slater's condition which says that, so if there exists an x bar or which is strictly feasible.

What do we mean by strict feasibility? That means for every inequality constraint it is actually satisfied with strict inequality. So, Slater condition says that if you are able to find an x bar such that the inequality constraints are strict. If you are able to find one such x bar, then under the assumptions that we looked at that So, what were the assumption here? So, assumptions was that these functions are convex and p star is finite under this assumption. So, let us call this assumption star. So, then assumption star plus strict feasibility.

So, this implies strong duality. So in the previous example, was there a strict feasibility? I mean the duality gap was there, so it wasn't, I mean it wasn't the case of strong duality, right? And if I look at all the set of feasible points, so x2 equal to 0, in fact, I mean it's not possible to have a strict feasibility, right? We cannot have a point which is strict, I mean which is entirely in the interior of x. this point because of x equal to 0 it lies at the boundary right. So, there is I mean we do not have strict feasibility here. So, the Slater condition Slater's condition are not met and therefore, we cannot have strong duality I mean we cannot conclude that way it is a sufficient condition, but the fact that we did not have a strong duality that in itself tells you that you would not have strict feasibility in the first place.

* Strong duality: $G = 0$ or $p^* = d^*$

Slater's condition: If $\exists\, \bar{x}$ which is strictly feasible,

$$h_i(\bar{x}) < 0 \quad \forall\, i \in \{1, 2, \ldots, m\}$$

then Assumption $(*)$ + Strict feasibility

$$\Downarrow$$

Strong duality.

So, under this assumption star and strict feasibility you would always have strong duality. So, that is the Slater's condition. and just a quick remark. So, this requirement for strict feasibility, it is there when you have non-linear inequality constraints. For linear inequality constraints, you always have a strong duality.

Yeah, just one point. So, something that must exist in the interior of the set. So, for linear inequality constraints, strong duality holds even without strict feasibility . So as long as you have linear equality constraints I mean you would have strong duality, but for in general non-linear inequality constraints you would want them to hold with you basically want them to be or you want the Slater's condition to hold true. So that means you should be able to find a strictly feasible point and as long as you are able to do so you can guarantee strong duality. And what are the consequences of strong duality? I mean something that we already looked at.

So if a strong duality holds that means we can work with dual optimization problem right. can work with dual optimization problems. So, something that we have not looked at is a KKT condition something which you which you are going to look at in the subsequent lectures, but if strong duality holds then KKT conditions which are always sufficient also become necessary. So, KKT conditions become if and only if kind of conditions under strong I mean we have not specified what KKT conditions are and I mean you do not have to worry about it for now, but then something that we are going to look at in the subsequent lectures ok. So, again for strong duality to hold for in I mean Slater's condition are sort of sufficient conditions.

So, if you are able to find strictly feasible point, you can guarantee for sure that strong duality holds and for linear inequality constraints you do not even have to do anything. I mean there you they I mean there will be an equivalent dual formulation by itself. as long as this assumption star is there ok. Any questions on this? So, what is the geometric meaning of Lagrange multipliers or the dual variable or not geometric meaning, but like in general how should you interpret dual variables? So, particularly again when we look at the equality constraint problem. So, this particular example that we looked at which was minimize f of x, x in R n subject to some equality budget constraint right.

So, what is the meaning of? So, we get some new star right which is the dual variable corresponding to this particular equality constraint, but what is the an intuitive understanding of this particular dual variable. So, let us write down the Lagrangian and Lagrangian turns out to be f of x plus nu times ok. and the constraints are or like I mean if we basically now I mean that. So, this basically becomes an unconstrained optimization problem. So, the first order condition for optimality is the gradients must vanish both with respect to x and nu.

* How should we interpret dual variables?

$$\left.\begin{array}{l} \min\limits_{x\in\mathbb{R}^n} f(x) \\ \text{s.t. } \underline{B(x)=b} \end{array}\right\} \longrightarrow \nu^*$$

$$L(x,\nu) = f(x) + \nu(B(x)-b)$$

$$B(x^*)=b$$

$$\nabla f(x^*) + \nu^* \nabla B(x^*) = 0$$

So, this basically gives you that v x star is equal to b why because the gradient with respect to nu should be 0 and the other constraint is nu star So, let us say m star is the optimal value. It is the optimal value of this objective function f of x that you are trying to minimize subject to this constraint. So, m star by definition is f of x star. because it is the optimal value and x star is the optimal solution which is because of the this particular constraint here this is nothing but L of x star nu star right ok is this clear. By the way if I change my budget I know that my x star and nu star these points are also going to vary right as I vary my budget little b.

So you should really view x star as a function of b and likewise nu star as a function of Because if you change your budget b that you have under which you are operating that that is also going to vary your optimal solution ok. So, let us see what this looks like. So, how does. So, the question is. So, now we are going to study how does this particular optimal value changes as my change my budget ok.

So, this is this is the derivative that I am going to evaluate. So, if I am going to be changing my budget, my optimal solution is going to be varying and why because these are also going to be function of b. So, this by definition is d ok, this total derivative is partial partial b. Is this clear? So, what is this quantity? Partial L partial x at x star. So, what is the first order condition for optimality? The gradient must vanish at x star and nu star, right? So, this is equal to 0 and this is equal to 0.

And what is partial L partial b? So, the derivative of this with respect to b is negative nu star. which is your dual variable right. So, so really like the way you should interpret this dual variable is as an incremental cost variable. So, as you change your budget how does your optimal solution vary? So, the rate of it is what this Lagrange this dual variable captures. So, this is also sometimes called incremental cost variable ok.

Let us say M^* is the optimal value.

$$M^* = f(x^*) = L(x^*, \nu^*)$$

$$x^*(b) \qquad \nu^*(b)$$

$$\frac{dM^*}{db} = \frac{d\,L(x^*,\nu^*)}{db} = \frac{\partial L(x^*,\nu^*)}{\partial b} + \cancel{\frac{\partial L}{\partial x^*}}\frac{\partial x^*}{\partial b} + \cancel{\frac{\partial L}{\partial \nu^*}}\frac{\partial \nu^*}{\partial b}$$

$$= -\nu^*$$

↳ Incremental cost variable.

So, your Lagrange multiplier nu star turns out to be dm star like this derivative of an optimal solution with respect to, ok. Is this clear to everyone? Alright. So, now we will conclude this with an example where we would try and convert a primal form to a dual form. So, we started with this particular problem minimize So, this is the primal form.

You also assume that Q is invertible. So, Q is positive definite. So, first of all does strong duality hold here? Right. I mean what kind of constraints are these? Linear inequality constraints right. So, strong duality always holds. So, and if Q is positive definite this function is convex.

So, these assumptions also hold. So, that means like the strong basically this strong duality holds right. So, we do not have to look for strictly feasible x as long as it is linear inequality constraint. So, that means in the dual objective value would be same as the primal objective value. So, first thing that we should notice is that strong duality holds.

Ex:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x, \quad Q \succ 0$$

$$\text{s.t. } Ax \leq b$$

$\longrightarrow$ Primal form

Strong duality holds $p^* = d^*$

$$g(\lambda) = \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} x^T Q x + c^T x + \lambda^T (Ax - b) \right\}$$

and p star turns out to be d star. So, now we will try and convert this to a dual optimization problem ok. ok is this clear. So, now this becomes an unconstrained optimization problem in x in order to be able to find g lambda right ok. So, how do we solve for this how do we find g lambda? So, we have to minimize this particular function and what is like if it is an unconstrained minimization what is the criteria to minimize with respect to x? Just set the derivative with respect to x and set it to 0 right take the derivative and set it to 0. So, what is the derivative with respect to x? Q x plus c transpose lambda that should be equal to 0 right.

$$Qx^* + c + A^T \lambda = 0$$

$$x^* = -Q^{-1}(c + A^T \lambda)$$

So, x star in because Q is positive definite, so this is invertible ok. Now you substitute the value of x star and everything would be in terms of like basically lambda right because this is defined only for minimum and you can verify I will just write down the solution you can verify this turns out to be. So, the equivalent g lambda which turns. So, this g lambda turns out to be negative lambda transpose P lambda minus a transpose lambda where this matrix P is defined to be something that you can verify.

So, I am not going to be deriving that ok. So, this is your dual problem and we know that dual function is always concave function right. if q is positive definite q inverse is also positive definite. So, this p turns out to be positive like negative semi definite at least right and that which is which is basically your concave function. So, this is always concave function and we look at the maximization of this completely defined in terms of lambda right and since it is defined in terms of lambda now and like if I look if I look at the problem that we started with which was which is this particular problem right and we are we were looking at the constraint when n is very large and r is very small everything is defined in terms of lambda which is in r dimensional right. And it is much easier to solve every agent can solve it locally and then you can use can use this particular constraint to evaluate once they have once they have agreed upon a common lambda then they can simply solve for this right and find the x star.

$$\boxed{\max_{\lambda \geq 0} g(\lambda) = -\frac{1}{2}\lambda^T P \lambda - a^T \lambda}$$

$$P := A Q^{-1} A^T$$

$$a := b + A Q^{-1} c$$

So, this is this is a very sort of neat way to work with the reduced order problem also the constraints that are going that you are going to be working with is much simpler than Ax less than equal to b right it just lambda greater than equal to 0. So, working with dual problems in certain cases turns out to be much easier than working with primal problem and this is something that we are also going to be looking at when we. So, remember in one of the lectures we looked at the mathematical formulation of support vector machines and we looked at the primal form. We are going to look at the dual form and then we are going to see how it also helps with implementing the kernel SVM something that that is often used in machine learning. So, yeah with this I would like to end today's lecture unless there are any more questions. Thank you very much.