PRINCETON COS 521: ADVANCED ALGORITHM DESIGN

Linear Programming Duality

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.

Weak LP Duality

Standard form

A generic linear program involves variables $x_1, \ldots, x_n$, and objective of form

$$\max / \min \sum_{i=1}^{n} c_i \cdot x_i$$

and constraints for $j = 1, \ldots, m$

$$\sum_{i=1}^{n} A_{ji} x_i \leq (or \geq or =) b_j.$$

It turns out that any LP has an equivalent LP in the *standard form*:

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

This can be done by observing that $\min \sum_i c_i x_i$ is equivalent to $\max \sum_i (-c_i x_i)$, $\sum_i A_{ji} x_i \geq b_j$ is equivalent to $\sum_i -A_{ji} x_i \leq -b_j$, and $\sum_i A_{ji} x_i = b_j$ can be written as two constraints with LHS $\leq b_j$ and LHS $\geq b_j$. Finally, by replacing an $x_i \in \mathbb{R}$ by $(x'_i - x''_i)$ for $x'_i, x''_i \geq 0$, we enforce all variables to be nonnegative.

Dual LP

Fix an LP, we'll call this the *primal* LP. $\vec{x}$ is called a *primal* solution. Say it is a maximization LP, then 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_js 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:

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

This is called the dual LP.

For LPs of general form, we can also use this approach to bound its optimal value, which gives its dual LP. For example, if a variable x_i is an arbitrary real, then we must require $\sum_j w_j \cdot A_{ji} = c_i$. If a constraint is equality $\sum_i A_{ji} x_i = b_i$, then its weight w_i can be an arbitrary real.

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. □

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.

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 section 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 can obtain the tight upper bound in this way.

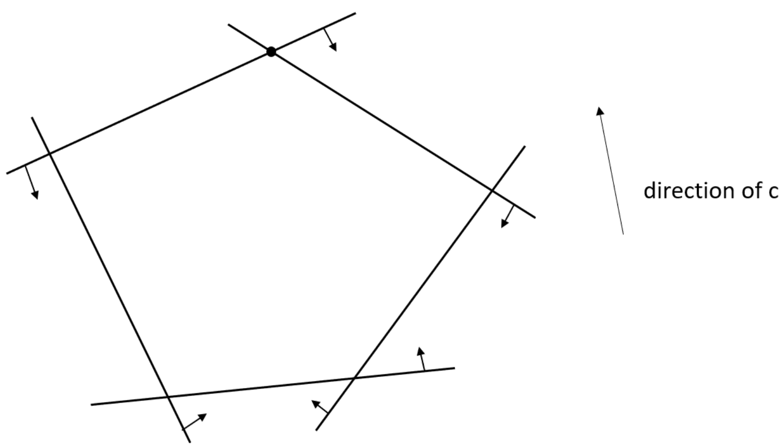
Theorem 2 (Strong LP Duality)**.** *Let LP1 be any maximization LP and LP2 be its dual (a minimization LP). Then:*

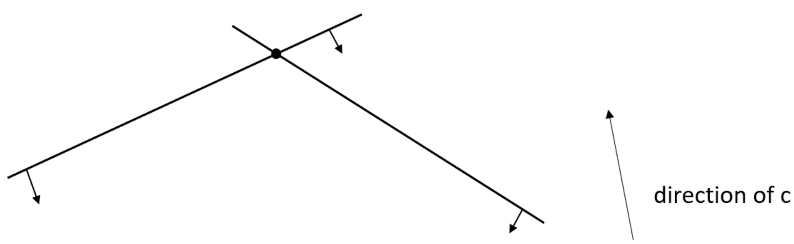
- *If the optimum of LP1 is unbounded ($+\infty$), the feasible region of LP2 is empty.*
- *If the feasible region of LP1 is empty, the optimum of LP2 is either unbounded ($-\infty$), or also infeasible.*
- *If optimum of LP1 finite, then the optimum of LP2 is also finite, and they are equal.*

The first bullet is covered in the weak duality. The second bullet holds, since if LP2 has finite optimum, then by the third bullet, and the fact that the dual of LP2 is LP1, LP1 would also have finite optimum, we get a contradiction. It remains to prove the last bullet.

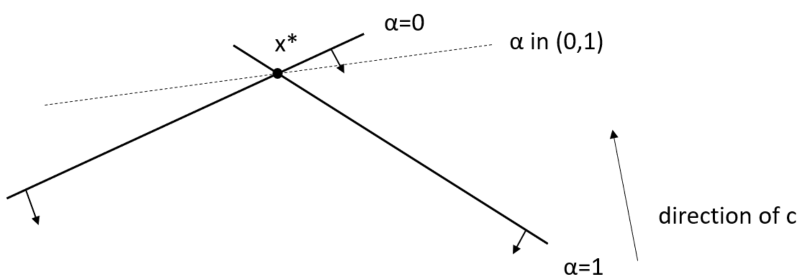
Intuition

To understand intuitively why the two LPs have the same optimum, we will take the geometric view of LP. For simplicity, let us imagine an LP in 2D (figure below). We try to maximize in the direction of c in the polygon (the objective is maximizing $\langle c, x \rangle$). First observe that the extreme point must be achieved at a vertex. Intuitively, this is because otherwise, we can always "slide" towards left or right, and one direction would allow us to go further in the direction of c (the only exception is when an edge is perpendicular to c, then the entire edge achieves OPT, in this case, one endpoint would also do). Suppose the optimal point is x^*. Now, we notice that for this particular LP and its objective, all constraints that do not go through x^* are not important in terms of proving strong duality. This is because, if we remove all those constraints, it does not affect x^* being the optimal solution (see figure on the right). Thus, if strong duality holds, then it holds for this new LP with fewer constraints – there is a nonnegative linear combination of the constraints that proves the optimality of x^*. In the other words, it is safe to ignore the other constraints for now.





Now, let us look at the two constraints in the figure that go through x^*. What does it mean to take a nonnegative linear combination of the two. When we give weights $w_1, w_2 \geq 0$ to them, and obtain a new inequality, first observe that if we multiply w_1, w_2 by the same (positive) constant, we would get the same new inequality. Thus, without loss of generality, let us assume that $w_1 = \alpha$ and $w_2 = 1 - \alpha$ for some $\alpha \in [0, 1]$. Now, when $\alpha = 1$, it is exactly the first constraint, when $\alpha = 0$, it is exactly the second constraint. When α goes from 0 to 1, the slope gradually changes, and the new inequality always goes through x^* (this is because x^* achieves equality in both constraints, and by taking a linear combination, equality should still be achieved). That is, the new inequality (constraint) rotates around x^*. Suppose we can find an $\alpha \in (0, 1)$ in the middle, such that this new constraint is perpendicular to c. Then this would be a proof of the optimality of x^* – every point on this constraint is as far in the direction of c as each other, and we can only go this far. We will need to use the optimality of x^* to show the existence of such an α (it is not too difficult to see in 2D, since the direction of c must be between the two constraints, otherwise we could slide along one edge to move further). This intuition holds in high dimensions, and we will give a formal proof below.



Proof

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

Theorem 3 (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 ε, 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 1. *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 ε, 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 ε 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.