15-750: Graduate Algorithms March 6, 2017

Lecture 21: Linear Programming Duality

Lecturer: David Witmer *Scribe(s): Gabriele Farina, Rui Silva*

1 Introduction

We have seen in Lecture 20 that every linear program can be put into *canonical form* as

$$\mathcal{P} \ : \ \begin{cases} \max\limits_{x} & c^\top x \\ \text{s.t.} & Ax \le b \\ & x \ge 0 \end{cases}$$

Today, we will see how to construct the *dual* of $\mathcal{P}$. Let's start with an example.

1.1 Example

Consider problem of Equation 1.

$$\begin{cases} \max\limits_{x_1,x_2} & & 2x_1 + 3x_2 \\ \text{s.t.} & ① & 4x_1 + 8x_2 \le 12 \\ & ② & 2x_1 + \ \ x_2 \le 3 \\ & ③ & 3x_1 + 2x_2 \le 4 \\ & ④ & x_1, x_2 \ge 0 \end{cases} \tag{1}$$

Suppose that we also suggested that the optimal solution might be attained at $(x_1, x_2) = (1/2, 5/4)$, achieving an objective value equal to $2x_1 + 3x_2 = 19/4$. We now construct a certificate for the optimality of the hint.

The problem is asking us to maximize $2x_1 + 3x_2$ (with x_1, x_2 non-negative), under some constraints. Let's start by considering constraint ① : since $4x_1 + 8x_2 \le 12$, we definitely know that the optimal value of $2x_1 + 3x_2$ is upper bounded by the value of $4x_1 + 8x_2$; hence, 12 is an upper bound on the optimal value of the objective. Now, let's consider ②+③, i.e. the sum of constraints ② and ③ : we have $5x_1 + 3x_2 \le 7$; since definitely $2x_1 + 3x_2 \le 5x_1 + 3x_2$, a better upper bound on the optimal value of the objective is 7. Let's now consider $\frac{1}{2}$①, that is, let's multiply constraint ① by 1/2: we conclude that $2x_1 + 4x_2 \le 6$; again, $2x_1 + 4x_2$ is an upper bound on the objective $2x_1 + 3x_2$ (remember that x_1 and x_2 are non-negative), and we improve our incumbent upper bound to 6. Now, consider instead $\frac{1}{3}$①+ $\frac{1}{3}$②: we get $2x_2 + 3x_2 \le 5$, improving the upper bound to 5. We still don't know if 19/4 is optimal, but we definitely know that whatever the optimal value, it cannot be larger than 5.

Finally, consider $\frac{5}{16}$①+ $\frac{1}{4}$③; we get $2x_1 + 3x_2 \le 19/4$. This proves that an upper bound on the optimal value of the objective is 19/4. This means that the hint was correct, and the optimal value of the problem is indeed 19/4. In other terms, our optimality certificate is the triple $(5/16, 0, 1/4)$, representing the multipliers that we applied to the constraints to obtain an upper bound matching the value we wanted to certify.

1.2 Dual problem

Take a second look at what we did in the example above. By selecting different choices of non-negative[1] constraint multipliers, we were able to upper bound the value of the objective function, finding better and better upper bounds until we found a choice that actually matched the optimal value of the program.

Now, let's try to generalize the process. Assume we are given the canonical-form problem $\mathcal{P}$ as defined above. We are interested in finding a non-negative vector y of constraint multipliers that, when applied to the set of constraints $Ax \le b$ has the property that:

- the left-hand side of the resulting constraint is an upper bound of the objective function $c^\top x$; Notice that since the resulting constraint is $(y^\top A)x = (y^\top b)$, it is enough that $(y^\top A) \ge c^\top$ element-wise. By taking the transpose, an equivalent condition is $A^\top y \ge c$.
- the right-hand side of the resulting constraint (that is, $y^\top b$) is as little as possible.

In other words, we can cast the problem of finding the lowest upper bound as the *linear* problem

$$\mathcal{D} \ : \ \begin{cases} \min\limits_{y} & y^\top b \\ \text{s.t.} & A^\top y \ge c \\ & y \ge 0 \end{cases}$$

The problem $\mathcal{D}$ is called the *dual* problem of $\mathcal{P}$.

1.3 Duality Recipes

While every linear problem can be brought in canonical form, it is practically convenient to remember how to deal with equality constraint and free variables when taking the dual of a problem. Consider the problem on the left in Equation 2: it's not hard to prove that its dual is the problem on the right in Equation 2.

$$\tilde{\mathcal{P}} \ : \ \begin{cases} \max\limits_{x} & \sum_j c_j x_j \\ \text{s.t.} & a_i^\top x \le b_i \quad \forall i \in I_\le \\ & a_i^\top x = b_i \quad \forall i \in I_= \\ & x_j \ge 0 \quad \forall j \in J_+ \\ & x_j \in \mathbb{R} \quad \forall j \in J_0 \end{cases} \qquad \leftrightarrow \qquad \tilde{\mathcal{D}} \ : \ \begin{cases} \min\limits_{y} & \sum_i b_i y_i \\ \text{s.t.} & a_j^\top y \ge c_j \quad \forall j \in J_+ \\ & a_j^\top y = c_j \quad \forall j \in J_0 \\ & y_i \ge 0 \quad \forall i \in I_\le \\ & y_i \in \mathbb{R} \quad \forall i \in I_= \end{cases} \tag{2}$$

2 Weak duality

The previous section introduced the concept of the dual problem of $\mathcal{P}$, the primal optimization problem. This section presents the Weak Duality Theorem in Linear Programming problems, which claims that the optimal solution to the dual problem serves as an upper bound to the optimal solution of the primal problem.

Theorem 2.1 (Weak Duality Theorem)**.** *Assume primal and dual problems $\mathcal{P}$ and $\mathcal{D}$, with feasible solutions x and y. Then, $c^\top x \le b^\top y$.*

[1] Notice that constraint multipliers have to be non-negative, not to alter the sense of the constraint inequality.

Proof. We have

$$c^\top x = x^\top c \leq x^\top (A^\top y) = (Ax)^\top y \leq b^\top y,$$

where the first and second inequalities follow from the feasibility of y and x respectively, and from the fact that x and y are non-negative. □

As a corollary, from this result we can extract further conclusions. In particular, given a primal problem $\mathcal{P}$ and its dual $\mathcal{D}$ one of the following applies:

- Both $\mathcal{P}$ and $\mathcal{D}$ are infeasible;
- $\mathcal{P}$ is unbounded and $\mathcal{D}$ is infeasible;
- $\mathcal{P}$ is infeasible and $\mathcal{D}$ is unbounded;
- Both $\mathcal{P}$ and $\mathcal{D}$ are feasible. In the next section we present an even stronger result that shows that when this is the case, the optimal values are also equal.

Notice that for the first case it suffices to find an example of primal and dual LPs that are both infeasible. The second and third cases follow directly from the Weak Duality Theorem—if $\mathcal{P}$ is unbounded then there is no y such that $b^\top y$ is an upper bound (and *vice versa*).

2.1 A second look via Lagrangian Duality

We now provide a second-look to the Weak Duality Theorem via Lagrangian Duality (Lemma 2.2), following the approach presented in the AM 221 Advanced Optimization Course, taught by Yaron Singer at Harvard. We start with a Lemma.

Lemma 2.2 (Lagrangian Weak Duality)**.** *For every function $f : \mathcal{X} \to \mathcal{Y}$ (including non-convex functions), the following inequality holds*

$$\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} f(x,y) \leq \min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} f(x,y)$$

Proof. Note that $\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} f(x,y) \leq \max_{x\in\mathcal{X}} f(x,\bar{y})$ for all $\bar{y} \in \mathcal{Y}$, including $y^* = \arg\min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} f(x,y)$. Thus, $\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} f(x,y) \leq \min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} f(x,y)$. □

In the case of Linear Programming, $f(x) = c^\top x$ is a linear function, and $\mathcal{X}$ is the polytope $\mathcal{X} = \{x \in \mathbb{R}^n \mid Cx \leq d\}$, for some C and d.[2] This is equivalent to the unconstrained maximization of function $\tilde{f}(x)$ equal to $f(x)$ on $\mathcal{X}$ and taking value $-\infty$ outside of $\mathcal{X}$. Formally,

$$\max_{x\in\mathcal{X}} f(x) = \max_{x\in\mathbb{R}^n} \tilde{f}(x), \text{ for } \tilde{f}(x) = \begin{cases} f(x), & \text{if } x \in \mathcal{X} \\ -\infty, & \text{otherwise.} \end{cases}$$

This result allows us now to formulate the maximization of $f(x)$ as a max-min problem:

[2] Note this definition is generic and also applies to any problem in canonical form. In particular we can merge the constraints $Ax \leq b$ and $x \geq 0$ into the one constraint $Cx \leq d$, where $C = \begin{bmatrix} A \\ -I \end{bmatrix}$, and $d = \begin{bmatrix} b \\ 0 \end{bmatrix}$.

$$\max_{x\in\mathcal{X}} f(x) = \max_{x\in\mathbb{R}^n} \tilde{f}(x) = \max_{x\in\mathbb{R}^n} \min_{\lambda\geq 0} \left\{ c^\top x + \lambda^\top (d - Cx) \right\}$$

Note that the inner minimization takes $\lambda^* = 0$ when $x \in \mathcal{X}$, and λ^* arbitrarily large otherwise.

Therefore, we conclude

$$\begin{aligned}\max_{x\in\mathcal{X}} f(x) = \max_{x\in\mathbb{R}^n} \tilde{f}(x) &= \max_{x\in\mathbb{R}^n} \min_{\lambda\geq 0} \left\{ c^\top x + \lambda^\top (d - Cx) \right\} \\ &\leq \min_{\lambda\geq 0} \max_{x\in\mathbb{R}^n} \left\{ \lambda^\top d + (c^\top - \lambda^\top C)x \right\}\end{aligned}$$

where the first inequality is due to Lemma 2.2, and the second inequality follows from the Lagrangian Weak Duality lemma (Lemma 2.2). Now, notice that

$$\tilde{g}(\lambda) := \max_{x\in\mathbb{R}^n} \left\{ \lambda^\top d + (c^\top - \lambda^\top C)x \right\} = \begin{cases} \lambda^\top d, & \text{if } c^\top = \lambda^\top C \\ +\infty, & \text{otherwise,} \end{cases}$$

so that

$$\min_{\lambda\geq 0} \max_{x\in\mathbb{R}^n} \left\{ \lambda^\top d + (c^\top - \lambda^\top C)x \right\} = \min_{\lambda\geq 0} \tilde{g}(\lambda) = \min_{\lambda\geq 0,\ A^\top\lambda = c} \lambda^\top d.$$

In conclusion, we have proved that

$$\max_{x\in\mathbb{R}^n\ :\ Cx\leq d} c^\top x \leq \min_{\lambda\geq 0,\ A^\top\lambda = c} \lambda^\top d,$$

that is exactly the Weak Duality Theorem for linear programs.

3 Strong duality

In the previous section we proved among other things that if the canonical-form primal problem $\mathcal{P}$ is feasible, so is the dual $\mathcal{D}$, with the value of the latter being an upper bound on the value of the former.

We now extend this result, showing that when the programs are feasible, their (optimal) values coincide.

Theorem 3.1. *If the canonical-form primal problem $\mathcal{P}$ is feasible, its dual $\mathcal{D}$ is also feasible, and the (optimal) values of the two programs are equal.*

Proof (sketch). Let x^* be one optimal vertex for the primal problem $\mathcal{P}$. Let I be the set of row indices for the active constraints at x^*: in other words, $i \in I \iff a_i^\top x^* = b_i$ where $\{a_i\}$ and $\{b_i\}$ represent the rows of A and b, respectively.

We prove that c belongs to the cone generated by the active rows $\{a_i\}_{i\in I}$. This has a strong geometric justification: c determines the direction of maximal growth of the objective function, and if x^* is optimal we expect c to be directed "outside" of the feasibility polytope (see Figure 1).

Lemma 3.2. *The vector c belongs to the convex cone generated by the active rows $\{a_i\}_{i\in I}$*

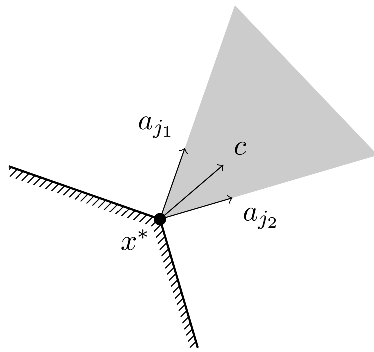

Figure 1: Geometric insight for the fact that $c \in \text{Cone}(\{a_i\}_{i\in I})$ at an optimal vertex. In this case, the active constraints correspond to rows j_1 and j_2.

Proof. Assume by contradiction that $c \notin \text{Cone}(\{a_i\}_{i\in I})$; then we can separate the convex cone and c with a hyperplane of the form $d^\top x = 0$, so that $d^\top a_i \le 0$ for all $i \in I$ and $d^\top c < 0$. We now show that x^* is not optimal, by "sliding" it by a tiny amount ϵ in the direction of d so that the new point remains feasible, and proving that the objective value at this new point $x^* + \epsilon d$ is larger than $c^\top x^*$. First of all, we prove that if ϵ is small enough, $x^* + \epsilon d$ remains feasible: for all row index i, we have $a_i^\top(x^* + \epsilon d) = a_i^\top x^* + \epsilon a_i^\top d$. Now, notice that if $i \in I$, $a_i^\top d \le 0$, so that $a_i^\top(x^* + \epsilon d) \le a_i^\top x^* = b_i$; if, on the other hand, $i \notin I$, then by hypothesis $a_i^\top x^* = b_i - \xi$, with $\xi > 0$ — this means that as long as $\epsilon > 0$ is such that $\epsilon a_i^\top d < \xi$, constraint i is not violated. In summary, there exists $\epsilon > 0$ such that the point $x^* + \epsilon d$ respects all constraints, and is therefore feasible. The objective value at that point is

$$c^\top(x^* + \epsilon d) = c^\top x^* + \epsilon c^\top d > c^\top x^*,$$

showing that x^* was not optimal, a contradiction. □

Let $\tilde{A}$ and $\tilde{b}$ be the portion of A and b corresponding to all active rows (i.e. those rows having index in I). By definition of convex cone, there exists $\tilde{\lambda} \ge 0$ such that $c = \tilde{A}^\top \tilde{\lambda}$. Now, since $\tilde{A}x^* = \tilde{b}$ by definition of active row, we can also write

$$\tilde{\lambda}^\top \tilde{b} = \tilde{\lambda}^\top(\tilde{A}x^*) = c^\top x^*.$$

This shows that that the dual problem $\mathcal{D}$ has a feasible solution (corresponding to the vector y equal to $\tilde{\lambda}$ on rows $\in I$ and 0 elsewhere), such that $y^\top b = c^\top x^*$. Therefore, the value of $\mathcal{D}$, being a minimization problem, is not greater than $c^\top x^*$. However, by the Weak Duality theorem, we know that the value of the dual is not less than $c^\top x^*$. Hence, the values of the primal and the dual coincide, concluding the proof.

□

4 Application: Max Flow – Min Cut duality

We now present an application of duality theory: the celebrated Max Flow – Min Cut theorem. In the previous lecture, we have seen that the Max Flow problem can be formulated as a Linear Program:

$$
\begin{aligned}
\max_{f} \quad & \sum_{u \in V} f(s,u) \\
\text{s.t.} \quad & \sum_{v \in V} f(v,u) = \sum_{v \in V} f(u,v), \quad \forall u \in V \setminus \{s,t\} \\
& f(u,v) \le c(u,v), \quad \forall (u,v) \in E \\
& f(u,v) \ge 0, \quad \forall (u,v) \in E.
\end{aligned}
$$

This time, however, we will formulate the problem in a slightly different yet equivalent way. Let P_{st} be the set of all paths from s to t and x_p the flow assigned to path $p \in P_{st}$, and consider the problem

$$
\mathcal{F} : \begin{cases} \max\limits_{x} & \sum\limits_{p \in P_{st}} x_p \\ \text{s.t.} & \sum\limits_{p \ni e} x_p \le c_e, \ \forall e \in E \\ & x_p \ge 0, \quad \forall p \in P_{st}. \end{cases}
$$

Note that the number of paths from s to t may be large, and as such, there may be exponentially many variables — this formulation is not really useful for finding a max-flow, but has a cleaner combinatorial structure.

The dual of $\mathcal{F}$ is $\mathcal{C}$, as shown below.

$$
\mathcal{C} : \begin{cases} \min\limits_{y} & & \sum\limits_{e \in E} c_e y_e \\ \text{s.t.} & \text{①} & \sum\limits_{e \in p} y_e \ge 1, \ \forall p \in P_{st} \\ & \text{②} & y_e \ge 0, \quad \forall e \in E. \end{cases}
$$

From the duality theory we know that the optimal values of $\mathcal{F}$ and $\mathcal{C}$ coincide. We now prove that the optimal value α of $\mathcal{C}$ is equal to the value β of the min-cut.

($\alpha \le \beta$) We start by proving that every cut $(S, \bar{S})$ of G, where $\bar{S} = V \setminus S$, can be easily transformed into a feasible solution of $\mathcal{C}$, preserving the value of the cut itself. To this end, assign

$$
y_e = \begin{cases} 1 & \text{if } e \text{ belongs to the cut, } \textit{i.e. } e = (u,v) \text{ with } u \in S, v \in \bar{S}, \\ 0 & \text{otherwise.} \end{cases}
$$

Clearly, ② is satisfied, and the value of $\mathcal{C}$ is equal to the value of the cut. Furthermore, being $(S, \bar{S})$ a cut of G, every path from s to t has to contain at least one edge of the cut, making ① trivially true. In particular, any min-cut is feasible for $\mathcal{C}$, proving that $\alpha \le \beta$.

($\beta \le \alpha$) Now, we prove that the existence of a feasible solution of $\mathcal{C}$ of value α implies the existence of a valid s-t cut $(S, \bar{S})$ of value not larger than α. Given a feasible solution for $\mathcal{C}$, we interpret the values of the variables y_e as edge lengths, so that it makes sense to reason about the minimum distance $d(v)$ from s to any vertex $v \in V$. Notice that $d(s) = 0$; furthermore, constraint ① implies that $d(t) \ge 1$. Therefore, given any $\rho \in [0,1)$, the two sets $S_\rho = \{v \in V \mid d(v) \le \rho\}$ and $\bar{S}_\rho = \{v \in V \mid d(v) > \rho\}$ define a valid s-t cut of G, as $s \in S_\rho$ and $t \in \bar{S}_\rho$.

Given $\rho \in [0,1)$, it makes sense to define the notion of *active cut* at "time" ρ, that is the set of edges $e = (u,v)$ with $u \in S_\rho$ and $v \in \bar{S}_\rho$. The key observation now is that given any edge

$e = (u, v)$, the set of time instants for which e is active is pretty small; more precisely, we show that the (Lebesgue) measure[3] μ of the set

$$T_e = \{\rho \in [0, 1) \mid u \in S_\rho,\ v \in \bar{S}_\rho\}$$

is not greater than y_e: $\mu(T_e) \le y_e$. In order to see this, notice that clearly $\rho < d(u) \implies u \notin S_\rho$; furthermore, since $d(v) \le d(u) + y_e$, $\rho \ge d(u) + y_e$ immediately implies that $v \notin \bar{S}_\rho$. In both cases, e would not be active: hence, $T_e \subseteq L_e \doteq [d(u), d(u) + y_e)$, and since $\mu(L_e) = y_e$ we conclude $\mu(T_e) \le y_e$.

With the above observation in mind, we prove that there must exist at least one $\rho \in [0, 1)$ for which the sum of the weight of all the active edges at time ρ is at most $\sum c_e y_e$. This makes intuitive sense: each edge can be active only for a short amount of time (the measure of T_e is upper-bounded), and therefore it is not possible to guarantee that the value of the cut $(S_\rho, \bar{S}_\rho)$ is large at all times.

Consider the function $c : [0, 1) \to \mathbb{R}$ indicating, for each ρ, the sum of the weights of all the active edges at time ρ. We can write

$$c = \sum_{e \in E} c_e \mathbb{I}_{T_e},$$

where $\mathbb{I}_\bullet$ is the indicator function for $\bullet$, *i.e.* the function equal to 1 for all $\rho \in \bullet$ and 0 otherwise. The integral of c is given by

$$\int c\, d\mu = \sum_{e \in E} c_e \int \mathbb{I}_{T_e}\, d\mu = \sum_{e \in E} c_e \mu(T_e) \le \sum_{e \in E} c_e y_e.$$

Therefore, there must exist at least one $\rho^* \in [0, 1)$ for which $c(\rho^*) \le w = \sum c_e y_e$ (or otherwise, the integral of c could not be $\le w$). Therefore, the cut $(S_{\rho^*}, \bar{S}_{\rho^*})$ is a valid s-t cut of G of value w, proving that the min-cut of G is upper bounded by the value of $\mathcal{C}$, and completing the proof.

[3] The measure of a set intuitively represents how "large" it is. The Lebesgue measure is the standard way of assigning a measure to a subset of $\mathbb{R}^n$. Depending on the dimension n of the space (in our case, $n = 1$), it coincides with the intuitive notion of length, area, or volume.