# Lecture 3: Linear Programming Duality

Lecturer: Last updated: September 13, 2022

LP Duality is an extremely useful tool for analyzing structural properties of linear programs. While there are indeed applications of LP duality to directly design algorithms, it is often more useful to gain structural insight (such as approximation guarantees, etc.).

In this lecture, we'll see statements of LP duality. We'll practice applying it in the homeworks.

**Make sure to be clear about what is a constraint (i.e. are $x_i \geq 0$ constraints with an index?), especially in the proofs for strong duality.**

## 1 Weak LP Duality

Consider a linear program of the form:

$$\begin{aligned} \max \sum_i c_i x_i \\ \sum_i A_{ji} x_i \leq b_j, \forall j \\ x_i \geq 0, \forall i. \end{aligned}$$

We'll call this the *primal* LP. $\vec{x}$ is called a *primal* solution, and our goal is to find a primal solution that maximizes our objective, subject to the feasibility constraints. On the other hand, instead of thinking about directly searching for good primal solutions, we could alternatively think about searching for good upper bounds on how good a primal can possibly be. This is called the *dual* problem: How can we derive an upper bound on how good a primal can possibly be?

Consider the following: if we have weights $w_j \geq 0$ for each inequality $j$, and take a linear combination of the feasibility constraints, we may directly conclude that any feasible $\vec{x}$ must satisfy:

$$\sum_i \left( \sum_j w_j \cdot A_{ji} \right) x_i \leq \sum_j w_j \cdot b_j.$$

Okay, so we can upper bound some linear function of any feasible $\vec{x}$, so what? Well, if we happen to have chosen our $w_j$s so that $\sum_j w_j A_{ji} = c_i$ for all $i$, now we're in business! We'll have directly shown that $\sum_i c_i x_i \leq \sum_j w_j \cdot b_j$. In fact, because $x_i \geq 0$, even if we only have $\sum_j w_j A_{ji} \geq c_i$ we're in business, as we'd have:

$$\sum_i c_i x_i \leq \sum_i \left( \sum_j w_j A_{ji} \right) \cdot x_i \leq \sum_j w_j \cdot b_j.$$

Note that the first inequality is only true because $x_i \geq 0$. So now we can think of the following "dual" approach: search over all weights $w_j$ to find the ones that induce the best upper bound. Note that our search is constrained to find weights such that $c_i \leq \sum_j w_j A_{ji}$, so this itself is a linear program:

$$\min \sum_j w_j \cdot b_j$$
$$\sum_j w_j \cdot A_{ji} \geq c_i, \forall i$$
$$w_j \geq 0, \forall j.$$

This is called the dual LP. As an exercise, verify that the dual of the dual LP is itself the primal. Note that we have already proved that *every* feasible solution of the dual provides an upper bound on how good any primal solution can possibly be. Therefore, we have established what is called weak LP duality:

**Theorem 1** (Weak LP Duality)**.** *Let LP1 be any maximization LP and LP2 be its dual (a minimization LP). Then if:*

- *The optimum of LP1 is unbounded ($+\infty$), then the feasible region of LP2 is empty.*
- *The optimum of LP1 finite, it is less than or equal to the optimum of LP2, or the feasible region of LP2 is empty.*

*Proof.* We have already proven the second bullet. To see the first bullet, observe that if the feasible region of LP2 is non-empty, then we have directly found a finite upper bound on LP1. So if LP1 is unbounded, LP2 must be empty. $\square$

In fact, we will see a stronger claim later. Weak Duality is easy to prove, and it's good to remember this intuition. Strong Duality (later) is good to know, but the intuition is largely captured by the proof of Weak Duality.

### 1.1 Complementary Slackness

We'll also want to discuss properties of optimal primal/dual pairs. One useful property is called *complementary slackness.* A $\vec{x}$ and $\vec{w}$ are said to satisfy complementary slackness if they satisfy condition 1) in the theorem statement below.

**Theorem 2.** *Consider a primal LP, LP1 and its dual LP, LP2, and feasible (not necessarily optimal) solutions $\vec{x}$ for the primal and $\vec{w}$ for the dual. Then the following are equivalent:*

1. *($w_j = 0$ OR $\sum_i A_{ji}x_i = b_j$ for all $j$) AND $\Big(x_i = 0$ OR $\sum_j A_{ji}w_j = c_i$ for all $i\Big)$.*
2. *$\sum_i c_i x_i = \sum_j w_j b_j$ (and therefore both $\vec{x}$ is an optimal primal and $\vec{w}$ is an optimal dual).*

*Proof.* Note that we can write:

$$\sum_i c_i \cdot x_i - \sum_j w_j b_j \leq \sum_i (\sum_j A_{ij} w_j) \cdot x_i - \sum_j w_j b_j = \sum_j w_j \cdot \left( \sum_i A_{ij} x_i - b_j \right).$$

The inequality is because $\vec{w}$ is a feasible solution to LP2. The equality is just rearranging the order of sums. Let's now analyze the RHS. Observe that $\sum_i A_{ij} x_i - b_j \leq 0$ for all $j$ as $\vec{x}$ is feasible for LP1. Observe also that $w_j \geq 0$ for all $j$, as $\vec{w}$ is feasible for LP2. So every term in the summand multiplies a non-negative number by a non-positive number and is therefore non-positive. **This means that the RHS is zero if and only if for all $j$, $w_j = 0$ or $\sum_i A_{ij} x_i - b_j = 0$.**

Now we turn our attention to the inequality. Note that because $c_i \leq \sum_j A_{ij} w_j$ for all $i$, the inequality is strict if and only if there exists an $i$ for which $x_i > 0$ and $c_i < \sum_j A_{ij} w_j$. **So the LHS is equal to the middle term if and only if for all $i$, $x_i = 0$ or $c_i = \sum_j A_{ij} w_j$.**

Taking the two bold-font claims together, this means that the LHS is equal to zero if and only if 1) holds. If 1) does not hold, then either the RHS is $< 0$, or the LHS is less than the middle term (which is $\leq 0$). Finally, observe that 2) holds if and only if the LHS above is equal to zero. $\square$

## 2 Weak "Partial Duality"

We'll discuss a slightly more general duality (it's not obvious that the previous duality is a special case of this, but it's a good exercise to show so). We'll again only prove the weak case for now.

**Definition 1.** *Consider an LP of the form:*

$$\begin{aligned} \max \sum_i c_i x_i \\ \sum_i A_{ji} x_i \leq b_j, \forall j \\ x_i \geq 0, \forall i. \end{aligned}$$

*Then a* Lagrangian relaxation *of the above LP for a subset $S$ of constraints and Lagrangian multiplies $\lambda_j \geq 0$ for all $j \in S$ is the following (which we'll refer to as $LP_S^\lambda$:*

$$\begin{aligned} \max \sum_i c_i x_i + \sum_{j \in S} \lambda_j \left( b_j - \sum_i A_{ji} x_i \right) \\ \sum_i A_{ji} x_i \leq b_j, \forall j \notin S \\ x_i \geq 0, \forall i. \end{aligned}$$

**Theorem 3** (Weak "Partial Duality"). *For all $S, \vec{\lambda}$, and any LP, the value of $LP_S^\lambda$ upper bounds the value of LP.*

*Proof.* Let $\vec{x}^*$ optimize LP. Then because $\vec{x}^*$ is feasible for LP, it is also feasible for $LP_S^\lambda$ (as the feasibility constraints in $LP_S^\lambda$ are a proper subset of those in LP). Also, because $\vec{x}^*$ is feasible for LP, we have $b_j - \sum_i A_{ji}x_i^* \geq 0$ for all $j$. As we also have $\lambda_j \geq 0$, this means that $\sum_{j\in S} \lambda_j \left(b_j - \sum_i A_{ji}x_i^*\right) \geq 0$. This directly implies that $\vec{x}^*$ is feasible for $LP_S^\lambda$, and also that $\vec{x}^*$ achieves a greater objective value when evaluated by $LP_S^\lambda$ than LP. $\square$

So every setting of $\vec{\lambda}$ again induces an upper bound on how good the solution to LP can possibly be. We can also think about searching for the best bound of this form (for a fixed $S$). We'll again call $\vec{\lambda}$ a candidate dual solution since it helps witness an upper bound on how good a primal solution can be. The problem below can be written as an LP in terms of the variables $\lambda_i$ (by introducing a variable $t$ constrained so that $t \geq \sum_i c_i x_i + \sum_{j\in S} \lambda_j \left(b_j - \sum_i A_{ji}x_i\right)$ and minimizing $t$). We'll refer to the following program as the partial Lagrangian w.r.t. $S$.

$$\min_{\{\lambda_i \geq 0, i\in S\}} \{\max \sum_i c_i x_i + \sum_{j\in S} \lambda_j \left(b_j - \sum_i A_{ji}x_i\right)$$
$$\sum_i A_{ji}x_i \leq b_j, \forall j \notin S$$
$$x_i \geq 0, \forall i\}.$$

### 2.1 Complementary Slackness

There's a similar definition of Complementary Slackness for this notion of duality. Property 1) below captures this definition.

**Theorem 4** (Complementary Slackness for Partial Lagrangian). *Let LP1 be a linear program and LP2 its Partial Lagrangian w.r.t. $S$. Let $\vec{x}$ be a candidate primal solution to LP1, and $\vec{\lambda}$ a candidate dual solution LP2. Then the following are equivalent:*

1. *For all $j \in S$, $\lambda_j = 0$ OR $A_{ji}x_i = b_j$, AND $\vec{x} = \arg\max_{\vec{x} \mid \sum_i A_{ji}x_i \leq b_j \forall j \notin S, x_i \geq 0 \forall i} \{\sum_i c_i x_i + \sum_{j\in S} \lambda_j \left(b_j - \sum_i A_{ji}x_i\right)\}$.*
2. *$\sum_i c_i x_i = \max_{\vec{x} \mid \sum_i A_{ji}x_i \leq b_j \forall j \notin S, x_i \geq 0 \forall i} \{\sum_i c_i x_i + \sum_{j\in S} \lambda_j \left(b_j - \sum_i A_{ji}x_i\right)\}$ (and therefore, $\vec{x}$ is optimal for LP1, and $\vec{\lambda}$ is optimal for LP2).*

Proof omitted, but similar to that in Section 1.1.

## 3 Strong Duality

(Proof adapted from Anupam Gupta's scribed lecture notes here:
https://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f11/www/notes/lecture05.pdf).

The previous sections discussed weak duality: using dual solutions as upper bounds on how good a primal solution could be. In fact, something quite strong is true: there is always a dual witnessing that the optimal primal is optimal. We'll give a proof, but note that most of the intuition (aside from geometry/linear algebra) is provided by Weak Duality. We'll just discuss the "classic" case, the "partial" case is similar and omitted.

**Theorem 5** (Strong LP Duality)**.** *Let LP1 be any maximization LP and LP2 be its dual (a minimization LP). Then:*

- *If the optimum of LP1 is unbounded ($+\infty$), the feasible region of LP2 is empty.*
- *If the feasible region of LP1 is empty, the optimum of LP2 is either unbounded ($-\infty$), or also infeasible.*
- *If optimum of LP1 finite, then the optimum of LP2 is also finite, and they are equal.*

The key ingredient in the proof will be what's called the Separating Hyperplane Theorem.

**Theorem 6** (Separating Hyperplane Theorem)**.** *Let $P$ be a closed, convex region in $\mathbb{R}^n$, and $\vec{x}$ be a point not in $P$. Then there exists a $\vec{w}$ such that $\vec{x}\cdot\vec{w} > \max_{\vec{y}\in P}\{\vec{y}\cdot\vec{w}\}$.*

*Proof.* Consider the point $\vec{y}\in P$ closest to $\vec{x}$ (that is, minimizing $||\vec{x}-\vec{y}||_2$ over all $\vec{y}\in P$. As distance is a positive continuous function, and $P$ is a closed region, such a $\vec{y}$ exists. Now consider the vector $\vec{w}=\vec{x}-\vec{y}$. We claim that the chosen $\vec{w}$ is the desired witness.

Observe first that $(\vec{x}-\vec{y})\cdot\vec{w} = ||\vec{w}||_2^2 > 0$, so indeed $\vec{x}\cdot\vec{w} > \vec{y}\cdot\vec{w}$. We just need to confirm that $\vec{y} = \arg\max_{\vec{z}\in P}\{\vec{z}\cdot\vec{w}\}$ and then we're done.

Assume for contradiction that $\vec{z}\cdot\vec{w} > \vec{y}\cdot\vec{w}$ and $\vec{z}\in P$. Then as $P$ is convex, $\vec{z}_\varepsilon = (1-\varepsilon)\vec{y}+\varepsilon\vec{z}\in P$ as well for all $\varepsilon > 0$. Observe that $||\vec{x}-\vec{z}_\varepsilon||_2^2 = ||\vec{x}-\vec{y}+\varepsilon(\vec{y}-\vec{z})||_2^2 = ||\vec{x}-\vec{y}||_2^2 - 2\varepsilon(\vec{x}-\vec{y})\cdot(\vec{y}-\vec{z}) + \varepsilon^2||\vec{y}-\vec{z}||_2^2 = ||\vec{x}-\vec{y}||_2^2 - 2\varepsilon(\vec{w})\cdot(\vec{y}-\vec{z}) + \varepsilon^2||\vec{y}-\vec{z}||_2^2$. By hypothesis, $\vec{w}\cdot(\vec{y}-\vec{z}) < 0$, and $||\vec{y}-\vec{z}||_2^2$ is finite, so for sufficiently small $\varepsilon$, we get $||\vec{x}-\vec{z}_\varepsilon||_2^2 < ||\vec{x}-\vec{y}||_2^2$, a contradiction. □

Now, consider the optimum $\vec{x}$ of LP1. Let $S$ denote the $j$ for which $\sum_i A_{ji}x_i = b_j$, and $\bar{S}$ the constraints for which $\sum_i A_{ji}x_i < b_j$. We claim that $\vec{c}$ can be written as a convex combination of the vectors $\vec{A}_j$, $j\in S$ (up to possible scaling).

**Lemma 7.** *Let $\vec{x}$ be the optimum of LP1, and let $S$ denote the $j$ for which $\sum_i A_{ji}x_i = b_j$. Then there exist $\{\lambda_j \geq 0\}_{j\in S}$ such that $c_i = \sum_{j\in S}\lambda_j A_{ji}$ for all $i$.*

*Proof.* Assume for contradiction that this were not the case. Let $X$ denote the space of all vectors $\vec{y}$ for which there exists $\{\lambda_j\geq 0\}_{j\in S}$ such that $y_i = \sum_{j\in S}\lambda_j A_{ji}$ for all $i$. Observe that $X$ is clearly closed and convex, so we can apply the separating hyperplane theorem. Therefore, if $\vec{c}\notin X$ (which we have assumed for contradiction), there exists some $\vec{\gamma}$ such that $\vec{c}\cdot\vec{\gamma} > \max_{\vec{y}\in X}\{\vec{y}\cdot\vec{\gamma}\}$.

Now, we will consider the vector $\vec{x}+\varepsilon\vec{\gamma}$ for sufficiently small $\varepsilon$, and argue that it is a strictly better solution to LP1, contradicting that $\vec{x}$ is optimal.

We first claim that for all $j\in S$, $\sum_i A_{ji}\gamma_i \leq 0$. Assume for contradiction that this is not the case for some $j$. Then, observe that setting $\lambda_j = +\infty$ and all other $\lambda_{j'} = 0$ results in a $\vec{\lambda} \in X$ such that $\vec{\lambda} \cdot \vec{\gamma} = +\infty$. In particular, this implies that $\max_{\vec{y} \in X}\{\vec{y} \cdot \vec{\gamma}\} = +\infty$, contradicting that $\vec{c} \cdot \vec{\gamma} > \max_{\vec{y} \in X}\{\vec{y} \cdot \vec{\gamma}\}$. So we conclude that for all $j \in S$, $\sum_i A_{ji}(x_i + \varepsilon \gamma_i) \leq b_j$.

Moreover, for all $i \notin S$, $\sum_i A_{ji} x_i < b_j$, and $\sum_i A_{ji} \gamma_i$ is finite. Therefore, there exists a sufficiently small $\varepsilon$ so that $\vec{x} + \varepsilon \vec{\gamma}$ is feasible for LP1.

Finally, observe that $\max_{\vec{y} \in X}\{\vec{y} \cdot \vec{\gamma}\} \geq 0$, as $\vec{0} \in X$. So $\vec{c} \cdot \vec{\gamma} > 0$, and we have now found a solution $\vec{x} + \varepsilon \cdot \vec{\gamma}$ such that: (a) for all $j$, $\sum_i A_{ji} \cdot (\vec{x} + \varepsilon \cdot \vec{\gamma}) \leq b_j$, and (b) $\vec{c} \cdot (\vec{x} + \varepsilon \cdot \vec{\gamma}) = \vec{c} \cdot \vec{x} + \varepsilon \vec{c} \cdot \vec{\gamma} > \vec{c} \cdot \vec{x}$. Therefore, we have found a strictly better feasible solution to LP1, contradicting that $\vec{x}$ was optimal. $\square$

Now with the lemma in hand, we want to show a dual whose value matches $\vec{c} \cdot \vec{x}$. Let $\vec{c} = \sum_{j \in S} \lambda_j \vec{A}_j$ with $\lambda_j \geq 0$ as guaranteed by the lemma. Set $w_j = \lambda_j$ for all $j \in S$, and $w_j = 0$ for all $j \notin S$. First, is it clear that $\vec{w}$ is feasible for LP2, as we have explicitly set $w_j$ so that $c_i = \sum_j w_j A_{ij}$ for all $i$. Now we just need to evaluate its value:

$$\sum_j b_j w_j = \sum_{j \in S} b_j w_j + \sum_{j \notin S} b_j \cdot 0 = \sum_{j \in S} (\sum_i A_{ji} x_i) w_j = \sum_i \left( \sum_{j \in S} A_{ji} w_j \right) x_i = \sum_i c_i x_i.$$

So its objective value is exactly the same as LP1.