Lecture 17 (Oct. 17): LP Duality

Lecturer: Zachary Friggstad *Scribe: Zachary Friggstad*

17.1 Certificates of Optimality

Consider the linear program

$$\begin{array}{rrcl} \text{maximize :} & \mathbf{c}^T \cdot \mathbf{x} & & \\ \text{subject to :} & \mathbf{A} \cdot \mathbf{x} & \leq & \mathbf{b} \\ & \mathbf{x} & \geq & \mathbf{0} \end{array}$$

with

$$\mathbf{A} = \begin{pmatrix} 4 & 5 & 1 \\ 1 & 8 & 1 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \text{ and } \mathbf{c} = \begin{pmatrix} 3 \\ 5 \\ 1 \end{pmatrix}.$$

Consider the following feasible solution with value $\mathbf{c}^{\mathrm{T}}\overline{\mathbf{x}} = 3$

$$\overline{x} = \begin{pmatrix} 0 \\ 1/3 \\ 4/3 \end{pmatrix}.$$

I claim that $\overline{\mathbf{x}}$ is an optimal solution, but how can I convince you easily?

Here is how to construct upper bounds. The first one constructs a somewhat weak upper bound, but then we discuss how to automate the process to find better upper bounds.

Consider the row vector $\mathbf{z} = \frac{1}{2}\mathbf{A}_1 + \mathbf{A}_2 = (3, 10\frac{1}{2}, 1\frac{1}{2})$. Note $\mathbf{z} \geq \mathbf{c}$. Therefore, for any feasible $\mathbf{x}'$ we have

$$\mathbf{c}^T\mathbf{x}' \leq \mathbf{z}\mathbf{x}' = \frac{1}{2}\mathbf{A}_1\mathbf{x}' + \mathbf{A}_2\mathbf{x}' \leq \frac{1}{2}\mathbf{b}_1 + \mathbf{b}_2 = 5\frac{1}{2}.$$

So the optimum value of this LP is somewhere between 3 and $5\frac{1}{2}$.

Can we come up with a better bound this way? The key parts are that we formed $\mathbf{z}$ as a nonnegative linear combination of the rows of $\mathbf{A}$ and that $\mathbf{z} \geq \mathbf{c}$. These are linear constraints, and the goal is to minimize the corresponding linear combination of $\mathbf{b}$ to get the best possible upper bound. More precisely, let $\mathbf{y}$ denote a vector over $\mathbb{R}^m$ of variables. Thinking of these as the nonnegative coefficients in the linear combination of the rows of $\mathbf{A}$ we arrive at the following **dual** linear program.

$$\begin{array}{rrcl} \text{minimize :} & \mathbf{b}^T \cdot \mathbf{y} & & \\ \text{subject to :} & \mathbf{A}^T \cdot \mathbf{x} & \geq & \mathbf{c} \\ & \mathbf{y} & \geq & \mathbf{0} \end{array}$$

Theorem 1 (Weak Duality) *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n$. *Let* $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}\}$ *be the* **primal polyhedron** *and* $\mathcal{D} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}^T \cdot \mathbf{y} \geq \mathbf{c}\}$ *be the* **dual polyhedron**.

If $\mathcal{P} \neq \emptyset$ and $\mathcal{D} \neq \emptyset$ then

$$\max_{\overline{\mathbf{x}} \in \mathcal{P}} \mathbf{c}^T \cdot \overline{\mathbf{x}} \leq \min_{\overline{\mathbf{y}} \in \mathcal{D}} \mathbf{b}^T \cdot \mathbf{y}.$$

Proof. Let $\overline{\mathbf{x}} \in \mathcal{P}, \overline{\mathbf{y}} \in \mathcal{D}$. Then

$$\mathbf{c}^T \cdot \mathbf{x} \leq (\mathbf{A}^T \cdot \mathbf{y})^T \cdot \mathbf{x} = \mathbf{y}^T \cdot \mathbf{A} \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{b}.$$

The first inequality holds because $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{A}^T \cdot \mathbf{y} \geq \mathbf{c}$. The second inequality holds because $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$. ■

Next lecture, we will see something stronger. That the primal has a finite optimum if and only if the dual has a finite optimum[1]. Furthermore, in this case the optimum solutions to the respective linear programs have the same objective function value (i.e. the weak duality bound above holds with equality if you consider optimal solutions).

In the example above, simply taking $\mathbf{y} = (1, 0)^T$ yields a feasible dual solution with value 3, so we now know the optimum primal (and dual) value for the example is in fact 3.

17.2 Duals of Arbitrary LPs

We talked about how to construct the dual of an LP in standard form. We also have rules for converting an arbitrary LP into an LP in standard form. So every LP has a dual (not just those in standard form), but it would be nice to compute the dual without going through the entire conversion to standard form.

In the following discussion, we call a variable $\mathbf{x}_i$ **unconstrained** in a given LP if $\mathbf{x}_i \geq 0$ is not a constraint. We also assume, by negating if necessary, that all inequality constraints in a minimization LP are of the form $\mathbf{A}_i \cdot \mathbf{x} \geq \mathbf{b}_i$ and all inequality constraints of a maximization LP are of the form $\mathbf{A}_i \cdot \mathbf{x} \leq \mathbf{b}_i$.

This table presents the information as if the primal was the minimization LP. If the primal is a maximization LP, just swap the columns.

Primal	$\leftrightarrow$	**Dual**
min	$\leftrightarrow$	max
variables	$\leftrightarrow$	CONSTRAINTS
constraints	$\leftrightarrow$	VARIABLES
$\mathbf{x}_j \geq 0$	$\leftrightarrow$	$\mathbf{A}_j^T \cdot \mathbf{y} \leq \mathbf{c}_j$
$\mathbf{x}_j$ unconstrained	$\leftrightarrow$	$\mathbf{A}_j^T \cdot \mathbf{y} = \mathbf{c}_j$
$\mathbf{A}_i \cdot \mathbf{x} \geq \mathbf{b}_i$	$\leftrightarrow$	$\mathbf{y}_i \geq 0$
$\mathbf{A}_i \cdot \mathbf{y} = \mathbf{b}_i$	$\leftrightarrow$	$\mathbf{y}_i$ unconstrained

For example, consider the following primal LP (left) and its corresponding dual LP (right).

$$\begin{array}{rrcl} \text{minimize}: & 2x_2 & & \\ \text{subject to}: & x_1 + 2x_2 & \geq & 3 \\ & 3x_1 + 4x_2 & \leq & 8 \\ & 5x_1 + 6x_2 & = & 10 \\ & x_1 & \geq & 0 \end{array}$$

Note x_2 is unconstrained.

[1] For a feasible minimization LP, we say it is unbounded if one can find solutions of arbitrarily low cost and, otherwise, we say it has a finite optimum.

To apply the above conversion, we negate any $\leq$ constraint so all constraints are of the form $\geq$ or $=$ (in a maximization LP, we would negate some constraints to ensure they are all of the form $\leq$ or $=$). This produces the equivalent LP

$$\begin{array}{rrcl} \text{minimize :} & 2x_2 & & \\ \text{subject to :} & x_1 + 2x_2 & \geq & 3 \\ & -3x_1 - 4x_2 & \geq & -8 \\ & 5x_1 + 6x_2 & = & 10 \\ & x_1 & \geq & 0 \end{array}$$

There are 3 constraints (apart from nonnegativity), so we use three dual variables y_1, y_2, y_3. The dual LP is then

$$\begin{array}{rrcl} \text{maximize :} & 3y_1 - 8y_2 + 10y_3 & & \\ \text{subject to :} & y_1 + 3y_2 + 5y_3 & \leq & 0 \\ & 2y_1 + 4y_2 + 6y_3 & = & 2 \\ & y_1, y_2 & \geq & 0 \end{array}$$

Note y_3 is unconstrained.

So the basic idea is the same as for a primal LP in standard form. The dual uses the transposed constraint matrix and the role of $\mathbf{b}$ and $\mathbf{c}$ swap. The particulars of which constraints are inequalities an which are equalities, as well as which variables are bound to be nonnegative and which are unconstrained, can be read off from the table above.

It is straightforward to show weak duality holds in this case.

17.3 A Separation Theorem for Convex Bodies

There are two commonly-followed approaches to proving strong duality (i.e. primal optimum = dual optimum if they are both feasible). One is seen as a consequence of why the simplex algorithm terminates. This would take too much effort to build up in this class, as the only LP solver we will see uses the ellipsoid method. Another is a surprisingly quick application of a general-purpose theorem about convex bodies. We now do this. In fact, the bulk of the effort will be in low-level details about basic topology in $\mathbb{R}^n$.

Definition 1 *A subset $\mathcal{P} \subseteq \mathbb{R}^n$ is* **convex** *if for any $\mathbf{x}, \mathbf{y} \in \mathcal{P}$ and any $0 \leq \lambda \leq 1$ we have $\lambda \cdot \mathbf{x} + (1 - \lambda) \cdot \mathbf{y} \in \mathcal{P}$.*

In other words, the line segment between any two points of $\mathcal{P}$ is completely contained in $\mathcal{P}$. The current assignment has you show that the set of feasible solutions to every linear program is convex.

We briefly need one standard definition from topology. Don't worry if you are not familiar with this field, this is only a very brief use of it and we will not use it again after proving strong duality.

Definition 2 *A subset $\mathcal{P} \subseteq \mathbb{R}^n$ is* **closed** *if for any $\mathbf{x} \notin \mathcal{P}$ there is some $\delta > 0$ such that $\{\mathbf{y} : ||\mathbf{x} - \mathbf{y}||_2 \leq \delta\} \cap \mathcal{P} = \emptyset$.*

That is, if $\mathbf{x} \notin \mathcal{P}$ then it is contained in some positive-radius ball that lies outside of $\mathcal{P}$.

We invoke the following classic result.

Theorem 2 *Suppose $\mathcal{P} \subseteq \mathbb{R}$ is closed and bounded (i.e. there is some $\Delta \geq 0$ such that $\mathcal{P}$ is contained in some cube with side length Δ). Then for every sequence $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \ldots$ of points in $\mathcal{P}$ there is some infinite subsequence $\mathbf{x}^{i_1}, \mathbf{x}^{i_2}, \ldots$ with $i_1 < i_2 < \ldots$ that converges to a point $\mathbf{x}^* \in \mathcal{P}$.*

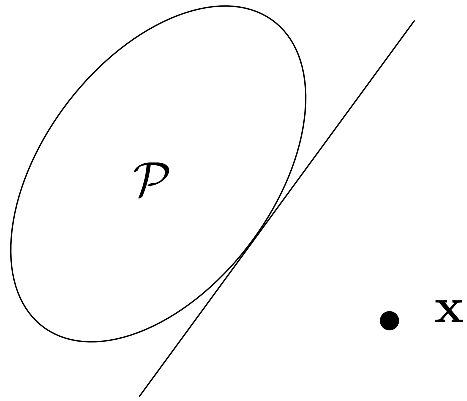

Figure 17.1: A hyperplane separating $\mathbf{x}$ from $\mathcal{P}$. All points $\mathbf{y}$ "above" the line satisfy $\mathbf{a}^T \cdot \mathbf{y} \geq v$.

Sets $\mathcal{P}$ that are closed and bounded are sometimes called **compact**. We did not discuss issues like this explicitly in the lecture, we simply stated that step 1 in the proof of Theorem 4 below follows from the assumption that $\mathcal{P}$ is closed.

However, for those that are curious to see the entire proof of strong duality from start to finish, the full proof is included the Appendix A. Feel free to skip it if you want to just take it for granted.

With this machinery at hand, we now prove the following fundamental result from convex geometry.

Theorem 3 (Separation Theorem) *Let $\mathcal{P} \subseteq \mathbb{R}$ be a closed and convex set. If $\mathbf{x} \notin \mathcal{P}$ then there is some $\mathbf{a} \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}$ and $v \in \mathbb{R}$ such that $\mathbf{a}^T \cdot \mathbf{x} < v$ and $\mathbf{a}^T \cdot \mathbf{y} \geq v$ for all $\mathbf{y} \in \mathcal{P}$.*

See Figure 17.1

Proof.

We assume $\mathcal{P} \neq \emptyset$, otherwise it is trivial: pick any $\mathbf{a} \neq \mathbf{0}$ and set $v = \mathbf{a}^T \cdot \mathbf{x} - 1$.

Step 1: find a point $y^* \in \mathcal{P}$ closest to $\mathbf{x}$

Let $\alpha = \inf_{\mathbf{y} \in \mathcal{P}} ||\mathbf{x} - \mathbf{y}||_2^2$. This is well-defined as $\mathcal{P} \neq \emptyset$ (so there is at least one point) and all distances are nonnegative.

By definition, there is a sequence of points $\mathbf{y}^1, \mathbf{y}^2, \ldots$ in $\mathcal{P}$ such that the values $||\mathbf{x} - \mathbf{y}^i||_2^2$ approach α from above. From Theorem 3, we may also assume that the points $\mathbf{y}^1, \mathbf{y}^2, \ldots$ themselves are converging to a point $\mathbf{y}^* \in \mathcal{P}$ (by restricting to a subsequence if necessary).

Because $||\mathbf{x} - \mathbf{y}||_2^2$ is a continuous function, $||\mathbf{x} - \mathbf{y}^*||_2^2 = \alpha$. So $\mathbf{y}^*$ is a closest point of $\mathcal{P}$ to $\mathbf{x}$. Also note $\alpha > 0$ because $\mathbf{x} \notin \mathcal{P}$.

Step 2: construct the vector $\mathbf{a}$ and value v

Simply put, let $\mathbf{a} = \mathbf{y}^* - \mathbf{x}$ and $v = \mathbf{a}^T \cdot \mathbf{y}^*$. Note

$$\mathbf{a}^T \cdot \mathbf{x} = \mathbf{a}^T \cdot (\mathbf{y}^* + \mathbf{x} - \mathbf{y}^*) = \mathbf{a}^T \cdot \mathbf{y}^* - \mathbf{a}^T \cdot (\mathbf{y}^* - \mathbf{x}) = v - ||\mathbf{x} - \mathbf{y}^*||_2^2 = v - \alpha < v.$$

Step 3: showing $\mathbf{a}^T \cdot \mathcal{P} \geq v$

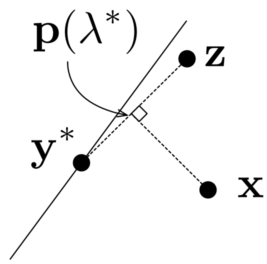

Figure 17.2: The solid line represents the halfspace delimited by $\mathbf{a}, v$. The point $\mathbf{p}(\lambda^*)$ on the line between $\mathbf{y}^*$ and $\mathbf{z}$. It is the point on this line that is closest to $\mathbf{x}$ and it is in $\mathcal{P}$ by convexity.

Suppose, for the sake of contradiction, there is some $\mathbf{z} \in \mathcal{P}$ such that $\mathbf{a}^T \cdot \mathbf{z} < v$. Let $\mathbf{p}(\lambda) = \lambda \mathbf{z} + (1-\lambda)\mathbf{y}^*$ be a parameterization of the line passing through $\mathbf{y}^*, \mathbf{z}$. Let λ^* be such that $(\mathbf{p}(\lambda^*) - \mathbf{x})^T \cdot (\mathbf{z} - \mathbf{y}) = 0$.

More precisely, let

$$\lambda^* = \frac{(\mathbf{x} - \mathbf{y}^*)^T(\mathbf{z} - \mathbf{y}^*)}{||\mathbf{z} - \mathbf{y}||_2^2} = \frac{\mathbf{a}^T \cdot (\mathbf{y}^* - \mathbf{z})}{||\mathbf{z} - \mathbf{y}||_2^2}.$$

One can check that $\mathbf{p}(\lambda^*)$ is the closest point to $\mathbf{x}$ among all points of the form $\mathbf{p}(\lambda)$ by seeing that the derivative of the quadratic function $f(\lambda) = ||\mathbf{p}(\lambda) - \mathbf{x}||_2^2$ vanishes at λ^*.

Note $\lambda^* > 0$ as $\mathbf{a}^T \cdot \mathbf{y}^* = v > \mathbf{a}^T \cdot \mathbf{z}$. If $\lambda^* \geq 1$, then we have $||\mathbf{p}(1) - \mathbf{x}||_2^2 < ||\mathbf{p}(0) - \mathbf{x}||_x^2$ as the quadratic $||\mathbf{p}(\lambda) - \mathbf{x}||_2^2$ is strictly decreasing in the interval $[-\infty, \lambda^*]$. That is, $\mathbf{z}$ would be closer to $\mathbf{x}$ than $\mathbf{y}^*$ contradicting our choice of $\mathbf{y}^*$.

So $0 < \lambda < 1$. But then $\mathbf{p}(\lambda^*)$ is itself a point in $\mathcal{P}$ (by convexity) and would then be strictly closer to $\mathbf{x}$ than $\mathbf{y}^*$. See Figure 17.2 for an illustration of this step.

Therefore for each $\mathbf{z} \in \mathcal{P}$ we have $\mathbf{a}^T \cdot \mathbf{z} \geq v$.

■

A Proof of Theorem 3

Proof. Suppose otherwise. Then for every $\mathbf{y} \in \mathcal{P}$ there is some $\delta_{\mathbf{y}} > 0$ such that the open ball $B(\mathbf{y}; \delta_{\mathbf{y}}) = \{\mathbf{y}' \in \mathbb{R}^n : ||\mathbf{y} - \mathbf{y}'|| < \delta_{\mathbf{y}}\}$ contains only finitely many terms from the sequence $\mathbf{x}^1, \mathbf{x}^2, \ldots$.

As $\mathbf{y} \in B(\mathbf{y}; \delta_{\mathbf{y}})$ for each $\mathbf{y} \in \mathcal{P}$, we have $\mathcal{P} \subseteq \cup_{\mathbf{y} \in \mathcal{P}} B(\mathbf{y}; \delta_{\mathbf{y}})$. We say$\{B(\mathbf{y}; \delta_{\mathbf{y}})\}_{\mathbf{y} \in \mathcal{P}}$ is an **open cover** of $\mathcal{P}$.

Claim: There is a finite subset $\mathcal{Q} \subseteq \mathcal{P}$ such that $\mathcal{P} \subseteq \cup_{\mathbf{y} \in \mathcal{Q}} B(\mathbf{y}; \delta_{\mathbf{y}})$.

That is, the open cover has a *finite subcover*. If this claim is true, we see a contradiction in that each ball $B(\mathbf{y}; \delta_{\mathbf{y}})$ contains only finitely many terms of the sequence yet this finite subcover is claimed to contain all points in $\mathcal{P}$.

To prove the claim, we suppose otherwise and arrive at a contradiction.

For $\mathbf{z} \in \mathbb{R}^n$ and $\kappa \geq 0$ let

$$C(\mathbf{z}, \kappa) = \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y}_i - \mathbf{z}_i| \leq \kappa, \ \forall 1 \leq i \leq n\}$$

be the cube with centre $\mathbf{z}$ with side lengths 2κ. So $\mathcal{P} \subseteq C(\mathbf{0}, \Delta)$ by the assumption that $\mathcal{P}$ is bounded. We construct a sequence of cubes $C_0 = C(\mathbf{0}, \Delta), C_1, C_2, \ldots$ where the side length of C_i is $\Delta/2^i$ inductively as follows.

We know, by assumpion, $C_0 \cap \mathcal{P} = \mathcal{P}$ cannot be covered by a finite subsequence of balls $B(\mathbf{y}, \delta_{\mathbf{y}})$ in the open cover. Inductively, we have that $C_i \cap \mathcal{P}$ cannot be covered by a finite subset of balls $B(\mathbf{y}; \delta_{\mathbf{y}})$ of points. Cut C_i into 2^n cubes with side length $\Delta/2^{i+1}$ by cutting C_i through the middle along each axis. At least one of these 2^n cubes cannot have all points in common with $\mathcal{P}$ being covered by a finite subset of balls $B(\mathbf{y}; \delta_{\mathbf{y}})$. Call this cube C_{i+1}.

This sequence of cubes $C_0, C_1, \ldots$ satisfies $C_i \subseteq C_{i-1}$ and the side length decreases by a factor of 2 in each step, so they converge around a point $\mathbf{y}^*$. in all C_i.

- If $y^* \notin \mathcal{P}$, then some positive-radius ball B around y^* is disjoint from $\mathcal{P}$ (as $\mathcal{P}$ is closed). But $C_i \subseteq B$ for large enough i, contradicting the fact that $C_i \cap \mathcal{P} = \emptyset$ cannot be covered by a finite set of balls $B(\mathbf{y}; \delta_{\mathbf{y}})$.
- If $y^* \in \mathcal{P}$, then for large enough i we have that $C_i \subseteq B(\mathbf{y}; \delta_{\mathbf{y}})$ which contradicts the fact that $C_i \cap \mathcal{P}$ cannot be covered by a finite set of balls.

■