ECE 490: Introduction to Optimization **Fall 2018**

Lecture 22

Duality

Lecturer: Bin Hu, Date:11/27/2018

Today we will present the duality theory for the following general constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \\ & g(x) \leq 0 \end{aligned} \tag{22.1}$$

Notice h and g both are vector functions. Specifically, we have

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_m(x) \end{bmatrix}, \quad g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_l(x) \end{bmatrix}$$

The duality theory here is very similar to the duality theory for optimization with equality constraints. Define the Lagrangian

$$L(x, \lambda, \mu) = f(x) + \lambda^\mathsf{T} h(x) + \mu^\mathsf{T} g(x)$$

From KKT condition, we only consider $\mu \geq 0$ (here μ is a vector and what we really mean is that each entry of μ is non-negative). Then the duality theory is built upon the following inequality:

$$\max_{\lambda \in \mathbb{R}^m, \mu \geq 0} D(\lambda, \mu) = \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} \min_{x \in \mathbb{R}^p} L(x, \lambda, \mu) \leq \min_{x \in \mathbb{R}^p} \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} L(x, \lambda, \mu) = \min_{x: h(x)=0,\, g(x) \leq 0} f(x) \tag{22.2}$$

where $D(\lambda, \mu) := \min_{x \in \mathbb{R}^p} L(x, \lambda, \mu)$ is the so-called dual function. (More precisely, we should replace min with inf, but for simplicity we abuse the notation and still use min here.) Now we explain the above statement:

1. $\max_{\lambda \in \mathbb{R}^m, \mu \geq 0} D(\lambda, \mu) = \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} \min_{x \in \mathbb{R}^p} L(x, \lambda, \mu)$: This follows from the definition of the dual function.

2. $\max_{\lambda \in \mathbb{R}^m, \mu \geq 0} \min_{x \in \mathbb{R}^p} L(x, \lambda, \mu) \leq \min_{x \in \mathbb{R}^p} \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} L(x, \lambda, \mu)$: This follows from the fact that we have $L(x, \lambda, \mu) \leq \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} L(x, \lambda)$ given any $\mu \geq 0$ and arbitrary vectors (x, λ). Consequently, we can take min over x on both sides and have $\min_{x \in \mathbb{R}^p} L(x, \lambda, \mu) \leq \min_{x \in \mathbb{R}^p} \max_{\lambda \in \mathbb{R}^m, \mu \geq 0} L(x, \lambda, \mu)$. This directly leads to the desired inequality.

3. $\min_{x\in\mathbb{R}^p}\max_{\lambda\in\mathbb{R}^m,\mu\geq 0} L(x,\lambda,\mu) = \min_{x:h(x)=0,\, g(x)\leq 0} f(x)$: This is a direct consequence of the following relation:

$$\max_{\lambda\in\mathbb{R}^m,\mu\geq 0} L(x,\lambda,\mu) = \begin{cases} f(x) & \text{if } h(x)=0 \text{ and } g(x)\leq 0 \\ +\infty & \text{Otherwise} \end{cases} \tag{22.3}$$

   Notice if we do not have $h(x)=0$, then we can always choose some arbitrarily large λ to make $f(x)+\lambda^{\mathsf{T}}h(x)+\mu^{\mathsf{T}}g(x)$ go to infinity. Similarly, if we do not have $g(x)\leq 0$, we can choose some μ to make $f(x)+\lambda^{\mathsf{T}}h(x)+\mu^{\mathsf{T}}g(x)$ go to infinity.

Concavity of dual function. Similar to the case where all the constraints are equality constraints, the dual function is always concave no matter what f we have. Please verify this fact by yourself. The only inequality you need to prove the concavity of D is $\min_{x\in\mathbb{R}^p}\{a(x)+b(x)\}\geq \min_{x\in\mathbb{R}^p} a(x)+\min_{x\in\mathbb{R}^p} b(x)$.

Strong duality. If the inequality in (22.2) holds as an equality, then we have the so-called strong duality. In general, the proofs of strong duality are case-dependent. There exist examples where strong duality holds for non-convex problems. Those proofs are non-trivial. When f is convex, g_j is convex, and h_i is linear, a sufficient condition guaranteeing the strong duality is that there exists a vector x satisfying $g_j(x)<0$ for all j and $h_i(x)=0$ for all i. This is the famous Slater's constraint qualification. There also exist other types of constraint qualifications that guarantee strong duality for various problems.

Dual problem. Dual problem refers to the following problem

$$\begin{aligned} \text{maximize} \quad & D(\lambda,\mu) \\ \text{subject to} \quad & \lambda\in\mathbb{R}^m \\ & \mu\geq 0 \end{aligned} \tag{22.4}$$

where D is the dual function. Based on the duality theory, the solution for the dual problem provides a lower bound for the solution of the primal problem (22.1). Sometimes this lower bound can be $-\infty$ and is completely useless. When strong duality holds, this solution for the dual problem becomes really useful and is also a solution for the primal problem. Quite often $D(\lambda,\mu)$ is only well-defined on a certain set and this poses some extra constraints to the dual problem. We will demonstrate this by examples. Now we demonstrate how to formulate dual problems by presenting a few example.

22.1 Dual of Linear Programming (LP)

Consider the following primal linear programming problem:

$$\begin{aligned} \text{minimize} \quad & c^{\mathsf{T}}x \\ \text{subject to} \quad & Ax=b \\ & x\geq 0 \end{aligned} \tag{22.5}$$

To formulate the dual problem, we first write out the Lagrangian:

$$L(x, \lambda, \mu) = c^\mathsf{T} x + \lambda^\mathsf{T}(Ax - b) + \mu^\mathsf{T}(-x) = \left(c^\mathsf{T} + \lambda^\mathsf{T} A - \mu^\mathsf{T}\right) x - \lambda^\mathsf{T} b$$

We have

$$D(\lambda, \mu) = \min_{x \in \mathbb{R}^p} L(x, \lambda, \mu) = \begin{cases} -\lambda^\mathsf{T} b & \text{if } c^\mathsf{T} + \lambda^\mathsf{T} A - \mu^\mathsf{T} = 0 \\ -\infty & \text{Otherwise} \end{cases} \tag{22.6}$$

Clearly $D(\lambda, \mu)$ is only well-defined for (λ, μ) satisfying $c + A^\mathsf{T}\lambda - \mu = 0$. This actually poses an extra constraint on the dual problem. Therefore, the dual problem is

$$\begin{aligned} \text{minimize} \quad & -b^\mathsf{T}\lambda \\ \text{subject to} \quad & c + A^\mathsf{T}\lambda - \mu = 0 \\ & \mu \geq 0 \end{aligned} \tag{22.7}$$

Notice we can eliminate μ by using the relation $\mu = c + A^\mathsf{T}\lambda$. The dual problem is then compactly rewritten as

$$\begin{aligned} \text{minimize} \quad & -b^\mathsf{T}\lambda \\ \text{subject to} \quad & c + A^\mathsf{T}\lambda \geq 0 \end{aligned} \tag{22.8}$$

We can see the dual problem for LP (22.5) is just another LP.

22.2 Dual of SDP

Now we consider the following semidefinite program (SDP) problem.

$$\begin{aligned} \text{minimize} \quad & c^\mathsf{T} x \\ \text{subject to} \quad & x_1 F_1 + x_2 F_2 + \ldots + x_p F_p - G \leq 0 \end{aligned} \tag{22.9}$$

Here $x \in \mathbb{R}^p$ and we have

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

where x_i $(i = 1, 2, \ldots, p)$ is just scaler. Here F_i $(i = 1, 2, \ldots, p)$ and G are all symmetric matrices. The inequality $x_1F_1 + x_2F_2 + \ldots + x_pF_p - G \leq 0$ just means $(x_1F_1 + x_2F_2 + \ldots + x_pF_p - G)$ is a negative semidefinite matrix. To derive the dual problem for SDP, we need the matrix version of Lagrangian formulations. Recall that the term $\mu^\mathsf{T} g(x)$ in the Lagrangian can be viewed as an inner product between the Lagrangian multiplier μ and the constraint function $g(x)$. For the SDP problem, the Lagrangian multiplier is a matrix Y and the inner product between Y and $(x_1F_1 + x_2F_2 + \ldots + x_pF_p - G)$ is trace$(Y(x_1F_1 + x_2F_2 + \ldots + x_pF_p - G))$.

Let's explain the inner product of two matrices first. Consider two symmetric matrices A and B. If we put augment the entries of A as a vector and also augment all the entries of B as a vector, then clearly the inner product of these two resultant vectors is $\sum_{i,j} A_{ij}B_{ij}$. This sum can be compactly rewritten as trace AB where trace just denotes the sum of the diagonal entries of a given matrix. Therefore, the Lagrangian for (22.9) can be written as

$$\begin{aligned} L(x, Y) =& c^{\mathsf{T}} x + \operatorname{trace}(Y(x_1F_1 + x_2F_2 + \ldots + x_pF_p - G)) \\ =& -\operatorname{trace}(YG) + \sum_{i=1}^{p} x_i(c_i + \operatorname{trace}(YF_i)) \end{aligned}$$

where c_i is the i-th entry of c. We have

$$D(Y) = \min_{x\in\mathbb{R}^p} L(x, Y) = \begin{cases} -\operatorname{trace}(YG) & \text{if } c_i + \operatorname{trace}(YF_i) = 0 \\ -\infty & \text{Otherwise} \end{cases} \tag{22.10}$$

Therefore, the dual problem for SDP is

$$\begin{aligned} \text{minimize} \quad & -\operatorname{trace}(GY) \\ \text{subject to} \quad & \operatorname{trace}(F_iY) + c_i = 0, \ \forall i = 1, \ldots, p \\ & Y \geq 0 \end{aligned}$$

Here $Y \geq 0$ just states that Y is a positive semidefinite matrix. Clearly G, F_i, and c_i are all given, and Y is the decision variable for this dual problem.