CS 466/666 Algorithm Design and Analysis . Spring 2020

Lecture 20 : Linear programming duality

We prove the strong duality theorem for linear programming.

Then we see some applications in combinatorial optimization and game theory.

Weak duality in linear programming

Consider the following linear program.

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 \\ & 4x_1 + 8x_2 \leq 12 \\ & 2x_1 + x_2 \leq 3 \\ & 3x_1 + 2x_2 \leq 6 \\ & x_1, x_2 \geq 0 \end{aligned}$$

This is a maximization problem.

To give a lower bound on the optimal value, we just need to give a feasible solution.

In this example, setting $x_1 = x_2 = 1$ shows that the optimal value is at least 5.

Could we find a better solution? If not, how do we prove that the optimal value is at most 5.

Looking at the first constraint, we know that $2x_1 + 4x_2 \leq 6$. Since $x_2 \geq 0$, this implies that $2x_1 + 3x_2 \leq 6$.

This shows that the optimal value is at most 6. Can we do better?

Add the first two constraints and divide by three, we obtain that any feasible solution must satisfy $2x_1 + 3x_2 \leq 5$.

Now, consider a general linear program

$$\begin{aligned} \max \quad & \langle c, x \rangle \\ & Ax \leq b \quad \text{where } A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c, x \in \mathbb{R}^n \\ & x \geq 0 \end{aligned} \qquad \Leftrightarrow \qquad \begin{aligned} \max \quad & \langle c, x \rangle \\ & \langle a_1, x \rangle \leq b_1 \\ & \vdots \\ & \langle a_m, x \rangle \leq b_m \\ & x \geq 0 \end{aligned}$$

We can generalize the same strategy to given an upper bound on the optimal value.

Take some positive linear combination of the constraints, i.e. $\langle y_1 \vec{a}_1 + y_2 \vec{a}_2 \rangle \leq y_1 b_1 + y_2 b_2$,

where y_1, y_2 being positive is to ensure that the inequality holds in the right direction.

Now, if $y_1 \vec{a}_1 + y_2 \vec{a}_2$ "dominates" $\vec{c}$, i.e. $y_1 \vec{a}_1 + y_2 \vec{a}_2 \geq \vec{c}$, then we know that $\langle y_1 \vec{a}_1 + y_2 \vec{a}_2, x \rangle \geq \langle \vec{c}, x \rangle$

as $x \geq 0$, and thus $y_1 b_1 + y_2 b_2$ is an upper bound on the optimal value.

How to find the best upper bound using this method? This is itself a linear programming problem!

Associate a non-negative number y_i to each constraint above. The best / smallest upper bound is:

$$\min \langle b,y\rangle \qquad\qquad \min \langle b,y\rangle$$

$$\sum_{i=1}^{m} y_i a_i \geq c \qquad \leftrightarrow \qquad y^T A \geq c$$

$$y_i \geq 0 \quad \forall i \qquad\qquad y \geq 0.$$

We call this pair of linear programs a primal-dual pair:

$$\max \langle c,x\rangle \qquad\qquad \min \langle b,y\rangle$$

$$Ax \leq b \qquad\qquad y^T A \geq c$$

$$x \geq 0 \qquad\qquad y \geq 0$$

From our discussion, the primal objective value is upper bounded by the dual optimal value.

This is called the weak duality theorem in linear programming.

Weak duality theorem Any feasible primal solution x and any feasible dual solution y satisfy $\langle c,x\rangle \leq \langle b,y\rangle$.

Proof $\langle b,y\rangle = y^T b \geq y^T A x \geq c^T x = \langle c,x\rangle$, where we use $y \geq 0$, $x \geq 0$ in the inequalities. □

Complementary slackness conditions

How to prove that a primal feasible solution x is an optimal solution?

One way is to find a dual feasible solution y such that $\langle c,x\rangle = \langle b,y\rangle$, in which case

we actually prove that both x is primal optimal and y is dual optimal.

When will $c^T x = b^T y$?

Recall from the proof of the weak duality theorem that $y^T b \geq y^T A x \geq c^T x$.

So, to have $c^T x = b^T y$, we must have both inequalities satisfy as equalities.

For the first inequality to hold as an equality, we should have $y_i > 0$ only if $\langle a_i, x\rangle = b_i$,

i.e we should use the i-th row only when the i-th constraint is tight.

For the second inequality to hold as an equality, we should have $\langle a^j, y\rangle = c_j$ where a^j denotes the

j-th column of A whenever $x_j > 0$, i.e. if $x_j > 0$ then the coefficient in the combination should

match exactly the coefficient in the objective function (in other words, the dual constraint is tight).

In fact, these are necessary and sufficient conditions for x, y to be optimal.

Complementary Slackness Conditions Let x be a primal feasible solution and y be a dual feasible solution.

Then x and y are optimal if and only if they satisfy:

Primal conditions: If $x_j > 0$, then $\langle a^j, y\rangle = c_j$, where a^j is the j-th column of A.

Dual conditions: If $y_i > 0$, then $\langle a_i, x\rangle = b_i$, where a_i is the i-th row of A.

The complementary slackness conditions can guide us to search for optimal primal and dual solutions using a combinatorial algorithm, e.g. the Hungarian algorithm for weighted bipartite matching.

These are called primal-dual algorithms, using the primal-dual pair without explicitly solving the LP, and they lead to simple and efficient algorithms for combinatorial problems.

There are also "approximate complementary slackness conditions" for the design of approximation algorithms.

Some well-known examples are for facility location and network design problems. See the book by Williamson & Shmoys.

Convex analysis

We have derived necessary and sufficient condition for a pair of primal and dual solutions to be optimal.

But we have not proved that they must exist, and this is the content of the strong duality theorem, which says that the dual program always provides a tight upper bound of the primal program.

There are different ways to prove the strong duality theorem.

One is based on the simplex algorithm, which is elementary and self-contained, but is specific to LP and not as insightful.

Instead, we will see a general approach based on simple convex geometry, but a slight disadvantage is that we need to assume some basic results in analysis and we will not see all the details of the proof.

Separation theorem

To prove strong duality theorems, we need the following fundamental result in convex analysis, showing that we can always find a seperating hyperplane to seperate a point not in a closed convex set from the convex set.

Recall that this is what we needed in the ellipsoid method as well.

Definition A set S is closed if every limit point of S is in S.

Definition A set S is convex if $x, y \in S$, then $\alpha x + (1-\alpha) y \in S$ for all $0 \le \alpha \le 1$.

Seperation theorem Let $S \subseteq \mathbb{R}^n$ be a non-empty closed convex set and $v \notin S$.

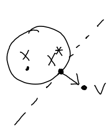
Then there exists $w \in \mathbb{R}^n$ such that $\langle w, v \rangle > \langle w, x \rangle$ for all $x \in S$.

The proof plan is simple and intuitive.

Given v, find the unique point $x^* \in S$ that is closest to v.

Then, argue that $\langle v - x^*, x - x^* \rangle \le 0$ for all $x \in S$, and this will give us the hyperplane with direction $v - x^*$.

The proof is summarized by the following picture



Lemma x^* is uniquely defined.

Lemma x^* is uniquely defined.

Proof Let x be an arbitrary point in S.

Consider $Y := \{ y \in S \mid \|y-v\|_2^2 \le \|x-v\|_2^2 \}$.

Then Y is bounded and closed, and thus compact.

A point closest to v is a minimizer of the continuous function $f(y) = \|y-v\|_2^2$ over compact set Y.

By Weierstrauss' theorem, the minimum is attained by some point $x^* \in S$.

To see that x^* is uniquely defined, suppose by contradiction that $x_1 \ne x_2$ and both are minimizers.

Consider $\bar{x} = \frac{1}{2}(x_1+x_2)$. By convexity, $\bar{x} \in S$. Let $\mu = \|x_1-v\|_2^2 = \|x_2-v\|_2^2$ be the minimum value.

Check that $\|\bar{x}-v\|_2^2 = \frac{1}{2}\|x_1-v\|_2^2 + \frac{1}{2}\|x_2-v\|_2^2 - \frac{1}{4}\|x_1-x_2\|_2^2 = \mu - \frac{1}{4}\|x_1-x_2\|_2^2$, a contradiction. □

Lemma x^* is the minimizer if and only if $\langle x-x^*, v-x^* \rangle \le 0$ for all $x \in S$.

Proof Let $x \in S$. Consider $y = (1-\varepsilon)x^* + \varepsilon x$. By convexity, $y \in S$.

Note that $\|y-v\|_2^2 = \|(x^*-v) - \varepsilon(x^*-x)\|_2^2 = \|x^*-v\|_2^2 - 2\varepsilon\langle x^*-v, x^*-x\rangle + \varepsilon^2\|x^*-x\|_2^2$.

Then x^* is the minimizer $\Leftrightarrow$ $-2\varepsilon\langle x^*-v, x^*-x\rangle + \varepsilon^2\|x^*-x\|_2^2 \le 0 \quad \forall x, \forall \varepsilon > 0$

$\Leftrightarrow$ $\langle x^*-v, x^*-x\rangle \le \frac{\varepsilon}{2}\|x^*-x\|_2^2 \quad \forall x, \forall \varepsilon > 0$

$\Leftrightarrow$ $\langle x^*-v, x^*-x\rangle \le 0. \quad \forall x.$ □

Proof of seperation theorem

Let $w = v - x^*$.

Then $\|v-x^*\| > 0 \Rightarrow \langle v, v-x^*\rangle > \langle x^*, v-x^*\rangle \Rightarrow \langle v, w\rangle > \langle x^*, w\rangle$.

For any $x \in S^*$, the previous lemma says $\langle x-x^*, v-x^*\rangle \le 0 \Rightarrow \langle x, v-x^*\rangle \le \langle x^*, v-x^*\rangle \Rightarrow \langle x, w\rangle \le \langle x^*, w\rangle$.

Therefore, $\langle v, w\rangle > \langle x, w\rangle$ for all $x \in S$. □

Farkas lemma and strong LP duality

We are going to use the separation theorem to prove Farkas lemma and then to derive strong LP duality.

Farkas lemma tells us exactly when a set of linear inequalities has no feasible solutions.

Theorem (Farkas lemma) The system $Ax=b$, $x \ge 0$ has no solutions if and only if

$$\exists y \text{ such that } y^T A \ge 0 \text{ and } y^T b < 0.$$

Proof $\Leftarrow$) If such a y exists, then the system must have no solutions, as otherwise

$$0 > y^T b = y^T A x \ge 0, \text{ a contradiction.}$$

$\Rightarrow$) Let $S = \{Ax \mid x \ge 0\}$. Then S is a closed convex set (see chapter 6 of Matousek-Gardner).

Note that the system $Ax=b,\ x \geq 0$ has no solutions is equivalent to saying that $b \notin S$.

By the separation theorem, there exists y such that $\langle y,b\rangle < \langle y,z\rangle \quad \forall z \in S$,

which can be rewritten as $\langle y,b\rangle < \langle y, Ax\rangle \quad \forall x \geq 0$.

Since $0 \in S$, we have $\langle y,b\rangle < 0$.

Also, we must have $y^TA \geq 0$, as otherwise there exists $x \geq 0$ such that $\langle y, Ax\rangle = \langle y^TA, x\rangle = -\infty$,

contradicting that $\langle y,b\rangle < \langle y,Ax\rangle \quad \forall x \geq 0$. □

Farkas lemma is often called a theorem of alternatives: Either the system has a solution or not, and in either case there is an easily verifiable condition for it.

In particular, the non-existence of x is equivalent to the existence of y!

We are ready to prove the strong LP duality theorem.

Theorem (strong LP-duality) If both the primal and dual are feasible, then they have the same objective value.

Proof Consider the primal linear program in the equality form:

$$\text{(primal)} \quad \begin{aligned} \max\ & c^Tx \\ & Ax=b \\ & x \geq 0 \end{aligned} \qquad\qquad \text{(dual)} \quad \begin{aligned} \min\ & b^Ty \\ & y^TA \geq c^T \end{aligned}$$

We would like to show that if the primal objective value is less than a value μ, then there is a dual feasible solution y with objective value less than μ, and this would imply strong duality.

The primal objective value is less than μ is equivalent to the following system is infeasible.

$$\begin{aligned} c^Tx - s &= \mu \\ Ax &= b \\ x \geq 0,\ s \geq 0&,\ s \in \mathbb{R} \end{aligned} \quad \Leftrightarrow \quad \begin{pmatrix} A & 0 \\ c^T & -1 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} = \begin{pmatrix} b \\ \mu \end{pmatrix}, \quad x \geq 0,\ s \geq 0$$

By Farkas lemma, this is infeasible iff $\exists y \in \mathbb{R}^m$ (where $A \in \mathbb{R}^{m\times n}$) and $z \in \mathbb{R}$ such that

$$(y^T\ \ z)\begin{pmatrix} A & 0 \\ c^T & -1 \end{pmatrix} \geq 0 \quad \text{but} \quad (y^T\ \ z)\begin{pmatrix} b \\ \mu \end{pmatrix} < 0.$$

More compactly, $y^TA + zc^T \geq 0$, $-z \geq 0$, and $y^Tb + z\mu < 0$.

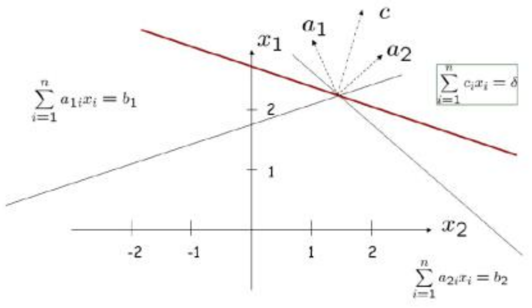
We consider two cases.

Suppose $z=0$, then this implies that $y^TA \geq 0$ and $y^Tb < 0$, which means that the primal linear program has no feasible solution at all, contradicting to our feasibility assumption.

Suppose $z \neq 0$. Then $z < 0$, and we have $\left(\frac{y^T}{z}\right) A \geq c^T$ and $\left(\frac{y^T}{-z}\right) b < \mu$.

Therefore, $\frac{-y^T}{z}$ is a dual feasible solution with objective value less than μ. □

Geometric interpretation



If x is an optimal solution in the direction c,
then the objective function should be in the cone of the tight constraints, i.e. $\exists y \geq 0$ s.t. $y^T A = c$.
Otherwise, we could find an improved solution in the direction c, and thus x is not optimal.

Min-max theorems in combinatorial optimization

Min-max theorems are fundamental results in combinatorial optimization, as they give nice characterizations of the optimal value of the problem. Some well-known examples include:

Max-bipartite-matching: For any bipartite graph, the maximum size of a matching is equal to the minimum size of a vertex cover. Note that Hall's theorem is implied by it.

Max-flow min-cut: For any directed graph and any two vertices s and t, the maximum value of an s-t flow is equal to the minimum value of an s-t cut.

We now see that these theorems can be proved systematically using LP duality.

Bipartite matching

(primal) $\max \sum_e x_e$
$x(\delta(v)) \leq 1 \quad \forall v \in V$
$x_e \geq 0 \quad \forall e \in E$

(dual) $\min \sum_v y_v$
$y_u + y_v \geq 1 \quad \forall uv \in E$
$y_v \geq 0 \quad \forall v \in V$.

Note that any integral solution of the dual program corresponds to a vertex cover.
It can be shown that the dual LP has integral optimal solutions (homework?).
So, the min-max theorem follows from the strong duality theorem.

Maximum flow

(primal) $\max f_{ts}$
$f(\delta^{in}(v)) - f(\delta^{out}(v)) \leq 0 \quad \forall v \in V$
$f_e \leq 1 \quad \forall e \in E$
$f_e \geq 0 \quad \forall e \in E$

(dual) $\min \sum_e d_e$
$d_{uv} + y_v - y_u \geq 0 \quad \forall uv \in E$
$y_s - y_t \geq 1$
$y_v \geq 0 \quad \forall v \in V$

$f_e \geq 0 \quad \forall e \in E$ $\qquad\qquad$ $y_v \geq 0 \quad \forall v \in V$

$d_e \geq 0 \quad \forall e \in E$

To understand the dual program, first observe that $d_{uv} = y_u - y_v \quad \forall uv \in E$ in any optimal solution.

Also, we can assume that $y_s = 1$ and $y_t = 0$, and $1 \geq y_v \geq 0$ for all $v \in V$.

Suppose $y_v \in \{0,1\} \quad \forall v$, then it corresponds to the cut where all vertices with $y_v = 1$ are on the source side, while all vertices with $y_v = 0$ are on the sink side, and the objective function counts the number of edges in this s-t cut.

Again, it can be proved that both the primal and the dual are integral (exercises?), and thus the max-flow min-cut theorem follows from the strong duality theorem.

Minimax theorem in game theory

There are also various uses of the strong duality theorem in game theory.

The most fundamental one is the following minimax theorem in two players zero-sum games.

A two players zero-sum game can be described by a matrix, where each row corresponds to a strategy of the row player, and each column corresponds to a strategy of the column player.

If the row player chooses strategy i and the column player chooses strategy j, then the payoff for the players is the (i,j)-th entry of the matrix.

The row player's goal is to maximize the payoff while the column player's goal is to minimize the payoff.

	P	S	R
P	0	-1	1
S	1	0	-1
R	-1	1	0

e.g. this is the payoff matrix of the paper-scissor-rock game.

A Nash equilibrium is a pair of strategies of the players, such that even if a player knows the strategy of the other player, he/she cannot gain by changing his/her own strategy.

It should be clear that there is no pure strategy solution in the paper-scissor-rock game, but the mixed strategy $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is an equilibrium solution for both players.

We now show that any two-players zero-sum game has a mixed strategy equilibrium solution.

Let $A \in \mathbb{R}^{m \times n}$ be the payoff matrix.

Let $x = (x_1, \ldots, x_m)$ be the probability distribution of the row strategies, s.t. $\sum_{i=1}^{m} x_i = 1$, $x_i \geq 0 \ \forall i$.

We denote it by $x \in \Delta^m$. Similarly, the column mixed strategy is denoted by $y \in \Delta^n$.

If the row player plays x and the column player plays y, then the payoff is simply $x^T A y$.

The minimax theorem says that if both players play optimally, then it doesn't matter who announce their mixed strategy first, and thus they form an equilibrium solution.

Theorem (minimax theorem) $\max_{x \in \Delta^m} \min_{y \in \Delta^n} x^T A y = \min_{y \in \Delta^n} \max_{x \in \Delta^m} x^T A y$.

On LHS, the row player announces first, while on RHS, the column player announces first.

Proof First, observe that once the row player fixes a strategy x, then the column player can compute $x^T A = (z_1, z_2, \dots, z_n)$ where z_i is the expected payoff if the column player plays strategy i.

Note that playing the pure strategy i with minimum z_i is a best response.

So, we can simplify $\max_{x \in \Delta^m} \min_{y \in \Delta^n} x^T A y$ as $\max_{x \in \Delta^m} \min_i (x^T A)_i$, which can be written as an LP:

$$\begin{aligned} \max \quad & t \\ & \sum_{i=1}^m x_i a_{ij} \geq t \quad && \forall\, 1 \leq j \leq n \\ & \sum_{i=1}^m x_i = 1 \\ & x_i \geq 0 && \forall\, 1 \leq i \leq m \end{aligned}$$

Similarly, $\min_{y \in \Delta^n} \max_{x \in \Delta^m} x^T A y$ can be simplified as $\min_{y \in \Delta^n} \max_j (Ay)_j$, which can be written as an LP:

$$\begin{aligned} \min \quad & r \\ & \sum_{j=1}^n a_{ij} y_j \leq r \quad && \forall\, 1 \leq i \leq m \\ & \sum_{j=1}^n y_j = 1 \\ & y_j \geq 0 && \forall\, 1 \leq j \leq n \end{aligned}$$

Now, check that these LP are a primal-dual pair. So the minimax theorem from strong LP duality.

Yao's minimax principle

Yao observed that the minimax principle can be used to prove lower bounds for randomized algorithms.

The worst case running time of a randomized algorithm is its running time on the worst input.

Suppose we want to prove a lower bound on the running time of any randomized algorithm to solve a problem.

Note that a randomized algorithm is just a distribution of deterministic algorithms, i.e. when the random bits are fixed, the algorithm behaves deterministically.

Think of it as a two-player zero-sum game where there is an adversary who likes to play the worst distribution of inputs to maximize the running time, while the randomized algorithm player wants to play the best distribution of algorithms to minimize the running time.

Suppose the number of inputs is finite and the number of deterministic algorithms is finite.

Let $A \in \mathbb{R}^{m\times n}$ be the payoff matrix, where each row corresponds to an input, and each column corresponds to a deterministic algorithm, and the (i,j)-th input is the running time of algorithm j on input i.

Let $\vec{i} \in \Delta^m$ be a probability distribution of the inputs, and $\vec{r} \in \Delta^n$ be a probability distribution of the algorithms.

The complexity we are interested in analyzing is $\min_{\vec{r}\in\Delta^n} \max_{\vec{i}\in\Delta^m} \vec{i} A \vec{r}$, the running time of the best randomized algorithms.

By the minimax theorem, it is equal to $\max_{\vec{i}\in\Delta^m} \min_{\vec{r}\in\Delta^n} \vec{i} A \vec{r}$.

So, to prove lower bound on the complexity of the best randomized algorithms, it is enough to come up with a distribution of the inputs and prove that any deterministic algorithm would take a long time.

It is usually easier to reason about deterministic algorithm, and this approach is widely used.

Note that any distribution of the inputs would give a lower bound (the weak duality theorem), but the minimax theorem says that one could prove the optimal lower bound this way.

See chapter 2 of "randomized algorithms" for an example of proving lower bound this way.

Tolls for multicommodity flow

We mention one more application with a game theory flavor.

Suppose there is a traffic network or a communication network.

There are n users in the network, where each user i wants to send d_i units of information from s_i to t_i.

If we let each user to choose their own paths to send the information, they may all send along the shortest paths and this may cause high congestions on some edges, and thus a bad utilization of the network.

Instead of directly controlling their behaviours, what we could do is to give prices on edges, and charge the users on the edges they used, and the hope is to avoid over congestion and to a better social welfare.

This problem can be formalized as follows.

Let P_i be the set of all possible s_i-t_i paths. For each path $p \in P_i$, we have a variable f_p^i.

Let c_e be the congestion of edge e that we would like to enforce.

Let l_p^c be the latency of the path p, given the current congestion pattern.

Our objective is to minimize the total latency while enforcing the congestion:

$$\min \sum_{i=1}^{n} \sum_{p\in P_i} l_p^c f_p^i$$

$$\sum_{p\in P_i}^{n} \sum_{e\in p} f_p^i \le c_e \qquad \forall e \qquad \text{(congestion constraint)}$$

$$\sum_{p\in P_i} f_p^i = d_i \qquad \forall i \qquad \text{(information constraint)}$$

$$f_p^i \ge 0 \qquad \forall p\ \forall i$$

The dual program is $\max \sum_{i=1}^{n} d_i z_i + \sum_e c_e t_e$

$$z_i - \sum_{e: e \in p} t_e \le l_p^c \quad \forall i \ \forall p \in P_i$$

$$t_e \ge 0 \quad \forall e$$

Now, if we set the price on each edge e to be t_e, and a user needs to pay $\sum_{e \in p} t_e$ dollars for sending one unit of information along path p.

Then, we can direct users to follow the optimal primal solution that minimizes congestion and the total latency.

The reason is that the paths with $f_p^i > 0$ would have $z_i = l_p^c + \sum_{e: e \in p} t_e$ by complementary slackness, while all other paths p would have $z_i \le l_p^c + \sum_{e: e \in p} t_e$.

So, the paths with $f_p^i > 0$ are the paths that the selfish users would have no incentive to switch, and so this is an equilibrium solution (and under some conditions the unique equilibrium solution).

To summarize, by setting price on edges, the network administrator can set the global optimal solution to be the (unique) equilibrium solution, so that the selfish users will be directed to this solution.

References

See chapter 6 of "understanding and using linear programming" for more about LP duality.

See chapter 7 of "the design of approximation algorithms" for primal-dual approximation algorithms.

See "combinatorial optimization: polyhedra and efficiency" for integrality proofs of LP relaxations.

See chapter 2 of "randomized algorithms" for a lower bound proof using Yao's minimax principle.

The last part is from the paper "tolls for heterogenous selfish users in multicommodity networks ...".

There are some very interesting algorithmic results using the primal-dual method to design online algorithms.

The survey "the design of competitive online algorithms via a primal-dual approach" by Buchbinder and Naor, and the paper "randomized primal-dual analysis of RANKING for online bipartite matching" are highly recommended.