Lecture 10 — September 30

Lecturer: Sanghavi *Scribe: Sara Mourad & Hardik Jain & Xiangkun Dai*

10.1 Last time

In the last few lectures, we have seen new algorithms for solving unconstrained minimization like Newton method and conjugate gradient method. However, many interesting problems in practice require to solve constrained minimization. In this lecture, we will introduce the concept of duality in context of constrained minimization. We would also look at the concepts of weak duality, strong duality and some applications of using duality.

10.2 Duality for linear program

10.2.1 Standard form for linear program

When the objective and constraint functions are all affine, the problem is called a *linear program* (LP). One such example is the following problem :

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Gx \succeq h \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{10.1}$$

where $G \in \mathbb{R}^{m\text{x}n}$ and $A \in \mathbb{R}^{p\text{x}n}$ and variable $x \in \mathbb{R}^n$.
The standard form of a linear programming problem is

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{10.2}$$

Any linear program with inequality constraint can be written in this standard form. For example, let's say we have an inequality constraint $Ax \succeq b$ for a linear program,

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax \succeq b \\ & x \geq 0 \end{aligned} \tag{10.3}$$

The inequality can be written as $-Ax \preceq -b$. Then, we can introduce a vector variable s and rewrite it as,

$$\begin{bmatrix} \min\limits_x & c^T x \\ \text{s.t.} & Ax \succeq b \\ & x \geq 0 \end{bmatrix} \equiv \begin{bmatrix} \min\limits_x & c^T x \\ \text{s.t.} & -Ax \preceq -b \\ & x \geq 0 \end{bmatrix} \equiv \begin{bmatrix} \min\limits_x & c^T x \\ \text{s.t.} & -Ax + s = -b \\ & s \geq 0 \\ & x \geq 0 \end{bmatrix} \tag{10.4}$$

With $A' = [-A \ I]$, $x' = \begin{bmatrix} x \\ s \end{bmatrix}$, $b' = -b$ and $c' = [c \ 0]$,

$$\begin{bmatrix} \min\limits_x & c^T x \\ \text{s.t.} & -Ax + s = -b \\ & s \geq 0 \\ & x \geq 0 \end{bmatrix} \equiv \begin{bmatrix} \min\limits_{x'} & c'^T x' \\ \text{s.t.} & A'x' = b' \\ & x' \geq 0 \end{bmatrix} \tag{10.5}$$

10.2.2 Dual Lagrangian function

Now, let's introduce Lagrangian function for the standard form of Linear program. The Lagrangian function for 10.2 can be defined as $L : \mathbb{R}^n \text{x} \mathbb{R}^p \to \mathbb{R}$,

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) \quad = c^T x + \sum_i \lambda_i (b_i - a_i^T x) \tag{10.6}$$

where a_i^T is the i^{th} row of A. The vector λ is the *dual variable* or *Lagrange multiplier vector* associated with the problem. Essentially, the Lagrangian function augments the objective function with a weighted sum of the constraint functions.

Lemma 10.1 (Minimax form of Primal). *Solving the original LP constrained problem (Primal problem) is equivalent to solving the unconstrained minimax problem*

$$\textit{(Primal)} \begin{bmatrix} \min\limits_x & c^T x \\ s.t. & A^T x = b \end{bmatrix} = \min_x \max_{\lambda \geq 0} L(x, \lambda) = \min_x \max_{\lambda \geq 0} c^T x + \sum_i \lambda_i (b_i - a_i^T x) \tag{10.7}$$

Proof (Lemma 10.1): The order of the variables is important. x is chosen first and for each value of x, the inner problem attempts to maximize the objective value. For any x not feasible, $a_i^T x < b_i \Rightarrow b_i - a_i^T x > 0$ and as $\lambda \geq 0$ can be chosen as large as required, the value of the term $\lambda_i(b_i - a_i^T x)$ in the inner problem will be $+\infty$. If the primal has a finite optimum, such an x cannot achieve the minimum value of the minimax. (If the optimal value of the primal is $+\infty$, the result holds trivially). So, the optimal $x^\star$ of Equation **??** will be feasible.

But, x feasible $\Rightarrow (b_i - a_i^T x) \leq 0 \Rightarrow \lambda_i = 0$ will be an optimal solution for the inner problem. Then, the minimax problem becomes

$$\min_x \max_{\lambda \geq 0} c^T x + \sum_i \underbrace{\lambda_i}_{=0} (b_i - a_i^T x) = \min_{x : a_i^T x \geq b_i} c^T x \tag{10.8}$$

So, the optimum value of the minimax will be equal to the optimum value of the primal. $\square$

Lemma 10.2 (Exchanging the order of the minimax).

$$\begin{gathered} \min_x \max_{\lambda \geq 0} L(x, \lambda) \geq \max_{\lambda \geq 0} \min_x L(x, \lambda) \\ \min_x \max_{\lambda \geq 0} c^T x + \sum_i \lambda_i (b_i - a_i^T x) \geq \max_{\lambda \geq 0} \min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x) \end{gathered} \tag{10.9}$$

Proof:

$$L(x, v) \geq \min_{y \geq 0} L(y, v) \; \forall x, v$$

$$\max_v L(x, v) \geq \max_v \min_{y \geq 0} L(y, v) \; \forall x \tag{10.10}$$

$$\min_x \max_v L(x, v) \geq \max_v \min_{y \geq 0} L(y, v)$$

$\square$

Lemma 10.3. *Solving the maxmin unconstrained problem is equivalent to solving the optimum of the dual problem*

$$\max_{\lambda \geq 0} \min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x) = \begin{bmatrix} \max_\lambda & \lambda^T b \\ s.t. & \sum_{i=1}^m \lambda_i a_i = c \\ & \lambda \geq 0 \end{bmatrix} \textit{(Dual)} \tag{10.11}$$

Proof:

$$\max_{\lambda \geq 0} \underbrace{\min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x)} \equiv \max_{\lambda \geq 0} q(\lambda) \tag{10.12}$$

with

$$q(\lambda) = \min_x c^T x + \sum_i \lambda_i \left(b_i - a_i^T x\right) = \left(c^T - \sum_i \lambda_i a_i^T\right) x + \sum_i \lambda_i b_i \tag{10.13}$$

Now, $\mathrm{dom}(q(\lambda)) := \{\lambda : q(\lambda) < +\infty\}$. So,

$$\mathrm{dom}(q(\lambda)) = \{\lambda : \lambda \geq 0 \wedge c = \sum_i \lambda_i a_i\} \tag{10.14}$$

Again, this holds because of the order. $\sum_i \lambda_i a_i^T \neq c^T \Rightarrow (c^T - \sum_i \lambda_i a_i^T) \neq 0 \Rightarrow q(\lambda) = -\infty$ $\square$

In Summary,

$$
\text{(P)} \begin{bmatrix} \min\limits_x & c^T x \\ \text{s.t.} & a_i^T x \geq b_i, 1 \leq i \leq m \end{bmatrix} = \min_x \max_{\lambda \geq 0} c^T x + \sum_i \lambda_i (b_i - a_i^T x)
$$

$$
\geq \max_{\lambda \geq 0} \min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x)
$$

$$
= \begin{bmatrix} \max\limits_\lambda & \lambda^T b \\ \text{s.t.} & \sum\limits_{i=1}^m \lambda_i a_i = c \\ & \lambda \geq 0 \end{bmatrix} \text{(D)}
$$

Corollary 10.4. *If x is feasible in the Primal problem, and v is feasible in the Dual problem, then* $c^T x \geq -b^T v$

10.3 Duality for general constrained optimization

Let's take a general constrained optimization problem,

$$
\begin{aligned}
\min_x \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0 \text{ for i = 1 to m} \\
& h_j(x) = 0 \text{ for j = 1 to p}
\end{aligned} \tag{10.15}
$$

We can define a Lagrangian function L,

$$
L(x, \lambda, \nu) = f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_i(x)
$$

where λ and ν are the Lagrange multipliers associated respectively with inequality and equality constraints.
So, the original problem is equivalent to solving :

$$
\min_x \max_{\lambda \geq 0, \nu} L(x, \lambda, \nu) \tag{10.16}
$$

From Lemma 10.2,

$$
\min_x \max_{\lambda \geq 0, \nu} L(x, \lambda, \nu) \geq \max_{\lambda \geq 0, \nu} \min_x L(x, \lambda, \nu) \tag{10.17}
$$

Let's define the dual function $g(\lambda, \nu) = \min\limits_x L(x, \lambda, \nu)$
The RHS of **??** is known as the dual problem:

$$
\begin{aligned}
\max_\nu \quad & g(\lambda, \nu) \\
\text{s.t.} \quad & \lambda \geq 0
\end{aligned} \tag{10.18}
$$

Claim : $g(\lambda, \nu)$ is concave.
Proof : $g(\lambda, \nu)$ is the pointwise minimum of $L(x, \lambda, \nu)$ over x. For each x, $L(\lambda, \nu)$ is linear

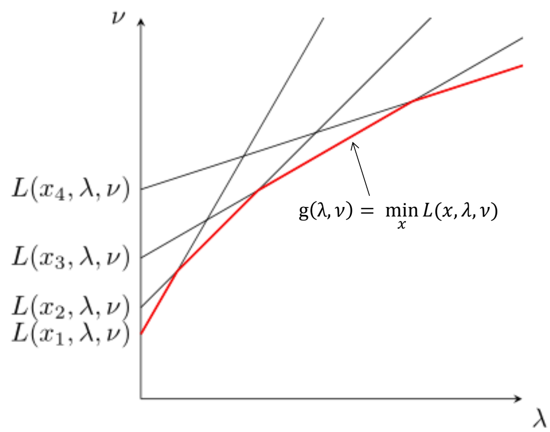

Figure 10.1.

in λ and ν so it is a concave function.Therefore, $g(\lambda, \nu)$ is concave as the pointwise minimum of concave functions is concave.
Let's denote by d^* the optimal dual point, i.e. the maximum of the concave function $g(\lambda, \nu)$, it is now possible to upper bound and lower bound the primal optimal point p^* :
$p^* \leq f_0(\hat{x})$, $\hat{x}$ being any primal feasible point, and $p^* \geq g(\lambda, \nu)$, λ and ν being any dual feasible point. In particular, $p^* \geq d^*$.

Lemma 10.5. *While the primal of an optimization problem may not be convex, the dual is always concave.*

Exercise: Prove this. (Hint: Assume $\mathfrak{X}$ is finite, $\mathfrak{X} = \{x_1, x_2, x_3, x_4, x_5\}$ and draw the picture)
Consequence: Since the dual is concave, the dual of the dual is also convex. We will later see that the dual of the dual is, in a precise sense, a convex relaxation of the primal, if the primal fails to be convex.

10.3.1 Slater's constraint Qualification Condition

Consider the convex optimization problem,

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \\ & Ax = b \end{aligned} \tag{10.19}$$

then, if $\exists$ feasible x such that $f_i(x) < 0\ \forall i$ then, we have that $p^* = d^*$. That is, the dual optimum corresponds to primal optimum.

Proof: For the detailed proof of Slater's condition, please refer to *Convex analysis and optimization* by D. Bertsekas. The basic idea is that for a convex set, if there exists a point which lies outside the set, we could find a hyperplane to seperate support the set. Actually, when f, g are convex, $C = \textbf{epigraph}(f(x) - f^*, g(x))$ is also convex. Note that $(0,0) \notin \textbf{relint}\, C$, so when slater's condition are satisfied, we could find hyperplane $(1, \lambda), \lambda \geq 0$ such that $f(x) - f^* + \lambda g(x) \geq 0$. $\square$

The insight that we gained from the proof of Slater's condition is that strong duality typically doesn't hold for non-convex problems. It's easy to give such examples.

A note on weak and strong duality : Weak duality refers to the case where $d^* \leq p^*$. Weak duality holds for convex and non-convex problems, and can be used to find non-trivial lower bounds for difficult problems.
Strong duality refers tot he case where $d^* = p^*$. Strong duality doesn't hold in general, but it is guaranteed in convex problems when the constraint qualifications are satisfied.
Next, we give two examples where Slater's condition is not satisfied and strong duality doesn't hold in the same time.
Example 1: Considering the following optimization problem:

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1 + x_2 \\ \text{s.t.} \quad & (x_1 + 1)^2 + x_2^2 - 1 = 0, \\ & (x_1 - 2)^2 + x_2^2 - 4 = 0 \end{aligned} \tag{10.20}$$

It's obvious that the feasible set is $\mathfrak{X} = (0,0)$. Figure **??** shows that the constraints are two circles with only one intersect point. Thus, the optimal solution $x^* = (0,0)$. We know that for convex optimization, the optimal solution should satisfy $0 \in \nabla f(x^*) + N_{\mathfrak{X}}(x^*)$, where $N_{\mathfrak{X}}(x^*)$ is the norm cone at x^*. Remember that this is always true. However, let us consider another expression $0 \in \nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$. In this case,$\nabla f(x^*) = (1,1)^T$, $\nabla g_1(x^*) = (2,0)^T$, $\nabla g_2(x^*) = (-4,0)^T$. We can't find such λ_1, λ_2 that:

$$0 \in \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -4 \\ 0 \end{pmatrix}.$$

This tells us the latter expression doesn't always holds which means the norm cone cannot always be written as the linear combination of constraints' gradient. For this example, it's easy to check that Slater's condition doesn't hold and there is no guarantee of strong duality.
Example 2: Let's consider this problem:

$$\begin{aligned} \min_{x} \quad & e^{x_2} \\ \text{s.t.} \quad & \|x\|_2 - x_1 \leq 0 \end{aligned} \tag{10.21}$$

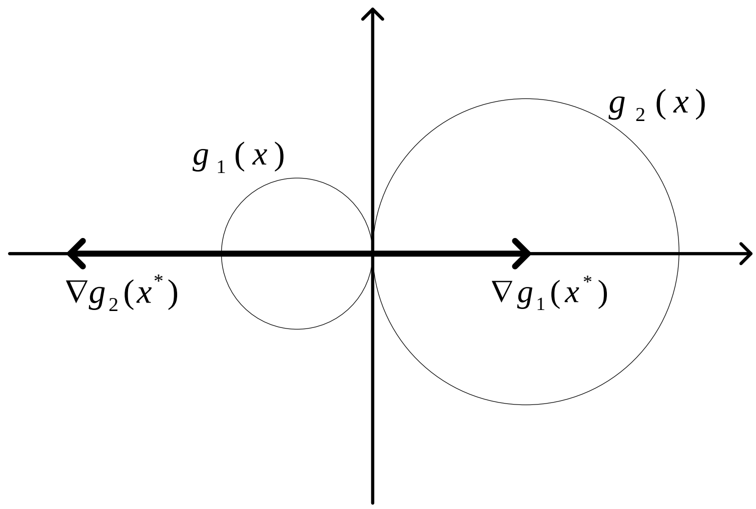

Figure 10.2. Illustration of Example 1

From the constraint we know the feasible region $X = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\}$. The constraint is always active so Slater's condition is not satisfied. Next we show that strong duality also doesn't hold. The dual:

$$q(\lambda) = \inf_x \left(e^{x_2} + \lambda(\|x\|_2 - x_1) \right). \tag{10.22}$$

It's obvious that $e^{x_2} + \lambda(\|x\|_2 - x_1) \geq 0$. Let $x_1 = x_2^4$. We have:

$$e^{x_2} + \lambda(\|x\|_2 - x_1) = e^{x_2} + \lambda x_2^4(\sqrt{1 + \frac{1}{x_2^6}} - 1).$$

When $x_2 \to -\infty$, $e^{x_2} \to 0$, $x_2^4\sqrt{1 + \frac{1}{x_2^6}} - 1 \sim \frac{1}{2x_2^2} \to 0$. Thus $q(\lambda) = 0$ for any $\lambda \geq 0$.

The previous two examples show that strong duality doesn't hold when Slater's condition is not satisfied. But it's worth to note that Slater's condition is just sufficient, not necessary. It's possible that strong duality holds when Slater's condition is not satisfied.

10.4 Applications

10.4.1 Example 1: Matrix Games

Let's say two players are playing a zero sum game. The player I gets to choose a row of the payoff matrix $P \in \mathbb{R}^{mxn}$. Player II gets to choose a column of the payoff matrix. The payoff is decided based on the row and column chosen by both the players. For example if player I chooses k^{th} row and player II chooses r^{th} column, then the payoff, which is from player I to player II, would be $P[k][r]$.
Let's now assume that the players are allowed to choose a vector describing the probability

distribution over the rows/columns. That is, player I can have a vector $u = (u_1, u_2, ..., u_k)$ and player II can have a vector $v = (v_1, v_2, ..., v_r)$. The vectors u and v are both probability distribution of choices of player I and player II respectively. These vectors follow the equation, $\mathbf{1}^T u = 1, \mathbf{1}^T v = 1, u \succeq 0, v \succeq 0$.
So, the expected payoff becomes,

$$u^T P v$$

Let's analyze the two cases:

- **Player I plays "first"**: Since player I decides u first, player II knows u and chooses v in order to maximize $u^T P v$. Player II can choose the maximum element of $P^T u$. Player I's best action is to choose u that minimizes his payoff. So, Player I's can find u by solving,

$$\begin{aligned} \min_u \max_i \quad & (P^T u)_i \\ s.t. \quad & u \geq 0 \\ & \mathbf{1}^T u = 1 \end{aligned} \tag{10.23}$$

- **Player II plays "first"**: Similar to the previous case, if player II plays first, his best strategy is to minimize the following equation :

$$\begin{aligned} \min_v \max_i \quad & (P^T v)_i \\ s.t. \quad & v \geq 0 \\ & \mathbf{1}^T v = 1 \end{aligned} \tag{10.24}$$

Let's now consider equation **??** and write it in the standard form,

$$\begin{bmatrix} \min\limits_x \max\limits_i & (P^T u)_i \\ \text{s.t.} & \mathbf{1}^T u = 1 \end{bmatrix} \equiv \begin{bmatrix} \min\limits_t & t & \\ \text{s.t.} & t \geq (P^T u)_i & \forall\ i \\ & \mathbf{1}^T u = 1 & \end{bmatrix} \tag{10.25}$$

We can now define the Lagrangian function as:

$$L(u, t, \lambda, \mu, \nu) = t + \lambda^T (P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - \mathbf{1}^T u) \tag{10.26}$$

The dual problem then becomes,

$$\begin{aligned} \max_\nu \quad & \nu \\ \text{s.t.} \quad & \lambda \geq 0 \\ & \mathbf{1}^T \lambda = 1 \\ & (P\lambda)_i - \nu_i \geq 0 \end{aligned} \tag{10.27}$$

This is clearly equivalent to **??**. Since the LPs are feasible, we have strong duality; so the optimal values of **??** and **??** are equal. Hence, it doesn't matter who plays first in this game.