EE 381V: Large Scale Optimization **Fall 2012**

Lecture 11 — October 4

Lecturer: Caramanis & Sanghavi *Scribe: Harsh Pareek*

These scribes notes are for the ghost lecture on October 4. The lecture video and instructor's notes are on Blackboard.

11.1 Last time

Last time we recapped optimality conditions in constrained problems. Recall the constrained convex optimization problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & x \in \mathfrak{X} \end{aligned} \tag{11.1}$$

where f is a convex function and $\mathfrak{X}$ is a convex set. $x^\star$ is a minimum for this program iff every feasible direction is an ascent direction:

$$0 \in \nabla f(x^\star) + N_{\mathfrak{X}}(x^\star) \tag{11.2}$$

where $N_{\mathfrak{X}}(x^\star)$ is the normal cone at $x^\star$. Also recall that we write it in this particular form to compare it with the unconstrained case where the optimality condition is $\nabla f(x^\star) = 0$. This form will also be useful when we generalize f to be a non-smooth function and its gradient will be replaced by a set.

In particular, we applied this to Linear Programming. In this lecture, we will prove weak and strong duality results for Linear Programs. Consider the linear program

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & a_i^T x \geq b_i, i = 1, 2, \ldots, m \end{aligned} \tag{11.3}$$

Figure 11.1 shows an example linear program with $x \in \mathbb{R}^2$ and $m = 6$ constraints.

11.2 Weak Duality

Recall that the normal cone is the polar of the tangent cone.

$$\begin{aligned} & N_{\mathfrak{X}}(x^\star) = (T_{\mathfrak{X}}(x^\star))^\circ \\ \Rightarrow\ & N_{\mathfrak{X}}(x^\star) = \{v : \langle v, x - x^\star \rangle \leq 0,\ \forall x \in \mathfrak{X}\} \end{aligned}$$

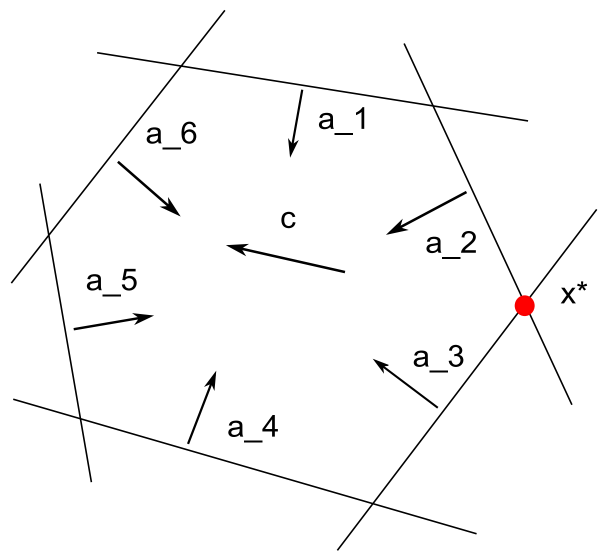

Figure 11.1. Example of a linear program. We can see from the figure that the optimality condition clearly holds, namely, $-\nabla f(x) \in N_{\mathcal{X}}(x)$.

For LPs, $\mathfrak{X}$ is defined by linear inequalities,

$$= \{v : \langle v, x - x^\star \rangle \leq 0,\ \forall x : a_i^T x \geq b_i\}$$

In Figure 11.1, at $x^\star$, $a_2^T x^\star = b_2$ and $a_3^T x^\star = b_3$ and the other inactive constraints are not needed,

$$= \{v : \langle v, x - x^\star \rangle \leq 0,\ \forall x : a_2^T x \geq b_2, a_3^T x \geq b_3\}$$

Subtracting off b_is and letting $z = x - x^\star$

$$= \{v : \langle v, z \rangle \leq 0,\ \forall z : a_2^T z \geq 0, a_3^T z \geq 0\}$$
$$\Rightarrow N_{\mathfrak{X}}(x^\star) = \{v : v = \lambda_2 a_2 + \lambda_3 a_3, \lambda_2, \lambda_3 \leq 0\}$$

Thus, the normal cone is the cone generated by the normals of the active constraints at the optimal point $x^\star$.

The gradient of f is $\nabla f(x^\star) = c$. Therefore, by Equation 11.2 we have

$$-c \in N_{\mathfrak{X}}(x^\star) \tag{11.4}$$
$$\Rightarrow -c = \lambda_2 a_2 + \lambda_3 a_3, \lambda_i \leq 0 \tag{11.5}$$
$$\Rightarrow c = \lambda_2 a_2 + \lambda_3 a_3, \lambda_i \geq 0 \tag{11.6}$$

λ_2 and λ_3 satisfy the following properties:

1. $\lambda_2 a_2 + \lambda_3 a_3 = c$
2. $\lambda_2, \lambda_3 \geq 0$
3. Optimality (discussed in the next paragraph)
4. Duality: Multiplying Equation 11.6 by $x^\star$, we get,

$$c^T x^\star = \lambda_2 a_2^T x^\star + \lambda_3 a_3^T x^\star \tag{11.7}$$

As $x^\star$ is on these constraints, $a_2^T x^\star = b_2$ and $a_3^T x^\star = b_3$

$$\Rightarrow c^T x^\star = \lambda_2 b_2 + \lambda_3 b_3 \tag{11.8}$$
$$\Rightarrow c^T x^\star = \lambda^T b \tag{11.9}$$

for $\lambda = (0, \lambda_2, \lambda_3, 0, 0, 0)$

Now, we describe the optimality of the λ_i. Consider any variables $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ with one λ_i per constraint $a_i^T x \geq b_i$. Observe the following: If $\lambda_i \geq 0$ and $\sum_i \lambda_i = c$, then we will have a lower bound on our objective $c^T x$ as, for any feasible x,

$$\left(\sum_i \lambda_i a_i^T\right) x = c^T x \text{ and } a_i^T x \geq b_i$$
$$\Rightarrow c^T x \geq \lambda^T b$$

Call any x which satisfies $a_i^T x \geq b_i$ *Primal feasible* and any λ that satisfies $\lambda \geq 0$ and $\sum_i \lambda_i a_i = c$ *Dual feasible*. Then, we have, for primal feasible x and dual feasible λ

$$c^T x \geq \lambda^T b \tag{11.10}$$

This gives the following theorem.

Theorem 11.1 (Weak Duality). *For all Primal feasible x and Dual feasible λ,*

$$\text{(Primal)} \begin{bmatrix} \min\limits_x & c^T x \\ s.t. & a_i^T x \geq b_i, 1 \leq i \leq m \end{bmatrix} \underset{\text{Weak Duality}}{\geq} \begin{bmatrix} \max\limits_\lambda & \lambda^T b \\ s.t. & \sum_{i=1}^m \lambda_i a_i = c \\ & \lambda \geq 0 \end{bmatrix} \text{(Dual)} \tag{11.11}$$

Proof (Using Lagrange Multipliers): We will now give an alternate proof of this theorem using Lagrange Multipliers.

Lemma 11.2 (Minimax form of Primal). *The optimal value of the (Primal) is the same as the following unconstrained minimax problem.*

$$\min_x \left[\max_{\lambda \geq 0} \left(c^T x + \sum_i \lambda_i (b_i - a_i^T x) \right)\right] \tag{11.12}$$

Proof (Lemma 11.2): 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 11.12 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

$$\min_x \max_{\lambda \geq 0} c^T x + \sum_i \underbrace{\lambda_i}_{=0} (b_i - a_i^T x) = \min_{x: a_i^T x \geq b_i} c^T x \tag{11.13}$$

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

Exercise: Show that you can exchange the order of the terms in the minimax to get $\max \min g \geq \min \max g$ for any g. (Hint: The inside term has more “power”).

So, we have that,

Lemma 11.3 (Exchanging the order of the minimax).

$$\min_x \max_{\lambda \geq 0} c^T x + \sum_i \lambda_i (b_i - a_i^T x) \geq \max_{\lambda \geq 0} \min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x) \tag{11.14}$$

We now look at conditions imposed by the RHS on λ and show that it is equal to the optimum of the dual problem. Rewriting,

$$\max_{\lambda \geq 0} \underbrace{\min_x c^T x + \sum_i \lambda_i (b_i - a_i^T x)} \equiv \max_{\lambda \geq 0} q(\lambda) \tag{11.15}$$

with

$$q(\lambda) = \min_x c^T x + \sum_i \lambda_i \left(b_i - a_i^T x\right) = \left(c^T - \sum_i \lambda_i a_i^T\right) x + \sum_i \lambda_i b_i \tag{11.16}$$

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

$$\mathrm{dom}(q(\lambda)) = \{\lambda : \lambda \geq 0 \wedge c = \sum_i \lambda_i a_i\} \tag{11.17}$$

Again, this holds because of the order. $\sum_i \lambda_i a_i^T \neq c^T \Rightarrow (c^T - \sum_i \lambda_i a_i^T) \neq 0 \Rightarrow q(\lambda) = -\infty$. Finally, this implies that the optimum of the RHS of Equation 11.14 is equal to the optimum of the dual problem.

Summarizing,

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

□

The natural question in the above setting is when equality holds in the above equation. This is called strong duality, and as we will show in the next section, it always holds for linear programs, though it does not hold in general for optimization problems.

11.2.1 Weak Duality in General Optimization Programs

We can follow the same steps as in the proof of Theorem 11.1 for a general program (not necessarily convex) and show the following:

$$
\begin{aligned}
\text{(Primal)} \begin{bmatrix} \min\limits_x & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in \mathfrak{X} \end{bmatrix} &= \min_{x \in \mathfrak{X}} \max_{\mu \geq 0, \lambda} \left[f(x) + \lambda^T h(x) + \mu^T g(x) \right] \\
&= \min_{x \in \mathfrak{X}} \max_{\mu \geq 0, \lambda} \mathcal{L}(x, \lambda, \mu) \\
&\geq \max_{\mu \geq 0, \lambda} \underbrace{\min_{x \in \mathfrak{X}} \mathcal{L}(x, \lambda, \mu)}_{q(\lambda, \mu)} \\
&= \begin{bmatrix} \max\limits_{\lambda, \mu} & q(\lambda, \mu) \\ \text{s.t.} & \mu \geq 0 \end{bmatrix} \text{(Dual)}
\end{aligned}
$$

This holds for any optimization problem without any conditions on $f, g, h, \mathfrak{X}$ and we can obtain the dual $q(\lambda, \mu)$. Thus every optimization problem has a dual.

Further, we can also show that,

Lemma 11.4. *While the primal of an optimization problem may not be convex, the dual is always convex*

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 convex, 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.

11.3 Strong Duality

In this section, we will show strong duality for linear programs. First, we will rewrite the linear program in the "Standard Form".

$$\begin{aligned}
\begin{bmatrix} \min\limits_{x} & c^T x \\ \text{s.t.} & Ax \geq b \end{bmatrix} &\equiv \begin{bmatrix} \min\limits_{x} & c^T x \\ \text{s.t.} & \tilde{A}x \leq \tilde{b} \end{bmatrix} \\
&\equiv \begin{bmatrix} \min\limits_{x} & c^T x \\ \text{s.t.} & \tilde{A}x + s = \tilde{b} \\ & s \geq 0 \end{bmatrix}
\end{aligned}$$

With a different A, x and b, this can be rewritten

$$\begin{aligned}
\min_{x} \quad & c^T x \\
\text{s.t.} \quad & Ax = b \\
& x \geq 0
\end{aligned} \tag{11.18}$$

This is known as the Standard Form LP.

Exercise: Derive the dual of the Standard form LP

Theorem 11.5 (Strong Duality). *If either (P) or (D) has a finite optimal solution, then the other is feasible and has a finite optimal solution. Moreover, Strong Duality holds, $(P) = (D)$*

Corollary 11.6. *If either problem has unbounded objective, then the other is infeasible.*

Exercise: Prove the corollary. (Hint: Use Weak Duality)

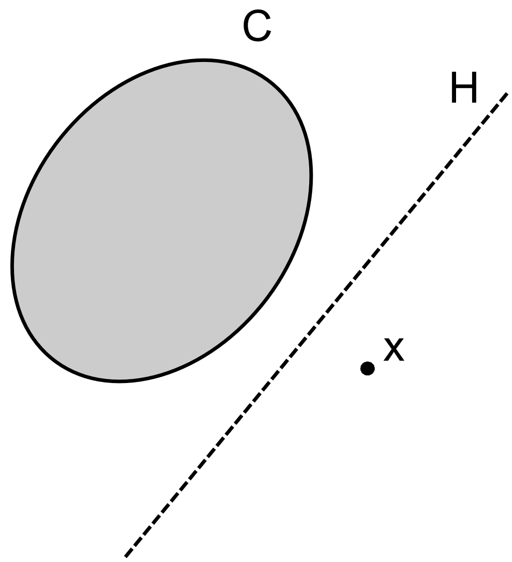

Figure 11.2. Separation of x from C by H

Proof (Proof of Strong Duality): The key to this result is the separation theorem (Figure 11.2) from Lecture 3: If C is a closed convex set and $x \notin C$, then we can separate x from C by a hyperplane H such that $C \subseteq H^+$ and $x \in \text{int}H^-$. Suppose that (P) is feasible with finite optimal value z_p.

Define:

$$C = \{(r, w) \in \mathbb{R}^{m+1} : r = tz_p - c^T x \in \mathbb{R}, w = tb - Ax \in \mathbb{R}^m, x \geq 0, t \geq 0\} \tag{11.19}$$

Recall that A is an $m \times n$ matrix where m is the number of constraints and n is the dimension of the optimization variable x

Claim: C is a convex set. In fact, it is a closed convex cone.

Exercise: Prove this claim. Check that C is a cone. (Hint: Prove that $(r, w) \in C \Rightarrow \alpha(r, w) \in C \; \forall \alpha$)

Lemma 11.7. *For C as defined in Equation 11.19,* $(1, 0) \notin C$

We will prove this lemma after the main theorem.

Corollary 11.8 (Consequence of Lemma 11.7). $\exists H$ *with* $(1, 0) \in \text{int}(H^-), C \subseteq H^+$.

As H is a hyperplane, it is defined by its normal vector and offset.

$$H = \{(r, w) : \left\langle \underbrace{(\lambda_0, \lambda_m)}_{\text{normal}}, (r, w) \right\rangle = \underbrace{\beta}_{\text{offset}}, \lambda_0 \in \mathbb{R}, \lambda_m \in \mathbb{R}^m\} \tag{11.20}$$

Then,

$$C \subseteq H^+ \Rightarrow \lambda_0 r + \langle \lambda_m, w \rangle \geq \beta \; \forall (r, w) \in C$$

Note that, C is a cone $\Rightarrow \beta = 0$ Also,

$$(1,0) \in \operatorname{int}(H^-) \Rightarrow \langle(\lambda_0, \lambda_m), (1,0)\rangle < \beta \Rightarrow \lambda_0 + 0 < \beta = 0 \Rightarrow \lambda_0 < 0$$

As $\lambda_0 < 0$ and $\beta = 0$, we can scale λ_0, λ_m to assume WLOG that $\lambda_0 = -1$ Now,

$$C \subseteq H^+ \Rightarrow -r + \lambda_m^T w \geq 0 \ \forall (r, w) \in C$$

By definition of C (Equation 11.19),

$$\begin{aligned} &\Rightarrow c^T x - t z_p + \lambda_m^T (tb - Ax) \geq 0 \ \forall t, x \geq 0 \\ &\Rightarrow (c^T - \lambda_m^T A)x - t(z_p - \lambda_m^T b) \geq 0 \ \forall t, x \geq 0 \end{aligned}$$

Since this holds for all x, t, we can set specific values to get,

$$t = 0 \Rightarrow \lambda_m^T A \leq c \tag{11.21}$$

So λ_m is dual feasible. By weak duality, $\lambda_m^T b \leq z_p$

$$x = 0, t = 1 \Rightarrow z_p \leq \lambda_m^T b \tag{11.22}$$

Combining, we get, $\lambda^T b = z_p = c^T x^\star \Rightarrow (P) = (D)$. So we have strong duality. $\square$

Finally, we prove that $(1,0) \notin C$ to complete the proof.

Proof (Proof of Lemma 11.7): Suppose to the contrary $(1,0) \in C$.

By the definition of C (Equation 11.19), this implies,

$$1 = t_0 z_p - c^T x_0 \text{ and } 0 = t_0 b - A x_0$$

for some $t_0 \in \mathbb{R} \geq 0, x_0 \in \mathbb{R}^m \geq 0$ If $t_0 > 0$, $\hat{x} = x_0/t_0$ is a feasible for the LP Equation 11.18 (as $A\hat{x} = b$ and $\hat{x} \geq 0$). As z_p is the optimal value,

$$c^T \hat{x} \geq z_p \Rightarrow 1 = t_0 z_p - c^T x_0 \leq 0$$

This contradiction $\Rightarrow t_0 = 0$ Then,

$$t_0 = 0 \Rightarrow 1 = t_0 z_p - c^T x_0 \Rightarrow 1 = -c^T x_0$$

and

$$0 = -A x_0, x_0 \geq 0$$

i.e. A has a nontrivial null space and $\exists x_0, x_0 \geq 0$ and $c^T x_0 < 0$, So, if $\tilde{x}$ is any feasible solution, then $\tilde{x} + \alpha x_0$ is feasible for any $\alpha > 0 \Rightarrow$ the primal (P) is unbounded. This is a contradiction. Hence, $(1,0) \notin C$. $\square$

In this class we established weak and strong duality for LPs. In the next class, we will discuss duality for other classes of functions.