PRINCETON UNIVERSITY F'02 COS 597D: A THEORIST'S TOOLKIT

Lecture 2: The Duality Theorem

Lecturer: *Sanjeev Arora* Scribe: *Satyen Kale*

1 Linear Programming and Farkas' Lemma

A Linear Program involves optimizing a linear cost function with respect to *linear* inequality constraints. They are useful for algorithm design as well as a tool in mathematical proofs. The typical program looks as follows.

GIVEN: vectors $\mathbf{c}, \mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_m \in \mathbf{R}^n$, and real numbers $b_1, b_2, \ldots b_m$.
OBJECTIVE: find $\mathbf{X} \in \mathbf{R}^n$ to minimize $\mathbf{c} \cdot \mathbf{X}$, subject to:

$$\begin{array}{rcl} \mathbf{a}_1 \cdot \mathbf{X} & \geq & b_1 \\ \mathbf{a}_2 \cdot \mathbf{X} & \geq & b_2 \\ & \vdots & \\ \mathbf{a}_m \cdot \mathbf{X} & \geq & b_m \\ \mathbf{X} & \geq & 0 \end{array} \tag{1}$$

The notation $\mathbf{X} > \mathbf{Y}$ simply means that $\mathbf{X}$ is componentwise larger than $\mathbf{Y}$. Now we represent the system in (1) more compactly using matrix notation. Let

$$A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Then the Linear Program (LP for short) can be rewritten as:

$$\begin{array}{rl} \min & \mathbf{c}^T\mathbf{X} : \\ A\mathbf{X} & \geq \mathbf{b} \\ \mathbf{X} & \geq 0 \end{array} \tag{2}$$

This form is general enough to represent any possible linear program. For instance, if the linear program involves a linear equality $\mathbf{a} \cdot \mathbf{X} = b$ then we can replace it by two inequalities

$$\mathbf{a} \cdot \mathbf{X} \geq b \quad \text{and} \quad -\mathbf{a} \cdot \mathbf{X} \geq -b.$$

If the variable X_i is unconstrained, then we can replace each occurence by $X_i^+ - X_i^-$ where X_i^+, X_i^- are two new non-negative variables.

The set of conditions in an LP may not be satisfiable, however. Farkas' Lemma tells us when this happens.

LEMMA 1
Farkas' Lemma. *The set of linear inequalities (1) is infeasible if and only if using positive linear combinations of the inequalities it is possible to derive* $-1 \geq 0$*, i.e. there exist* $\lambda_1, \lambda_2, \ldots \lambda_m \geq 0$ *such that*

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i < 0 \quad \textit{and} \quad \sum_{i=1}^{m} \lambda_i b_i > 0.$$

2 The Duality Theorem

With every LP we can associate another LP called its *dual*. The original LP is called the *primal*. If the primal has n variables and m constraints, then the dual has m variables and n constraints.

$$\begin{array}{rl|rl} \multicolumn{2}{c|}{\underline{\text{Primal}}} & \multicolumn{2}{c}{\underline{\text{Dual}}} \\ \min & \mathbf{c}^T\mathbf{X}: & \max & \mathbf{Y}^T\mathbf{b}: \\ A\mathbf{X} & \geq \mathbf{b} & \mathbf{Y}^T A & \leq \mathbf{c^T} \\ \mathbf{X} & \geq 0 & \mathbf{Y} & \geq 0 \end{array} \tag{3}$$

(Aside: if the primal contains an equality constraint instead of inequality then the corresponding dual variable is unconstrained.)

It is an easy exercise that the dual of the dual is just the primal.

THEOREM 2
The Duality Theorem. *If both the Primal and the Dual of an LP are feasible, then the two optima coincide.*

PROOF: The proof involves two parts:

1. Primal optimum $\geq$ Dual optimum.
   This is the easy part. Suppose $\mathbf{X}^*, \mathbf{Y}^*$ are the respective optima. This implies that

   $$A\mathbf{X}^* \geq \mathbf{b}.$$

   Now, since $\mathbf{Y}^* \geq 0$, the product $\mathbf{Y}^* A\mathbf{X}^*$ is a non-negative linear combination of the rows of $A\mathbf{X}^*$, so the inequality

   $$\mathbf{Y}^{*T} A\mathbf{X}^* \geq \mathbf{Y}^{*T}\mathbf{b}$$

   holds. Again, since $\mathbf{X}^* \geq 0$ and $\mathbf{c}^T \geq \mathbf{Y}^{*T} A$, the inequality

   $$\mathbf{c}^T\mathbf{X}^* \geq (\mathbf{Y}^{*T} A)\mathbf{X}^* \geq \mathbf{Y}^{*T}\mathbf{b}$$

   holds, which completes the proof of this part.

2. Dual optimum $\geq$ Primal optimuml.
   Let k be the optimum value of the primal. Since the primal is a minimization problem, the following set of linear inequalities is infeasible for any $\epsilon > 0$:

   $$\begin{array}{rl} -\mathbf{c}^T\mathbf{X} & \geq -(k-\epsilon) \\ A\mathbf{X} & \geq \mathbf{b} \end{array} \tag{4}$$

Here, ϵ is a small positive quantity. Therefore, by Farkas' Lemma, there exist $\lambda_0, \lambda_1, \ldots \lambda_m \geq 0$ such that

$$-\lambda_0 \mathbf{c} + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i < 0 \tag{5}$$

$$-\lambda_0 (k - \epsilon) + \sum_{i=1}^{m} \lambda_i b_i > 0. \tag{6}$$

Note that $\lambda_0 > 0$ omitting the first inequality in (4) leaves a feasible system by assumption about the primal. Thus, consider the vector

$$\Lambda = (\frac{\lambda_1}{\lambda_0}, \ldots \frac{\lambda_m}{\lambda_0})^T.$$

The inequality (5) implies that $\Lambda^T A \leq \mathbf{c}^T$. So Λ is a feasible solution to the Dual. The inequality (6) implies that $\Lambda^T \mathbf{b} > (k - \epsilon)$, and since the Dual is a maximization problem, this implies that the Dual optimal is bigger than $k - \epsilon$. Letting ϵ go to zero, we get that the Dual optimal $\geq k =$ Primal optimal. Thus, this part is proved, too. Hence the Duality Theorem is proved.

□

Sanjeev's thoughts on this business: (1) Usually textbooks bundle the case of infeasible systems into the statement of the Duality theorem. He feels that this muddies the issue. Usually all applications of LPs fall into two cases: (a) We either know (for trivial reasons) that the system is feasible, and are only interested in the value of the optimum or (b) We do not know if the system is feasible and that is precisely what we want to determine. Then it is best to just use Farkas' Lemma. (2) The proof of the Duality theorem is interesting. The first part shows that for any dual feasible solution $\mathbf{Y}$ the various Y_i's can be used to obtain a *weighted* sum of primal inequalities, and thus obtain a lowerbound on the primal. The second part shows that this method of taking weighted sums of inequalities is *sufficient* to obtain the best possible lowerbound on the primal: there is no need to do anything fancier (e.g., taking products of inequalities or some such thing).

3 Example: Max Flow Min Cut theorem in graphs

The input is a directed graph $G(V, E)$ with one source s and one sink t. Each edge e has a capacity c_e. The flow on any edge must be less than its capacity, and at any node apart from s and t, flow must be conserved: total incoming flow must equal total outgoing flow. We wish to maximize the flow we can send from s to t. The maximum flow problem can be formulated as a Linear Program as follows:

Let $\mathcal{P}$ denote the set of all (directed) paths from s to t. Then the max flow problem

becomes:

$$\max \sum_{P \in \mathcal{P}} f_P : \tag{7}$$

$$\forall P \in \mathcal{P} : f_P \geq 0 \tag{8}$$

$$\forall e \in E : \sum_{P : e \in P} f_P \leq c_e \tag{9}$$

Going over to the dual, we get:

$$\min \sum_{e \in E} c_e y_e : \tag{10}$$

$$\forall e \in E : y_e \geq 0 \tag{11}$$

$$\forall P \in \mathcal{P} : \sum_{e \in P} y_e \geq 1 \tag{12}$$

Notice that the dual in fact represents the Fractional min $s - t$ cut problem: think of each edge e being picked up to a fraction y_e. The constraints say that a total weight of 1 must be picked on each path. Thus the usual min cut problem simply involves $0 - 1$ solutions to the y_e's in the dual.

EXERCISE 1 Prove that the optimum solution does have $y_e \in \{0, 1\}$.

Thus, the max- st-flow = (capacity of) min-cut.

PRINCETON UNIVERSITY F'02 COS 597D: A THEORIST'S TOOLKIT

Lecture 3: Using LP Duality: Approximate Inclusion-Exclusion

Lecturer: *Sanjeev Arora* Scribe: *Satyen Kale*

1 Approximate Inclusion-Exclusion

Today we will see an application of linear programming and the duality theorem. The Inclusion-Exclusion formula for the cardinality of the union of n finite sets $A_1, A_2, \ldots, A_n$ is given by

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| - \cdots + (-1)^{n+1} |A_1 \cap A_2 \cap \ldots \cap A_n| \quad (13)$$

The question is, suppose we know only the first k terms of the formula, how good an approximation can we get? The simple idea of truncating the formula to the first k terms doesn't work: for instance, consider the case when all A_i are identical.

The answer (due to Linial and Nisan, 1988) is that for $k \geq \Omega(\sqrt{n})$, we can get a good approximation, correct upto a multiplicative factor $1 - \mathcal{O}(\exp(\frac{-2k}{n}))$, while for $k \leq \mathcal{O}(\sqrt{n})$, no good approximation is possible.

Our approach is to look at a related question. Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ and $\mathcal{B} = (B_1, B_2, \ldots, B_n)$ be two collections of sets. Assume that for any set $S \subset [n]$ with $|S| \leq k$ we have $\left|\bigcap_{i \in S} A_i\right| = \left|\bigcap_{i \in S} B_i\right|$. We want to estimate how far apart $|\bigcup A_i|$ and $|\bigcup B_i|$ can be.

Without loss of generality, we may assume that the A_i and the B_i are events in a probability space, and we will consider a slightly more general question, viz. we would like to estimate

$$E(k, n) = sup(\mathbf{Pr}[\bigcup_{i=1}^{n} A_i] - \mathbf{Pr}[\bigcup_{i=1}^{n} B_i]), \quad (14)$$

where $(A_1, A_2, \ldots, A_n)$ and $(B_1, B_2, \ldots, B_n)$ satisfy the condition

$$\mathbf{Pr}[\bigcap_{i \in S} A_i] = \mathbf{Pr}[\bigcap_{i \in S} B_i] \quad (15)$$

for every $S \subset [n]$ such that $|S| \leq k$.

Now, we define a j-atom to be an intersection of exactly j of the A_i's and the complements of the remaining $(n - j)$. Note that any other event consisting of intersections of events A_i or their complements can be expressed in terms of these atoms, and that all the atoms are disjoint events. We call a collection of n events *symmetric* if all the j-atoms occur with the same probability.

LEMMA 3
The optimum value of $E_{k,n}$ is attained for some $\mathcal{A}$ and $\mathcal{B}$ that are symmetric.

PROOF: For any $\mathcal{A}$ we construct a *symmetric* collection $\mathcal{A}'$ whose each term in the Inclusion-Exclusion formula is the same. The Lemma then follows.

Obtain $\mathcal{A}'$ by setting the probability of each j-atom of $\mathcal{A}'$ to be the average of the probabilities of the j-atom of $\mathcal{A}$. There is only one n-atom, so $\Pr[A'_1 A'_2 \cdots A'_n] = \Pr[A_1 A_2 \cdots A_n]$. Now consider $\Pr[A'_1 A'_2 \cdots A'_{n-1}]$. It may be expressed as

$$\Pr[A'_1 A'_2 \cdots A'_{n-1} \overline{A_n}'] + \Pr[A'_1 A'_2 \cdots A'_{n-1} A'_n].$$

Doing a similar rewriting for all conjunctions of $(n-1)$ events, and adding, we see that the $(n-1)$th term of the inclusion-exclusion for $\mathcal{A}'$ is the same as for $\mathcal{A}$. Proceeding this way, a simple induction shows the equivalence of all terms. □

Now let

$$a_j = \text{sum of all } j\text{-atoms of } A_1, A_2, \ldots, A_n; \tag{16}$$

$$b_j = \text{sum of all } j\text{-atoms of } B_1, B_2, \ldots, B_n; \tag{17}$$

$$r_j = \text{sum of all } j\text{-intersections of } A_1, A_2, \ldots, A_n; \tag{18}$$

$$q_j = \text{sum of all } j\text{-intersections of } B_1, B_2, \ldots, B_n. \tag{19}$$

LEMMA 4
The following relation connects the r_j and the a_j:

$$r_j = \sum_{i=j}^{n} \binom{i}{j} a_i.$$

PROOF: Consider a generic term, say $\mathbf{Pr}[A_1 \cap A_2 \cap \ldots A_j]$ contributing to r_j. Expand this out as the summation of atoms:

$$\mathbf{Pr}[A_1 \cap A_2 \cap \ldots A_j] = \sum \mathbf{Pr}[A_1 \cap A_2 \cap \ldots A_j \cap s],$$

where s in the summation runs over all the atoms of $A_{j+1}, \ldots, A_n$.

In this summation, each i-atom is counted exactly $\binom{i}{j}$ times: once for each j-subset of the i uncomplemented events. The lemma follows. □

Now, we will construct an LP that solves (14).

LEMMA 5
The solution to (14) is given by solving the following LP:

$$max \quad \sum_{i=1}^{n} x_i : \tag{20}$$

$$\forall j \le k : \quad \sum_{i=j}^{n} \binom{i}{j} x_i = 0 \tag{21}$$

$$\forall S \subseteq [n] : \quad \sum_{i \in S} x_i \le 1 \tag{22}$$

$$-\sum_{i \in S} x_i \le 1 \tag{23}$$

PROOF: First, we check that $x_i = a_i - b_i$ is feasible: for (21),

$$\sum_{i=j}^{n} \binom{i}{j} x_i = r_j - q_j = 0 \quad \forall i \leq k,$$

and (22, 23) follow easily from noting that the a_j and b_j are probabilities of *disjoint* events, and so for any $S \subseteq [n]$ the sum $\sum_{i \in S} x_i$ represents a difference in probabilities and is therefore bounded in absolute value by 1.

For the reverse direction, let $x_1, x_2, \ldots, x_n$ be a feasible solution to the LP. Now define the a_j and b_j by

$$a_j = \begin{cases} x_i \text{ if } x_i > 0 \\ 0 \text{ otherwise} \end{cases} \tag{24}$$

$$b_j = \begin{cases} -x_i \text{ if } x_i < 0 \\ 0 \text{ otherwise} \end{cases} \tag{25}$$

It is easy to check that this defines proper probabilities using (22, 23), and that the sequences of events they define satisfy (15) because of (21). □

Now we look at the dual of the LP. It involves finding $\mu_S, \lambda_S \geq 0$ and y_i's such that

$$\min \quad \sum_{S \subseteq [n]} (\lambda_S + \mu_S) : \tag{26}$$

$$\forall 1 \leq i \leq n : \quad \sum_{S \ni i} (\lambda_S - \mu_S) + \sum_{j \leq min(i,k)} \binom{i}{j} y_j \geq 1 \tag{27}$$

Clearly, the optimum solution will, for each subset S, never need to make both λ_S and μ_S positive. For example, if $\lambda_S = a > 0$ and $\mu_S = b > a$ then making $\lambda_S = 0$ and $\mu_S = b - a$ still satisfies all constraints while lowering the objective.

For any given $y_1, y_2, \ldots, y_n$, what is the best choice of λ_S, μ_S? For each i let $c_i = 1 - \sum_{j \leq min(i,k)} \binom{i}{j} y_j$. Let $I = \{i : c_i > 0\}$ and $J = \{j : c_j < 0\}$. Let $c_+ = \max\{c_i : i \in I\}$ and $c_- = \min\{-c_j : j \in J\}$. (If I or J is empty the corresponding max or min is defined to be 0.) Then there is a dual solution of cost $c_+ + c_-$, namely, $\lambda_I = c_+$, $\mu_J = c_-$ and the variables associated with all other subsets are zero. Furthermore, every feasible solution must have *some* set with $\lambda_S \geq c_+$ and some other set with $\mu_S \geq c_-$, and thus have cost at least $c_+ + c_-$. Finally, we claim that at the optimum, the y_i's are such that $J = \emptyset$, and hence $c_- = 0$. Suppose not, and $c_- > 0$. Then divide all y_i by $1 + c_-$; a simple calculation shows that the new c_- is 0 whereas the new c_+ is $\frac{c_+ + c_-}{1 + c_-}$. Thus the objective function has gone down, which contradicts optimality.

We have thus proved:

LEMMA 6

The dual solution is the following optimization problem.

$$\min_{y_1,\ldots,y_n} \left[\max_{1\le i\le n} \left(1 - \sum_{j\le min(i,k)} \binom{i}{j} y_j \right) \right] \; : \tag{28}$$

$$\left(1 - \sum_{j\le min(i,k)} \binom{i}{j} y_j \right) \ge 0 \qquad \forall i. \tag{29}$$

LEMMA 7
The optimum value of the program in Lemma 6 is given by

$$\inf_q \left(\max_{m\in[n]} (1 - q(m)) \right) \tag{30}$$

where the infimum ranges over all polynomials q of degree atmost k with constant term 0 such that $q(m) \le 1$ for all $m \in \{1, \ldots, n\}$.

PROOF: Recall that $\binom{x}{i} = \frac{x(x-1)(x-2)\cdots(x-i+1)}{i!}$, which is a degree i polynomial whose constant term is 0. (That is, at $x = 0$ its value is 0.) It is also a polynomial that is 0 at $x = 1, 2 \ldots, (i-1)$. We note that any polynomial of degree at most k and constant term 0 can be written as a linear combination of $\binom{x}{i}$, for $1 \le i \le k$. (The proof is by induction. If the polynomial is $cx^i + q(x)$ where $q(x)$ is a polynomial of degree at most $i-1$, then it may be expressed as $ci!\binom{x}{i} + r(x)$ where $r(x)$ has degree at most $i-1$.) Finally, note that if we define the polynomial $q(x) = \sum_{j=1}^{k} \binom{x}{j} y_j$, then (28) becomes simply

$$\inf_q \left(\max_{m\in\{1,\ldots,n\}} (1 - q(m)) \right)$$

as required. The proof follows. □

To prove an upperbound on the primal, we construct a suitable *feasible* solution to the dual (28). We use the *Chebyshev polynomials*. Here are some of their properties:

1. Recall that for each integer $m \ge 0$, $\cos(m\theta)$ is a polynomial in $\cos(\theta)$ of degree m. Thus $\cos(m \cos^{-1}(x))$ is a degree m polynomial in x, called the m'th Chebyshev polynomial $T_m(x)$. It is also given by

$$T_m(x) = \frac{(x + \sqrt{x^2-1})^m + (x - \sqrt{x^2-1})^m}{2}. \tag{31}$$

2. For $x \in [-1, 1]$, we have $-1 \le T_m(x) \le 1$.

Consider the following polynomial of degree k,

$$q_{k,n}(x) = 1 - \frac{T_k(\frac{2x-(n+1)}{n-1})}{T_k(\frac{-(n+1)}{n-1})} \tag{32}$$

Note that $q(0) = 0$ (i.e. its constant term is 0) and when $x \in [1, n]$, $T_k(\frac{2x-(n+1)}{n-1}) \in [-1, 1]$ so $|q_{k,n}(x) - 1| \leq 1/D$ where $D = \left|T_k(\frac{-(n+1)}{n-1})\right|$.

Then $p(x) = Dq_{k,n}(x)/(1 + D)$ satisfies $\frac{D-1}{1+D} \leq p(m) \leq 1$ for all $m \in [1, n]$. Thus it is a dual feasible solution and we conclude that $E(k, n) \leq 1 - \frac{D-1}{1+D} = \frac{2}{1+D}$. Thus the maximum ratio for $\Pr[\cup_i A_i]$ and $\Pr[\cup_i B_i]$ is $\frac{1}{1-E(k,n)} \leq \frac{D+1}{D-1}$. It only remains to estimate this quantity. Since $D = \left|T_k(-\frac{n+1}{n-1})\right| = (\lambda^k + \lambda^{-k})/2$ where $\lambda = \frac{\sqrt{n}+1}{\sqrt{n}-1} \approx 1 + \frac{2}{\sqrt{n}}$ for large n, we can upperbound this by

$$\frac{D+1}{D-1} \leq 1 + O(k^2/n).$$

2 A note on algorithms

We have emphasized the use the duality theorem as a tool for proving theorems. Of course, the primary use of LPs is to solve optimization problems. Several algorithms exist to solve LPs in polynomial time. We want to mention Khachiyan's *ellipsoid algorithm* in particular because it can solve even exponential size LPs provided they have a polynomial time *separation oracle*. (There is an additional technical condition that we need to know a containing ball for the polytope in question, and the ball should not be too much bigger than the polytope. Usually this condition is satisfied.)

A separation oracle for an LP decides whether a given input $(x_1, x_2, \ldots, x_n)$ is feasible or not, and if it isn't, outputs out one constraint that it violates.

For example, consider the dual of max-flow (viz. fractional min-cut) that was discussed in the previous lecture:

$$\min \sum_{e \in E} c_e y_e : \tag{33}$$

$$\forall e \in E : y_e \geq 0 \tag{34}$$

$$\forall P \in \mathcal{P} : \sum_{e \in P} y_e \geq 1 \tag{35}$$

This can be solved in many ways, but the simplest (if we do not care about efficiency too much) is to use the Ellipsoid method, since we can design a polytime separation oracle for this problem using the *shortest path algorithm*. Suppose the oracle is given as input a vector $(y_e)_{e \in E}$. To decide if it is feasible, the oracle computes the shortest path from s to t with edge weights $= y_e$, and checks if the length of this shortest path is atleast 1. (Of course, before anything else one should check if all the $y_e > 0$.) Clearly, $(y_e)_{e \in E}$ is feasible iff the shortest path has length at least 1, and if it is infeasible then the shortest path constitutes an unsatisfied constraint.