CS 521: Linear Programming **Rutgers: Fall 2022**

Lecture 4

September 30, 2022

Instructor: Sepehr Assadi *Scribes: Xiang Ao, Wensen Mao, Chen Wang, Yuning Wang*

Topics of this Lecture

1 A Motivating Example 1

2 Duality in Linear Programming 3

2.1 Primal and Dual LPs . . . 3

2.2 Weak and Strong Duality Theorems . . . 4

3 LP Duality Applications 7

3.1 Optimizing vs Feasibility Checking LPs . . . 7

3.2 $\mathsf{LP} \in \mathsf{NP} \cap \mathsf{coNP}$ . . . 7

3.3 Bipartite Matching and Vertex Covers . . . 8

4 Proof of Strong Duality (Theorem 6) 9

4.1 Farkas Lemma . . . 9

4.2 Strong Duality From Farkas Lemma . . . 11

1 A Motivating Example

Consider the following linear program:

$$\begin{aligned} \max_{x_1,x_2} \quad & 2x_1 + 3x_2 \\ \text{subject to} \quad & 4x_1 + 8x_2 \leqslant 12 \\ & 2x_1 + x_2 \leqslant 3 \\ & 3x_1 + 2x_2 \leqslant 4 \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

While the optimal solution to this LP may not be immediately clear, it is easy to see that it is at most 12 just given the first constraint:

$$\underbrace{2x_1 + 3x_2}_{\text{objective function}} \leqslant \underbrace{4x_1 + 8x_2 \leqslant 12}_{\text{first constraint}}.$$

In fact, we can divide both sides of the same constraint by two and obtain an even better upper bound of 6 on the objective function:

$$\underbrace{2x_1 + 3x_2}_{\text{objective function}} \leqslant 2x_1 + 4x_2 = \underbrace{\frac{1}{2} \cdot (4x_1 + 8x_2) \leqslant 6}_{\text{first constraint}}.$$

Yet another way is to sum up the first two constraints and then divide by three to obtain an upper bound of 5 as follows:

$$\underbrace{2x_1 + 3x_2}_{\text{objective function}} = \frac{1}{3} \cdot (6x_1 + 9x_2) = \frac{1}{3} \cdot (\underbrace{4x_1 + 8x_2}_{\text{LHS of first constraint}} + \underbrace{2x_1 + x_3}_{\text{LHS of second constraint}}) \leqslant \frac{1}{3} \cdot (12 + 3) = 5.$$

We can continue playing this game to find better and better upper bound. In fact, we can formulate this as game follows:

1. Multiply variables y_1, y_2, y_3 separately on the three inequalities, which leads to
$$\begin{aligned} &y_1 \cdot (4x_1 + 8x_2) \leqslant 12 \cdot y_1 \\ &y_2 \cdot (2x_1 + x_2) \leqslant 3 \cdot y_2 \\ &y_3 \cdot (3x_1 + 2x_2) \leqslant 4 \cdot y_3 \\ &x_1, x_2 \geqslant 0 \\ &y_1, y_2, y_3 \geqslant 0. \end{aligned}$$
(We ensures the order of inequalities do not change by making y_1, y_2, y_3 non-negative).

2. Sum up the above inequality constraints to obtain the following linear combination of the constraints:
$$(4y_1 + 2y_2 + 3y_3) \cdot x_1 + (8y_1 + y_2 + 2y_3) \cdot x_2 \leqslant 12y_1 + 3y_2 + 4y_3. \tag{1}$$

3. Ensure that the coefficient of x_1 is at least 2 and the coefficient of x_2 is at least 3 in the above equation:
$$\begin{aligned} 4y_1 + 2y_2 + 3y_3 &\geqslant 2 \\ 8y_1 + y_2 + 2y_3 &\geqslant 3. \end{aligned}$$

4. With these constraints, we now have that RHS of Eq (1) will be an *upper bound* on the objective value of the original program; thus, to obtain the best upper bound, we should minimize this expression as much as possible. Putting all these together gives us the following formulation:
$$\begin{aligned} \min_{y_1,y_2,y_3} \quad & 12y_1 + 3y_2 + 4y_3 \\ \text{subject to} \quad & 4y_1 + 2y_2 + 3y_3 \geqslant 2 \\ & 8y_1 + y_2 + 2y_3 \geqslant 3 \\ & y_1, y_2, y_3 \geqslant 0. \end{aligned}$$

It is easy to see that the formulation of our game for finding an upper bound on the original LP, it itself another LP. This LP is referred to as the *dual* of the original program which is called the *primal* LP in this context. Thus, for our maximization primal LP, the dual is a minimization LP that upper bounds the objective value of the primal.

An immediate consequence of our dual construction is that *any* feasible solution to the dual LP already gives us a valid upper bound on the optimal objective of the primal LP (this is often referred to as *"weak duality"*). In other words, finding an upper bound for the primal LP reduces to finding a feasible solution in the dual. But how tight is this approach, i.e., how well can we upper bound the objective of the primal LP if we manage to minimize the dual? The answer is *completely tight*! Minimum value of the dual LP coincides with the maximum value of the primal LP (this is often referred to as *"strong duality"*).

In the rest of this lecture, we formalize these notions and prove weak and strong duality of LPs. We will then see some simple applications of duality.

2 Duality in Linear Programming

2.1 Primal and Dual LPs

Consider a primal LP as follows:

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & A \cdot x \leqslant b \\ & x \geqslant 0, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, i.e., we have m constraints (beside non-negative constraints). Recall that this is one of the standard representation of LPs and any LP can be written in this form. For now, we work with this representation although **everything discussed in this lecture can be directly extended to any other representation** as well.

1. Let $y = [y_1, \ldots, y_m]$ be the vector of non-negative coefficients in the four step approach of previous section for writing the dual. Let a_i denote the i-th row of matrix A. We can thus write:

$$\begin{gathered} y_1 \cdot \langle a_1 \, , \, x \rangle \leqslant b_1 \cdot y_1 \\ y_2 \cdot \langle a_2 \, , \, x \rangle \leqslant b_2 \cdot y_2 \\ \vdots \\ y_m \cdot \langle a_m \, , \, x \rangle \leqslant b_m \cdot y_m. \end{gathered}$$

   In the matrix form, this can be written as:

$$y^T \cdot A \cdot x = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \odot \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \\ a_{m1} & & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \leqslant \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \odot \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = y^T \cdot b,$$

   where $\odot$ stands for the element-wise multiplication.

2. Now, by adding all these constraints together and factor out $x_1, \ldots, x_n$, we obtain that:

$$\langle y \, , \, a_1 \rangle \cdot x_1 + \langle y \, , \, a_2 \rangle \cdot x_2 + \ldots + \langle y \, , \, a_m \rangle \cdot x_m \leqslant \langle b \, , \, y \rangle.$$

3. In order for the LHS to form an upper bound on the objective value of the primal, we need to have,

$$\begin{gathered} \langle y \, , \, a_1 \rangle \geqslant c_1 \\ \langle y \, , \, a_2 \rangle \geqslant c_2 \\ \vdots \\ \langle y \, , \, a_m \rangle \geqslant c_m. \end{gathered}$$

   In the matrix form, this is simply

$$A^T y \geqslant c.$$

4. Finally, we would like to minimize the RHS of the equation in step (2), which gives us the following:

$$\begin{aligned} \min_{y \in \mathbb{R}^m} \quad & b^T y \\ \text{subject to} \quad & A^T \cdot y \geqslant c \\ & y \geqslant 0. \end{aligned}$$

This way, we get:

Definition 1. For any primal LP (P) as below, the dual LP (D) is defined as:

Primal LP:	**Dual LP:**
$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & A \cdot x \leqslant b \\ & x \geqslant 0, \end{aligned}$$	$$\begin{aligned} \min_{y \in \mathbb{R}^m} \quad & b^T y \\ \text{subject to} \quad & A^T \cdot y \geqslant c \\ & y \geqslant 0. \end{aligned}$$

An immediate observation given the above formula is that:

Observation 2 ("Dual of the dual is primal")**.** Given an LP (P) and its dual (D), the dual of (D) is (P).

Remark. There are other equivalent ways of getting to the dual of linear program (for instance, via Lagrangian multipliers). We will revisit some these approaches later in the course.

Remark. As stated earlier in this section, we can write the dual of any LP even if it is not in the representation we worked with in this section. For instance, if the original primal LP does not have the non-negativity constraint, we will have:

Primal LP:	**Dual LP:**
$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & A \cdot x \leqslant b. \end{aligned}$$	$$\begin{aligned} \min_{y \in \mathbb{R}^m} \quad & b^T y \\ \text{subject to} \quad & A^T \cdot y = c \\ & y \geqslant 0. \end{aligned}$$

This can be proven easily by first representing the LP with $2n$ non-negative variables (using the same method as in the transformation to the equational form) and then applying the previous approach.

2.2 Weak and Strong Duality Theorems

Weak duality. The weak duality theorem states that for any maximization primal LP (P), any feasible solution to its dual LP (D) gives an upper bound on the objective value of (P). Formally,

Theorem 3 ("Weak Duality")**.** *Let (P) be any maximization LP and (D) be its dual in Definition 1. For any feasible point x in (P) and any feasible point y in (D):*

$$c^T \cdot x \leqslant b^T \cdot y.$$

Proof. Proof of this theorem is straightforward basically by our construction of the dual (and the whole motivation behind it). For any $i \in [m]$ and $j \in [n]$, let a_i denote the i-th row and the j-th column of the matrix $A \in \mathbb{R}^{m \times n}$, respectively. This way, the primal and dual constraints will be

$$\forall i \in [m] \quad \langle a_i \, , \, x \rangle \leqslant b_i \qquad \text{and} \qquad \forall j \in [m] \quad \langle a^j \, , \, y \rangle \geqslant c_j. \tag{2}$$

We now have

$$
\begin{aligned}
c^T \cdot x = \sum_{j=1}^{n} c_j \cdot x_j \leqslant \sum_{j=1}^{n} \langle a^j \,,\, y \rangle \cdot x_j & \qquad \text{(by the right inequality of Eq (2))} \\
= \sum_{j=1}^{n} \sum_{i=1}^{m} a_{i,j} \cdot y_i \cdot x_j = \sum_{i=1}^{m} \langle a_i \,,\, x \rangle \cdot y_i & \qquad \text{(notice that the RHS is equal to } y^T A x\text{)} \\
\leqslant \sum_{i=1}^{m} b_i \cdot y_i = b^T \cdot y. & \qquad \text{(by the left inequality of Eq (2))}
\end{aligned}
$$

□

We have the following immediate corollary of Theorem 3.

Observation 4. Let (P) be any maximization LP and (D) be its dual in the form of Definition 1. Then,

1. If the objective value of (P) can go to $+\infty$ (is unbounded from above), then (D) has no feasible solution.
2. If the objective value of (D) can go to $-\infty$ (is unbounded from below), then (P) has to feasible solution.

Proof. For the first one, suppose (D) has a feasible solution y. Then, by weak duality, we need $b^T \cdot y$ to be larger than any feasible objective value of (P), but the latter is going to $+\infty$, a contradiction. The second part holds via the same argument. □

We also have a simple example that shows it is possible for both primal and dual LP to be infeasible.

Observation 5. The following is an example of a primal (P) and dual (D) pair which are both infeasible[1]:

(P)

$$
\begin{aligned}
\max_{x_1, x_2 \in \mathbb{R}^2} \quad & x_1 - x_2 \\
\text{subject to} \quad & x_1 + x_2 = 1 \\
& 2x_1 + 2x_2 = -2
\end{aligned}
$$

(D)

$$
\begin{aligned}
\min_{y_1, y_2 \in \mathbb{R}^2} \quad & y_1 - 2y_2 \\
\text{subject to} \quad & y_1 + 2y_2 = 1 \\
& y_1 + 2y_2 = -1.
\end{aligned}
$$

Just using weak duality and the above observations, we can infer the following possibilities between the primal and dual LPs:

(P)/(D)	Unbounded	Infeasible	Feasible
Unbounded	no	yes	no
Infeasible	yes	yes	???
Feasible	no	???	???

Table 1: Here, for the primal (P), being unbounded means having objective value going to $+\infty$ (first row) and for the dual (D), means having objective value going to $-\infty$ (first column).

What happens in the remaining cases? For instance, can it be the case that both primal and dual are infeasible? What can we say about primal and dual when they are both feasible and bounded? Strong duality handles these cases (and more!).

[1] You should first convince yourself that these two LPs are indeed primal and dual of each other.

Strong duality. Strong duality states that whenever at least one of the primal or dual is feasible and bounded, then so is the either one and more importantly, their objective values are equal. Formally,

Theorem 6 ("Strong Duality")**.** *Let (P) be any maximization LP and (D) be its dual in Definition 1. Suppose at least one of these LPs is both feasible and bounded. Then, both LPs are feasible and bounded and have optimal solutions x^* for (P) and y^* for (D) such that*

$$c^T \cdot x^* = b^T \cdot y^*.$$

We will prove Theorem 6 later in this lecture in Section 4. For now, we mention two consequences of this theorem. The first one is that we can now complete Table 2 as follows:

(P)/(D)	Unbounded	Infeasible	Feasible
Unbounded	no	yes	no
Infeasible	yes	yes	no
Feasible	no	no	yes†

Table 2: †: In this case, both optimal objective value of (P) and (D) is the same.

The second consequence, referred to as the **complementary slackness** condition, allows us to infer "some information" about the optimal assignment of primal and dual from each other (in addition to the fact that they both achieve the same objective value).

Theorem 7 ("Complementary slackness condition")**.** *Let (P) be any maximization LP and (D) be its dual in Definition 1. Suppose x' is a feasible solution to (P) and y' is a feasible solution to (D). Then, the following two conditions are equivalent:*

(i) *x' and y' are optimal solutions of their respective LPs.*

(ii) *$y'^T \cdot (Ax' - b) = 0$ and $x'^T \cdot (A^T y' - c) = 0$. I.e., only dual variables y'_i for $i \in [m]$ (resp. primal variables x'_j for $j \in [n]$) can be non-zero that for their corresponding constraint' $\langle a_i \, , \, x \rangle \leqslant b_i$ in the primal (resp. $\langle a^j \, , \, y \rangle \geqslant c_j$) holds with equality*

Proof. We prove each part separately.

- ***(i)* $\implies$ *(ii)*.** Recall the proof of the weak duality (Theorem 3) from earlier and let x^* and y^* be optimal solutions of (P) and (D), respectively. We showed that

$$c^T x^* \leqslant {y^*}^T A x^* \leqslant {y^*}^T b^*.$$

By strong duality (Theorem 6), we have that $c^T x^* = y^T b^*$, which implies both inequalities above are tight. As such

$$c^T x^* = {y^*}^T A x^* \implies ({y^*}^T A - c^T) \cdot x^* = 0 \implies {x^*}^T (A^T y^* - c) = 0$$
$${y^*}^T A x^* = {y^*}^T b^* \implies {y^*}^T (Ax^* - b) = 0,$$

as desired.

***(ii)* $\implies$ *(i)*.** Suppose now that x' and y' satisfy the given equations in *(ii)*. We have,

$$x'^T \cdot (A^T y' - c) = 0 \implies c^T x' = y'^T A x$$
$${y^*}^T (Ax^* - b) = 0 \implies y'^T A x' = y'^T b,$$

which implies that $c^T x' = b^T y'$, namely the objective value of x' in (P) and y' in (D) is the same. Since by weak duality (Theorem 3), we always have $c^T x \leqslant b^T y$ for any primal-dual pair (x, y), we have that both x' and y' need to be optimal.

□

Remark. Complementary slackness condition is a useful tool for inferring **some** information about optimal assignment of dual variables from the ones for the primal and vice versa. However, in general, one cannot recover the entirety of the optimal primal solution, given only the dual solution, or vice versa (more on this later in this lecture).

3 LP Duality Applications

Before getting to the proof of the strong duality theorem, let us visit some basic applications of LP duality.

3.1 Optimizing vs Feasibility Checking LPs

In Lecture 3, we saw that just checking whether a given LP (technically a polyhedron as the objective function does not matter here) has any feasible solution or not is almost as hard as solving the whole problem of finding an optimal point. The idea was that given any LP (LP1), include a new constraint to (LP1) that lower bounds the objective function, check whether this new LP has any feasible solution (using the blackbox algorithm for feasibility check), and based on that run a binary search by updating the extra inserted condition. We now show that using the strong duality of LPs, we can build a LP solver without the binary search strategy.

Concretely, suppose we are given a LP of the following form

$$\begin{aligned} \max_{x\in\mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & A\cdot x \leqslant b \\ & x \geqslant 0, \end{aligned}$$

where A is in $\mathbb{R}^{m\times n}$. We build a $(m+n)$-dimensional polyhedron as below by combining the LP and its dual and add one more constraint that requires the objective function of the LP and its dual LP to be equal.

Polyhedron (P1):

$$\begin{aligned} A\cdot x &\leqslant b \\ A^T\cdot y &\geqslant c \\ c^T\cdot x &= b^T\cdot y \\ x, y &\geqslant 0. \end{aligned}$$

Then we give this polyhedron to the black box algorithm as input. If the algorithms decides that the polyhedron has a feasible point (x^*, y^*), then we know the original LP has an optimal solution x^* by the strong duality theorem (Theorem 6), and if this polyhedron is infeasible, then the original LP should also be either unbounded or infeasible.

3.2 $\mathsf{LP} \in \mathsf{NP} \cap \mathsf{coNP}$

Consider the decision version of the LP problem as the set

$$\mathsf{LP} := \{(P, q) \mid \text{(P) is an LP of the form in Definition 1 with objective value at least } q\}.$$

For simplicity, we further assume that in this problem we are promised that (P) is bounded and feasible (this assumption is not necessary). In Lecture 2 and 3 we showed that this problem is at least decidable and in fact can be solved in exponential time. But what complexity class this problem belongs to?

An immediate consequence of strong duality is that $\mathsf{LP} \in \mathsf{NP} \cap \mathsf{coNP}$, meaning that both the "yes" instances and "no" instances of the LP problem can be efficiently *verified* in polynomial time:

- $\mathsf{LP} \in \mathsf{NP}$: The certificate for verification is a feasible solution x' for (P) such that $c^T x' \geqslant q$. The algorithm simply verifies that x' is a feasible solution of (P) by checking each constraint respectively and verify that $c^T x' \geqslant q$. Then, we can conclude the optimal value of (P) is at least q and the algorithm only takes polynomial time[2].
- $\mathsf{LP} \in \mathsf{coNP}$: To show this, we need to be able to prove that the objective value of (P) is less than the given parameter q. To do this, we take the dual (D) and the certificate is a feasible solution y' for (D) such that $b^T y' < q$. By strong duality (Theorem 6), such a certificate must always exist and by weak duality (Theorem 3), given y', we have the proof that $c^T x \leqslant b^T y' < q$ for all feasible x in (P). We can verify y' is a feasible solution in (D) exactly as in the first part.

Remark. While we have not covered this yet in the class, we in fact have $\mathsf{LP} \in \mathsf{P} \subseteq \mathsf{NP} \cap \mathsf{coNP}$, i.e., there are polynomial time algorithms for optimizing any linear program, in particular, Ellipsoid algorithms and Interior-point methods. We shall cover these algorithms later in this course.

3.3 Bipartite Matching and Vertex Covers

We defined the bipartite matching and vertex cover problems in Lecture 1: in the former problem we want to pick the maximum number of edges that do not share any vertices, and in the latter the goal is to pick the minimum number of vertices so that every edge has at least one chosen endpoint. We then proved that size of maximum matching is equal to minimum vertex cover in bipartite graphs. We now reprove this result here using LP duality.

Recall that the LP for the fractional matching in a bipartite graph $G = (L \sqcup R, E)$ is as follows:

$$\begin{aligned} \max_{x \in \mathbb{R}^E} \quad & \sum x_e \\ \text{subject to} \quad & \sum_{e \ni v} x_e \leqslant 1 && \forall\, v \in L \sqcup R, \\ & x_e \geqslant 0 && \forall\, v \in L \sqcup R. \end{aligned}$$

Let (P) denote this LP. We can obtain the dual (D) of (P) using the same approach as before to get:

$$\begin{aligned} \min_{y \in \mathbb{R}^V} \quad & \sum y_v \\ \text{subject to} \quad & y_u + y_v \geqslant 1 && \forall\, (u,v) \in E, \\ & y_v \geqslant 0 && \forall\, v \in L \sqcup R. \end{aligned}$$

As before, *if* assume the values in this dual (D) are integral, then we get a program for minimum vertex cover, because $y_v = 1$ can be interpreted as picking the vertex v in the solution, and the $y_u + y_v \geqslant 1$ constraints ensures that from every edge we are picking at least one vertex. Of course, in general, there is no reason (D) picks its solution integrally and thus (D) corresponds to the *fractional* vertex cover problem. However, we show that the optimum value of this LP is equal to the minimum vertex cover of G still.

Proposition 8. *Any feasible solution $y \in \mathbb{R}^V$ of* (D) *can be rounded to a vertex cover of the input graph with size at most $\sum_{v \in V} y_v$.*

Proof. We use a similar proof as that of the min-cut problem from the previous lecture. Pick $\theta \in [0,1]$ uniformly at random and define the sets:

$$L^* = \{u \mid u \in L \text{ and } y_u \geqslant \theta\} \qquad \text{and} \qquad R^* = \{v \mid v \in R \text{ and } y_v \geqslant 1 - \theta\}.$$

[2] We are a bit *cheating* here because we are ignoring the *bit-complexity* of x', namely, how large the proof itself needs to be. This is a topic that is addressed in homework 1, where we show that the bit-complexity of the optimal solution of the LP is polynomial in the input size

We claim the set $L^* \cup R^*$ is a vertex cover of G because for an arbitrary edge $(u,v) \in E$ such that $u \in L, v \in R$, at least one of $u \in L^*$ or $v \in R^*$ is true. If $u \notin L^*$ and $v \notin R^*$, then we have $y_u < \theta$ and $y_v < 1-\theta$, thus $y_u + y_v < 1$ which contradicts the feasibility of y for (D).

We now bound the size of the solution in expectation.

$$\begin{aligned}
\mathbb{E}\,|L^* \cup R^*| &= \sum_{u \in L} \Pr(u \in L^*) + \sum_{v \in R} \Pr(v \in R^*) \\
&= \sum_{u \in L} \Pr(y_u \geqslant \theta) + \sum_{v \in R} \Pr(y_v \geqslant 1-\theta) && \text{(by the definition of } L^* \text{ and } R^*\text{)} \\
&= \sum_{u \in L} y_u + \sum_{v \in R} y_v && (\theta \text{ is chosen uniformly at random}) \\
&= \sum_{v \in V} y_v.
\end{aligned}$$

This concludes the proof. □

By the results we have proven for fractional matching and vertex cover, plus strong duality (Theorem 6), we have that for any *bipartite* graph:

size of maximum matching = optimal value of (P) = optimal value of (D) = size of minimum vertex cover;

here, the first equality is by Proposition 3 of Lecture 1, the second one is by strong duality, and the third one is by Proposition 8.

Remark. It is worth examining what happens to the complementary slackness condition in the case of bipartite matching and vertex cover, when the input graph has a perfect matching. In this case, one optimal solution for the dual is to set $y^*_u = 1$ for all $u \in L$ and $y^*_v = 0$ for all $v \in R$. The complementary slackness condition now only implies that for all $u \in L$, $\sum_{e \ni u} x^*_e = 1$, which is an obvious condition on any graph that has a perfect matching. Thus, in this case, knowing the optimal dual solution does not tell us anything meaningful about the optimal primal assignment[a]

[a] This step can be made even more formal by showing that if we had an algorithm that could find a perfect matching in a graph that has a perfect matching given an optimal dual solution, then we can run the algorithm at most $O(\log n)$ times and obtain a maximum matching of any given graph without any assumption on perfect matching or access to an optimal dual solution. Such an algorithm can be easily constructed using a binary search approach and is left as an exercise.

4 Proof of Strong Duality (Theorem 6)

We now move to the proof of the strong duality. There are different proofs of this theorem and we shall revisit some of them later in the course. In particular, one can prove this result using the correctness of the Simplex algorithm; however, that approach is quite mechanical and uninformative (and relies on establishing correctness of the Simplex algorithm rigorously which is beyond the scope of this course). Instead, in this lecture, we are going to prove this result using an important tool in convexity, namely *Farkas Lemma*, which has a simple intuitive geometric interpretation.

4.1 Farkas Lemma

Consider the linear system $Ax = b$ for an $n \times n$ matrix A. What property ensures that this system has a solution? The answer is simple: A needs to be full rank and in that case $x = A^{-1} \cdot b$. What if A is an $m \times n$ matrix in general (where x may no longer be unique)? We can still solve this problem using Gaussian elimination by figuring out whether b is in the column space of A or not. What if we are now interested in checking whether $Ax = b$ has a solution which is *non-negative*?[3] This is the content of the Farkas Lemma.

[3] Recall that we visited this problem in the context of feasibility check for LPs in the equational form.

Proposition 9 (Farkas Lemma [1])**.** *For any $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$,* exactly one *of the following is true:*

1. *there exists some $x \geqslant 0$, such that $Ax = b$;*
2. *there exists some $y \in \mathbb{R}^m$ such that $y^T A \geqslant 0$ and $y^T b < 0$.*

It is easy to see that both statements of Proposition 9 cannot be true (prove this!). But, showing at least one of these two needs to be true is non-trivial and is the main part of the proof of Proposition 9. We will defer the proof of Proposition 9 to a later lecture. For now, we point out a highly intuitive explanation of Farkas lemma.

Consider the columns of A as vectors in $\mathbb{R}^m$ and the *convex cone* generated by them, i.e.,

$$\text{convex cone of } A := \{A \cdot z \mid z \geqslant 0\}\,.$$

Now, Farkas lemma states that either b belongs to the convex cone of A or there exists a hyperplane y that places all of A on one side, while placing b on the other side, namely, b can be separated from the convex cone of A via a hyperplane[4].

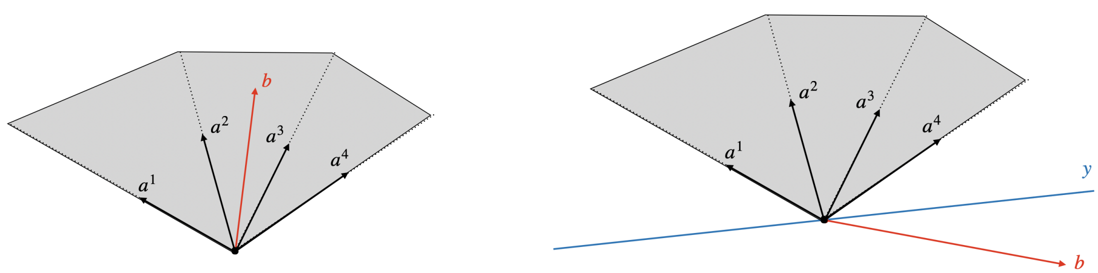

Figure 1: An illustration of Farkas Lemma in the plane, where columns of A are a^1, a^2, a^3, a^4. The figure on the left corresponds to the case when b belongs to the convex cone of A and the figure on the right is the case when b is not in the convex cone and can be separated from A via a hyperplane y.

While the above form of Farkas Lemma is perhaps most intuitive (from a geometric point of view at least), for our proof of Theorem 6, we need the following alternative but equivalent form.

Proposition 10 (Farkas Lemma – Alternative Form)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. $Ax \leqslant b$ has a non-negative solution $x \geqslant 0$* if and only if *for every $y \geqslant 0$ such that $y^T A \geqslant 0$, we also have $y^T b \geqslant 0$.*

Proof (assuming Proposition 9). Suppose $Ax \leqslant b$ has a solution $x \geqslant 0$. We are going to use Proposition 9 to show that the for all $y \geqslant 0$ such that $y^T A \geqslant 0$, there is $y^T b \geqslant 0$. Note that if $Ax \leqslant b$, we can always define slack variables $s \geqslant 0$ such that

$$\underbrace{\begin{bmatrix} A & I_m \end{bmatrix}}_{\tilde{A}} \cdot \underbrace{\begin{bmatrix} x \\ s \end{bmatrix}}_{\tilde{x}} = b.$$

There, the system $\tilde{A} \cdot \tilde{x} = b$ has a non-negative solution $\tilde{x} \geqslant 0$. This corresponds to the condition 1 of Proposition 9. Thus, for any vector $\tilde{y}$, we have that if $\tilde{y}^T \cdot \tilde{A} \geqslant 0$, then $\tilde{y}^T \cdot \tilde{b} \geqslant 0$ also. If we have a non-negative $y \geqslant 0$ with $y^T \cdot A \geqslant 0$, then we also have that $y^T \cdot \tilde{A} \geqslant 0$ because non-negativity of y ensures that $y^T \cdot I_m \geqslant 0$ as well. But then we should have that $y^T \cdot b \geqslant 0$ as well, concluding the proof. ☐

[4]Such a result is typically called a *Separation Theorem*. A prototypical example of such results is the following: for any compact convex body and any point, either the point belongs to the convex body or can be separated from it by a hyperplane.

4.2 Strong Duality From Farkas Lemma

We are now ready to prove Theorem 6.

Proof. By symmetry (as dual of the dual is primal), we are going to assume that the primal (P) is feasible and bounded. As we proved in Lecture 2 (in Theorem 9), this implies that (P) has an optimal solution x^*. Let $opt := c^T x^*$ denote the optimal value of (P) and $\varepsilon > 0$ be any arbitrary small parameter. Consider the following two polyhedra:

$$\text{(PH1)} \qquad\qquad\qquad \text{(PH2)}$$

$$\begin{aligned} c^T x &\geqslant opt \\ Ax &\leqslant b \\ x &\geqslant 0. \end{aligned} \qquad\qquad \begin{aligned} c^T x &\geqslant opt + \varepsilon \\ Ax &\leqslant b \\ x &\geqslant 0. \end{aligned}$$

We have that (PH1) is feasible as x^* belongs to this polyhedron, and (PH2) is not feasible as otherwise optimal value of (P) can be increased to $opt + \varepsilon$.

Define the following matrix and vectors:

$$\bar{A} = \begin{bmatrix} A \\ -c^T \end{bmatrix}, \qquad \bar{b} = \begin{bmatrix} b \\ -\text{OPT} \end{bmatrix} \qquad \bar{b}_\varepsilon = \begin{bmatrix} b \\ -\text{OPT} - \varepsilon \end{bmatrix}.$$

We thus have

$$\text{(PH1)} = \left\{\bar{A}x \leqslant \bar{b}, x \geqslant 0\right\} \qquad \text{and} \qquad \text{(PH2)} = \left\{\bar{A}x \leqslant \bar{b}_\varepsilon, x \geqslant 0\right\}$$

Now since (PH2) is infeasible, by Proposition 10, there is a vector $\bar{y} = [y\ z]$ for $y \in \mathbb{R}^m$ and $z \in \mathbb{R}$ such that

$$\bar{y} \geqslant 0 \qquad , \qquad \bar{y}^T \bar{A} \geqslant 0 \qquad \text{and} \qquad \bar{y}^T \bar{b}_\varepsilon < 0.$$

This implies that

$$y^T \cdot A - z \cdot c^T \geqslant 0 \qquad \text{and} \qquad y^T b - z \cdot (\text{opt} + \varepsilon) < 0.$$

Let us **assume** for now that $z > 0$ (we will lift this assumption at the end of the proof). Define the vector $w = \frac{1}{z} \cdot y \in \mathbb{R}^m$. We have that $w \geqslant 0$ and

$$A^T \cdot w \geqslant c \qquad \text{and} \qquad b^T \cdot w < \text{opt} + \varepsilon.$$

This implies that w is a feasible solution for the dual (D). As the dual is not unbounded (since (P) is feasible) and is also feasible, we get that it also has an optimal solution. Moreover, we showed that for every $\varepsilon > 0$, the objective value of the dual is less than $\text{opt} + \varepsilon$. This implies that the value of the dual can only be opt as well, implying that both primal and dual have the same objective value. It thus only remains to remove the assumption on $z > 0$.

Lifting the assumption on $z > 0$. We are going to show that in fact $z > 0$ is always going to hold and so our assumption was without a loss of generality. Consider the vector $\bar{y}$ again. Since (PH1) is feasible, by Proposition 10 applied to the vector $\bar{y}$, we have that

$$y^T \cdot A \geqslant c^T \qquad \text{and} \qquad y^T b \geqslant z \cdot \text{opt}.$$

Thus, we need the following to hold:

$$y^T b \geqslant z \cdot \text{opt} \qquad \text{but} \qquad y^T b < z \cdot \text{opt} + z \cdot \varepsilon \quad \text{for all } \varepsilon > 0.$$

This can only happen when $z > 0$, concluding the proof. ∎

References

[1] J. Farkas. Theorie der einfachen ungleichungen. *Journal für die reine und angewandte Mathematik*, 124:1–27, 1902. 10