Optimization October 2, 2024

Lecture 4: Duality

Dan A. Iancu

1 Preliminaries

In this class, we will discuss duality theory, one of the most important concepts in optimization. Duality theory will help us construct bounds on the optimal values of optimization problems and will provide optimality certificates (i.e., help us know that a specific solution is optimal) as well as optimality gaps (i.e., know how suboptimal a given solution is). Moreover, duality will also provide alternative algorithms to solve optimization problems and will lead to important applications in economics, finance, and engineering, which we will discuss more amply in our next lecture.

Note. Our discussion here is inspired to a large extent by Chapter 4 in the Bertsimas & Tsitsiklis book, but we adjusted several proofs to make them self-contained and emphasize more general concepts that will be useful beyond ***linear*** *optimization.*

1.1 Notation

For today's class, we will try to be very consistent with mathematical notation. For a matrix $A \in \mathbb{R}^{m\times n}$, we use A_j to denote the j-th column, A_S to denote the submatrix obtained by retaining the columns $j \in S$, and $a_i^\intercal$ to denote the i-th row. For a vector $x \in \mathbb{R}^n$, we can then view the expression Ax either as a linear combination of the columns A_j or as having components corresponding to the inner products $a_i^\intercal x$, i.e.,

$$Ax = \sum_{j=1}^{n} A_j x_j = \begin{bmatrix} a_1^\intercal x \\ a_2^\intercal x \\ \vdots \\ a_m^\intercal x \end{bmatrix}.$$

We also let $\|\cdot\|$ be the Euclidean norm defined by $\|x\| = (x^\intercal x)^{1/2}$.

1.2 Motivation

To start, let us consider a linear optimization problem in the most general form possible:

$$\begin{array}{lll} (\mathscr{P})\ \text{minimize}_x & c^\intercal x & \\ & a_i^\intercal x \geq b_i, & i \in M_1, \\ & a_i^\intercal x \leq b_i, & i \in M_2, \\ & a_i^\intercal x = b_i, & i \in M_3, \\ & x_j \geq 0, & j \in N_1, \\ & x_j \leq 0, & j \in N_2, \\ & x_j \text{ free}, & j \in N_3. \end{array} \tag{1}$$

which we henceforth call the **primal problem** and concisely refer to as problem $(\mathscr{P})$. We also denote its feasible set with P and we let x^* be an optimal solution, assumed to exist.

Because $(\mathscr{P})$ is a minimization, we are interested in constructing **lower bounds** on its optimal value. One thought is to simply **remove** some constraints! Although that would lead to a lower bound, it might lose too much information from the problem and give us poor lower bounds, such as $-\infty$. A better approach is to **relax** some of the constraints – specifically, we should remove the constraints and instead add them in the objective, with a suitable penalty. To that end, let us consider a relaxed problem in which we associate with every constraint $i \in M_1 \cup M_2 \cup M_3$ a **price** or **penalty** p_i that should penalize us when that constraint is violated. The objective in this relaxed problem, which is referred to as the **Lagrangean** function, can be written as follows:

$$\mathcal{L}(x,p) = c^\intercal x - \sum_{i \in M_1 \cup M_2 \cup M_3} p_i^\intercal (a_i^\intercal x - b_i) = p^\intercal b + (c^\intercal - p^\intercal A)x. \tag{2}$$

We note that the choice of $-$ sign in front of the penalty and the choice to write the penalty as $p^\intercal(Ax-b)$ rather than $p^\intercal(b-Ax)$ is quite arbitrary. Our choice above has two advantages: (i) it gives us the "nice" term $p^\intercal b$ (rather than $-p^\intercal b$) in the expression for $\mathcal{L}$, and (ii) it makes it easy to figure out what **requirements we need to impose on p to ensure that it a valid penalty**, or equivalently, that $\mathcal{L}(x,p)$ leads to a valid relaxation. To that end, note that we must impose the following constraints on p:

$$\begin{array}{llll} \forall\, i \in M_1, & a_i^\intercal x - b_i \geq 0 \;\text{ for } x \text{ feasible in } (\mathscr{P}) & \Rightarrow & p_i \geq 0 \\ \forall\, i \in M_2, & a_i^\intercal x - b_i \leq 0 \;\text{ for } x \text{ feasible in } (\mathscr{P}) & \Rightarrow & p_i \leq 0 \\ \forall\, i \in M_3, & a_i^\intercal x - b_i = 0 \;\text{ for } x \text{ feasible in } (\mathscr{P}) & \Rightarrow & p_i \text{ free.} \end{array} \tag{3}$$

Clearly, any p satisfying these leads to a valid lower bound on the primal objective:

$$p \text{ satisfying (3)} \;\Rightarrow\; \mathcal{L}(x,p) \leq c^\intercal x,\; \forall\, x \in P. \tag{4}$$

This allows us to define a lower bound on the optimal objective of the primal by considering the function $g(p)$ defined as:

$$\begin{aligned} g(p) := \min\nolimits_x \;& \mathcal{L}(x,p) \\ \text{s.t. } & x_j \geq 0,\; j \in N_1, \\ & x_j \leq 0,\; j \in N_2, \\ & x_j \text{ free},\; j \in N_3. \end{aligned} \tag{5}$$

Because $\mathcal{L}(x,p) \leq c^\intercal x$, $\forall x \in P$ by (4) and problem (5) has fewer constraints than $(\mathscr{P})$, we can immediately infer that $g(p)$ is a valid lower bound on the optimal primal cost:

$$g(p) \leq c^\intercal x^*, \text{ for any } p \text{ satisfying (3).}$$

Moreover, because we obtain a valid bound for any price p, we might as well look for the *best* such lower bound, which leads us to consider the problem:

$$\text{maximize}_p \,\{g(p) \,:\, p \text{ satisfying (3).}\} \tag{6}$$

Problem (6) is called the **dual** of the primal problem ($\mathscr{P}$); for conciseness, we also refer to it as problem ($\mathscr{D}$). Let us try to rewrite ($\mathscr{D}$) to make it clear that it is also a linear optimization problem. First, we rewrite the objective. Note that we have:

$$
\begin{aligned}
g(p) := \min_x \; & \left[p^\intercal b + (c^\intercal - p^\intercal A)x\right] \\
\text{s.t. } & x_j \geq 0,\; j \in N_1, \\
& x_j \leq 0,\; j \in N_2, \\
& x_j \text{ free},\; j \in N_3.
\end{aligned}
$$

$$
= \begin{cases} p^\intercal b, & \text{if } \operatorname{sign}(c_j - p^\intercal A_j) = \operatorname{sign}(x_j),\, \forall\, j \in N_1 \cup N_2 \\ -\infty, & \text{otherwise.} \end{cases}
$$

Because we are interested in maximizing $g(p)$, we can restrict attention to those values of p for which $g(p) > -\infty$. Therefore, the dual is equivalent to the linear programming problem:

$$
\begin{array}{rlll}
(\mathscr{D}) \text{ maximize} & p^\intercal b & & \\
\text{subject to} & p_i \geq 0, & i \in M_1, & \\
& p_i \leq 0, & i \in M_2, & \\
& p_i \text{ free}, & i \in M_3, & \\
& p^\intercal A_j \leq c_j, & j \in N_1, & \\
& p^\intercal A_j \geq c_j, & j \in N_2, & \\
& p^\intercal A_j = c_j, & j \in N_3. &
\end{array} \tag{7}
$$

Putting everything together, we obtain the following primal-dual pair of problems:

$$
\begin{array}{rlll}
(\mathscr{P}) \text{ minimize} & c^\intercal x & \\
(p_i \rightarrow) & a_i^\intercal x \geq b_i, & i \in M_1, \\
(p_i \rightarrow) & a_i^\intercal x \leq b_i, & i \in M_2, \\
(p_i \rightarrow) & a_i^\intercal x = b_i, & i \in M_3, \\
& x_j \geq 0, & j \in N_1, \\
& x_j \leq 0, & j \in N_2, \\
& x_j \text{ free}, & j \in N_3.
\end{array}
\qquad
\begin{array}{rlll}
(\mathscr{D}) \text{ maximize} & p^\intercal b & \\
\text{subject to} & p_i \geq 0, & i \in M_1, \\
& p_i \leq 0, & i \in M_2, \\
& p_i \text{ free}, & i \in M_3, \\
(x_j \rightarrow) & p^\intercal A_j \leq c_j, & j \in N_1, \\
(x_j \rightarrow) & p^\intercal A_j \geq c_j, & j \in N_2, \\
(x_j \rightarrow) & p^\intercal A_j = c_j, & j \in N_3.
\end{array} \tag{8}
$$

There are simple mnemonic rules to help you memorize this primal-dual formulation so that you can avoid going through all the steps above with the Lagrangean each time. Specifically, note that we introduce a dual decision variable p_i for every constraint in the primal except the sign constraints; so every constraint $i \in M_1 \cup M_2 \cup M_3$ has a corresponding dual variable indicated by the symbol $p_i \rightarrow$ on the left of the constraint. Symmetrically, for every decision variable x_j in the primal with $j \in N_1 \cup N_2 \cup N_3$, there is a constraint in the dual (and the mapping is indicated by the symbol $x_i \rightarrow$ on the left of the dual constraints). As for the signs, the following table summarizes all the cases that can arise:

PRIMAL	minimize	maximize	DUAL
constraints	$\geq b_i$ $\leq b_i$ $= b_i$	≥ 0 ≤ 0 free	**variables**
variables	≥ 0 ≤ 0 free	$\leq c_j$ $\geq c_j$ $= c_j$	**constraints**

Note. There are intuitive rules to derive the signs in the table above. To understand what sign you need for the dual variable p_i, think of it as a shadow price that records the marginal change in the primal optimal objective value when the right-hand side b_i of the primal constraint is changing infinitesimally. Increasing the right-hand-side in a "$\geq b_i$" constraint would (weakly) shrink the feasible set and therefore **reduce** *the objective (because the primal is a minimization), hence the positive shadow price $p_i \geq 0$ for a primal constraint "$\geq b_i$". Similarly, increasing the right-hand-side in a "$\leq b_i$" constraint would (weakly) enlarge the feasible set and therefore* decrease *the objective (in our primal minimization), hence the negative shadow price $p_i \leq 0$ for the primal constraint "$\leq b_i$". These rules will become more clear once we formalize this interpretation of dual variables as gradients of the primal objective value with respect to the right-hand-side vector b.*

The main result in duality theory asserts that when the primal ($\mathscr{P}$) admits an optimal value, it will be equal to the optimal value of the dual problem ($\mathscr{D}$). And this also implies that a choice of penalties/prices p exists so that the relaxed problem (RP) with penalty p has exactly the same optimal value as the primal ($\mathscr{P}$). We will prove this result in the subsequent sections.

The following implication (from the table above and our earlier derivation) will be useful subsequently:

$$\forall\, x \in P, \forall\, p \in D \quad : \quad \text{sign}(a_i^\intercal x - b_i) = \text{sign}(p_i), \quad \text{sign}(x_j) = \text{sign}(c_j - p^\intercal A_j). \tag{9}$$

1.3 Duals of Equivalent Primals and Duals of Duals

It is a rather tedious exercise, but it can be readily checked that for linear optimization problems, the following result holds.

Theorem 1 *If we transform a primal linear optimization problem P_1 into an equivalent formulation P_2 by transformations such as*

- *replacing a free variable with a difference of two non-negative variables, $x_i = x_i^+ - x_i^-$;*
- *replacing an inequality constraint with an equality constraint by introducing a slack variable;*
- *for a* feasible *LP in standard form, removing any rows $a_i^\intercal$ that are linearly dependent on other rows,*

then the duals of (P_1) and (P_2) are equivalent, i.e., they are either both infeasible or they have the same optimal objective.

The proof involves simple algebra and is not very enlightening, so we omit it. The result should be consistent with the intuition that the precise formulation of the primal should bear no impact on its optimal value, so the duals of equivalent primal formulations should also be equivalent.

The following result is slightly more subtle, and concerns a natural question: *"what if we formed the dual of a dual? would we recover the primal?"* For linear optimization, the answer is "yes."

Theorem 2 (The dual of the dual is the primal) *If we transform the dual into an equivalent minimization problem and then form its dual, we obtain a problem equivalent to the original primal optimization problem.*

We leave the proof to the reader. An important word of caution here is that this result is **not** true more generally. It does hold (with qualifiers) for the broader class of convex optimization problems, but it does not hold for non-convex optimization problems.

2 Weak Duality

We already argued at the start of this section that for a primal ($\mathscr{P}$) in standard form, the cost $g(p)$ of any dual solution p provides a lower bound on the optimal primal objective. The following result is a slightly more general restatement.

Theorem 3 (Weak Duality) *Consider any primal-dual pair in the general form* (8). *If x is feasible for the primal and p is feasible for the dual, then:*

$$p^\intercal b \leq c^\intercal x.$$

Proof: For any x and p, we define:

$$u_i = p_i(a_i^\intercal x - b_i),$$
$$v_j = (c_j - p^\intercal A_j)x_j.$$

Recall from (8) that for any feasible $x \in P$ and $p \in D$:

$$\text{sign}(a_i^\intercal x - b_i) = \text{sign}(p_i), \quad \text{sign}(x_j) = \text{sign}(c_j - p^\intercal A_j). \tag{10}$$

Therefore, the sign of p_i equals the sign of $a_i^\intercal x - b_i$ and the sign of $c_j - p^\intercal A_j$ equals the sign of x_j, and therefore: $u_i \geq 0$, $v_j \geq 0$. Also, we have:

$$\sum_i u_i = p^\intercal Ax - p^\intercal b, \qquad \sum_j v_j = c^\intercal x - p^\intercal Ax.$$

Add these equalities and using the non-negativity of u_i and v_j then proves the result. $\square$

As the name suggests, weak duality is not a powerful result and it will hold for many optimization problems, including non-convex ones. In our context, it has the following immediate corollaries.

Corollary 1 *The following results hold:*
(a) If the optimal cost in the primal is $-\infty$*, then the dual problem must be infeasible.*
(b) If the optimal cost in the dual is $+\infty$*, then the primal problem must be infeasible.*
(c) If x *is primal-feasible and* p *is dual-feasible and* $p^\intercal b = c^\intercal x$ *holds, then* x *and* p *are optimal solutions to the primal and dual problems, respectively.*

A practical implication of these results is worth pointing out: **weak duality enables us to assess the degree of suboptimality for a given solution**. More specifically, suppose we have a primal-feasible solution x. Then, any dual-feasible solution p will lead to a suboptimality guarantee for x, because the optimal solution x^* for $(\mathscr{P})$ must satisfy:

$$c^\intercal x \geq c^\intercal x^* \geq p^\intercal b.$$

Therefore, if the gap $c^\intercal x - p^\intercal b$ is small, we may be satisfied with the current solution x and not need to worry about finding *the* optimum! However, weak duality cannot guarantee that such optimality gaps become small, so these may not be practically meaningful.

Similarly, Part (c) provides a (weak) form of optimality certificate: it states that if we can produce two solutions x and p satisfying these conditions, we are guaranteed that these are optimal solutions for the primal and the dual, respectively. However, weak duality **does not guarantee that such a pair of x and p even exists!**

3 Strong Duality

We would now like to prove a more powerful result, referred to as **strong duality**: if the primal and dual are both feasible – which, by Corollary 1, implies they both admit optimal solutions – then they will have the same optimal values. There are several proofs possible for this result. We adopt an approach here that is significantly more general and involves results that will be useful later in the course, when we discuss convex optimization. Chapter 4 of the Bertsimas & Tsitsiklis book also has an alternative proof that relies on the iterations in the simplex algorithm (and is therefore tailored to linear optimization problems).

3.1 A Few Results from Real Analysis

Our proof requires a few basic facts from analysis. First, recall the definition of a closed set. A set $S \subset \mathbb{R}^n$ is called **closed** if it has the following property: if $x_1, x_2, \ldots$ is a sequence of elements of S that converges to some $x \in \mathbb{R}^n$, then $x \in S$. That is, S contains the limit of any sequence of elements of S.

The first result we need is that any polyhedron is a closed set.

Theorem 4 *Every polyhedron is closed.*

Proof: Consider a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ (the representation does not matter because the set of points is the same; so we adopt a representation with inequalities without loss of generality). Suppose that $x_1, x_2, \ldots$ is a sequence of elements of P that converges to some x^*. For each k, we have $x_k \in P$, and therefore, $Ax_k \geq b$. Taking the limit, we obtain $Ax^* = A\left(\lim_{k\to\infty} x_k\right) = \lim_{k\to\infty} Ax_k \geq b$, so x^* belongs to P. $\square$

It is important to note that this is **not true for any convex set**! For instance, consider a circle with a full interior and remove one point from its boundary. Then, the remaining points will form a convex set that is not closed.

The following result – which we state without proof – is fundamental in real analysis. It states that any continuous function achieves its minimum and maximum value on a nonempty, compact (i.e., closed and bounded) set of points.

Theorem 5 (Weierstrass' Theorem) *If $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function, and if S is a nonempty, closed, and bounded subset of $\mathbb{R}^n$, then there exists some $\underline{x} \in S$ such that $f(\underline{x}) \leq f(x)$ for all $x \in S$ and there exists some $\bar{x} \in S$ such that $f(\bar{x}) \geq f(x)$ for all $x \in S$.*

This result is not valid if the set S is not closed. A classic example in the half-line $S = \{x \in \mathbb{R} \mid x > 0\}$, for which the problem of minimizing x does not achieve its minimum. The reason the set is not closed is because we used a strict inequality to define it. The definition of polyhedra and linear programming problems does not allow for strict inequalities in order to avoid precisely situations of this type.

3.2 The Separating Hyperplane Theorem

The first step in our proof is to show that if a point x^* lies outside a polyhedron P, then there exists a hyperplane that strictly separates x from P, i.e., there exists a vector c such that $c^\intercal x^* < c^\intercal x$ for all $x \in P$. This has a very clear geometric intuition, as depicted in the left panel of Figure 1. On first sight, you may think the fact is obvious, but it is actually an important result in linear programming. Instead of showing this, we prove a more general result that concerns the separation of two convex sets.

Theorem 6 (Separating Hyperplane Theorem for Convex Sets) *Let S and U be two nonempty, closed, convex subsets of $\mathbb{R}^n$ such that $S \cap U = \emptyset$ and S is bounded. Then, there exists a vector $c \in \mathbb{R}^n$ such that $c^\intercal x < c^\intercal y$ holds for all $x \in S$ and $y \in U$.*

The geometric intuition for this case (and for the proof) appears in the right panel of Figure 1. Our result will obviously follow as a special case with $S = \{x^*\}$ and $U = P$.

Proof: Consider the following optimization problem:

$$\begin{aligned} &\text{infimum } \|x - y\| \\ &\text{such that } x \in S,\, y \in U. \end{aligned} \tag{11}$$

We claim that the infimum is achieved, i.e., $\exists (x^*, y^*) \in S \times U$ such that $\|x^* - y^*\| \leq \|x - y\|$ for any $(x, y) \in S \times U$. To see this, we will invoke the Weierstrass Theorem. The theorem does not immediately apply because the set U is not required to be bounded. But we will try to apply the theorem to the following function $f : S \to \mathbb{R}$:

$$f(x) := \inf_{y \in U} \|x - y\|.$$

Intuitively, $f(x)$ is the shortest distance from x to U. If f were continuous, then the Weierstrass Theorem would be readily applicable because the domain of f is the compact,

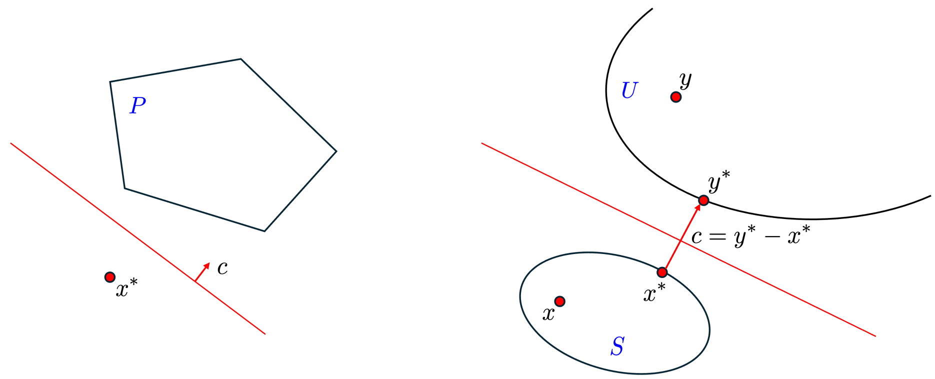

Figure 1: Separating Hyperplane Theorem. Left: separating a point from a polyhedral set. Right: more general result on separating two convex sets.

convex set S. So all we need is to argue that $f(x)$ is continuous. To that end, consider any $x, x' \in S$ and $y \in U$ and the following inequalities derived from the triangle inequality:

$$\|x - y\| \leq \|x - x'\| + \|x' - y\|$$
$$\|x' - y\| \leq \|x' - x\| + \|x - y\|.$$

These imply that

$$|(\|x - y\| - \|x' - y\|)| \leq \|x - x'\|$$

and therefore

$$|f(x) - f(x')| \leq \|x - x'\|,$$

which proves that f is continuous and that the minimum is achieved in (11). Let (x^*, y^*) denote an optimal solution in that problem. *(Q: Is such a solution guaranteed to be unique? Does that matter?)*

We will show that the vector $c := y^* - x^*$ gives the separating hyperplane. Specifically, we prove that for any $x \in S$ and any $y \in U$:

$$c^\intercal x \leq c^\intercal x^* < c^\intercal y^* \leq c^\intercal y.$$

See Figure 1 for the corresponding geometric intuition.

First, observe that:

$$c^\intercal y^* - c^\intercal x^* = (y^* - x^*)^\intercal (y^* - x^*) > 0,$$

where the inequality holds because the optimal value in problem (11) must be strictly positive because S and U are closed and have non-empty intersection.

We next argue that $c^\intercal y^* \leq c^\intercal y$ holds for any $y \in U$. Consider any $y \in U$. For any $\lambda \in (0, 1]$, we have that $y^* + \lambda(y - y^*) \in U$ because U is convex. Because y^* minimizes $\|y - x^*\|$ over all $y \in U$, we have:

$$\begin{aligned}\|y^* - x^*\|^2 &\leq \|y^* + \lambda(y - y^*) - x^*\|^2 \\ &= \|y^* - x^*\|^2 + 2\lambda(y^* - x^*)^\intercal (y - y^*) + \lambda^2 \|y - y^*\|^2\end{aligned}$$

which implies that

$$2\lambda(y^* - x^*)^\intercal(y - y^*) + \lambda^2\|y - y^*\|^2 \geq 0$$

Dividing by λ and taking the limit as λ approaches zero, we obtain $c^\intercal(y - y^*) \geq 0$. The proof that $c^\intercal x \leq c^\intercal x^*$ is analogous. $\square$

The following result, which is necessary for our purposes, is an immediate corollary.

Corollary 2 *If P is a polyhedron and x^* satisfies $x \notin P$, there exists a hyperplane that strictly separates x from P, i.e., there exists $c \neq 0$ such that $c^\intercal x^* < c^\intercal x$ for all $x \in P$.*

Note. The strict separation result above is much stronger than what we needed here, but it is very enlightening to see the proof one time and understand its inner workings. (Moreover, the proof is not significantly harder than a direct proof of Corollary 2, and the generalization is substantial and will be useful later in our course, when we discuss convex optimization.)

3.3 Farkas Lemma

We are now equipped to prove the building block that will provide us with certificates of feasibility and optimality and will lead to a quick proof of strong duality. This result is named after Gyula Farkas, a Hungarian mathematician, and has played a pivotal role in the development of mathematical optimization (and it even has interesting connections to quantum mechanics!)

Theorem 7 (Farkas' Lemma) *Consider $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then, exactly one of the following two alternatives holds:*

(a) There exists some $x \geq 0$ such that $Ax = b$.

(b) There exists some vector p such that $p^T A \geq 0$ and $p^T b < 0$.

Before proving this result, let's develop a bit of geometric intuition. Figure 2 depicts the two alternatives: on the left, the vector b belongs to the cone generated by the columns A_i of the matrix A, so there exists $x \geq 0$ such that $Ax = b$. In contrast, on the right, the vector b does not belong to the cone generated by the columns of A, so a separating hyperplane given by the normal vector p exists.
Proof: **"(a) $\Rightarrow$ not (b)."** This direction is easy. If there exists some $x \geq 0$ satisfying $Ax = b$ and if we have p such that $p^T A \geq 0$, then $p^T b = p^T Ax \geq 0$, so (b) cannot hold.
"not (a) $\Rightarrow$ (b)." This is the more subtle direction, but the separating hyperplane theorem will make our life easy. Assume that there exists no vector $x \geq 0$ satisfying $Ax = b$. This implies that $b \notin S$ where the set S is defined as

$$S := \{Ax : x \geq 0\} = \{y : \exists\, x \geq 0 \,\text{such that}\, y = Ax\}.$$

The set S is clearly convex. However, to apply the separating hyperplane theorem, we must show that it is also **closed**. The set S is the projection of the polyhedral set

$$\bar{S} := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \;\; x \geq 0,\, y = Ax$$

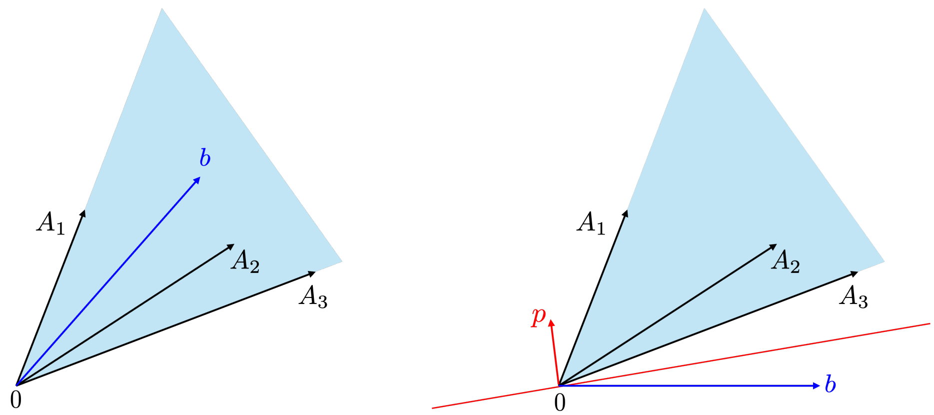

Figure 2: The two alternatives possible in the Farkas Lemma. Left: the vector b belongs to the cone generated by the columns A_i of the matrix A, so there exists $x \geq 0$ such that $Ax = b$. Right: the vector b does not belong to the cone generated by the columns of A, so a separating hyperplane exists.

on the last m coordinates. Because the projection of a polyhedral set on a subset of coordinates is another polyhedral set[1] and because every polyhedral set is closed, we can indeed apply Theorem 6 to conclude that there must exist a vector p such that $p^\intercal b < p^\intercal y$ for every $y \in S$. Because $0 \in S$, we must have $p^\intercal b < 0$. Moreover, because every column A_i of A satisfies $\lambda A_i \in S$ for every $\lambda > 0$, we have

$$\frac{p^\intercal b}{\lambda} < p^\intercal A_i, \forall \lambda > 0,$$

and taking the limit as $\lambda \to \infty$ we see that it must be the case that $p^\intercal A_i \geq 0$. We conclude that there exists p such that $p^\intercal A \geq 0$ and $p^\intercal b < 0$, which completes the proof. $\square$

To appreciate the power of the Farkas Lemma, note that its statement provides an immediate certificate of infeasibility for the primal problem. Recall that in our original (standard form) primal problem, we are interested in points x satisfies $Ax = b, x \geq 0$. The Farkas Lemma essentially states that either the primal problem is feasible or there exists a vector p satisfying alternative (b). Therefore, p constitutes a **certificate of infeasibility**: if we have such a p, we know for a fact that the primal problem is **infeasible**.

3.4 The Strong Duality Theorem

We are now ready to derive our main result – the strong duality theorem – as a direct corollary of the Farkas Lemma. Without loss of generality, we prove this for a primal problem with constraints in inequality form, $Ax \geq b$. (This is without loss because the optimal solution in an optimization problem is the same irrespective of the representation of the feasible set and any polyhedron admits an inequality representation like the one we

[1] This result can be shown in several ways, including via the Fourier-Motzkin procedure for eliminating variables in a linear program; see the Bertsimas and Tsitsiklis book for details on this.

consider.) So we consider here the following pair of primal and dual problems:

$$\begin{array}{ll} \text{Primal Problem } (P_1): & \text{Dual Problem } (D_1): \\ \text{minimize } c^\intercal x & \text{maximize } p^\intercal b \\ \text{subject to } Ax \geq b, & \text{subject to } p^\intercal A = c^T,\ p \geq 0. \end{array} \tag{12}$$

Theorem 8 *If a primal linear programming problem has an optimal solution, so does its dual, and the respective optimal values are equal.*

Proof: Assume that the primal (P_1) in(12) has an optimal solution x^*. We prove that the dual problem admits a feasible solution p such that $p^\intercal b = c^\intercal x^*$.

Let $\mathcal{F} = \{i \mid a_i^\intercal x^* = b_i\}$ be the indices of active constraints at x^*. We claim that the cost vector c can be written as a conic combination of the active constraints $\{a_i : i \in \mathcal{F}$. (See Figure 3 for a visualization.)

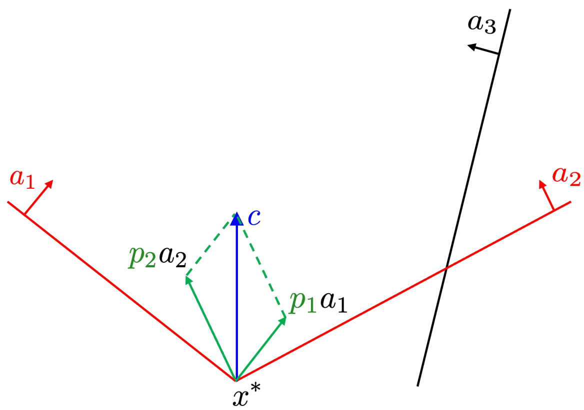

Figure 3: Interpretation of optimality conditions at x^*. The red constraints a_1, a_2 are active at x^*, whereas the black constraint a_3 is not active. The proof shows that in this case, the cost vector c can be generated as a conic combination of the active constraints a_1, a_2, with coefficients p_1 and p_2.

As a first step, we show that for any vector d, the following implication holds:

$$a_i^\intercal d \geq 0, \forall\, i \in \mathcal{F} \quad \Rightarrow \quad c^\intercal d \geq 0.$$

To see this, consider any d satisfying the premise on the left-hand-side. For a sufficiently small $\epsilon > 0$, we claim that the point $x^* + \epsilon d$ is feasible for P. We have that $a_i^\intercal(x^* + \epsilon d) \geq b_i, \forall i \in \mathcal{F}$; moreover, because $a_i^\intercal x^* > b_i$ for all the constraints $i \notin \mathcal{F}$, we will have that $a_i^\intercal(x^* + \epsilon d) \geq b_i$ also holds for $i \notin \mathcal{F}$ provided that ϵ is sufficiently small. So $x^* + \epsilon d$ is feasible and moreover, $c^\intercal(x^* + \epsilon d) < c^\intercal x^*$, which contradicts the optimality of x^*.

This implies that we cannot find any vector d such that $a_i^\intercal d \geq 0, \forall\, i \in \mathcal{F}$ and $c^\intercal d < 0$. In the context of the Farkas Lemma (Theorem 7), this means alternative (b) is not true, so alternative (a) must be true: c can be expressed as a nonnegative linear combination of the vectors $a_i, i \in \mathcal{F}$. That is, there exist nonnegative scalars $p_i, i \in \mathcal{F}$, such that:

$$c = \sum_{i \in \mathcal{F}} p_i a_i.$$

Letting $p_i = 0$ for $i \notin \mathcal{F}$, we conclude that $\exists p \geq 0$ feasible for the dual $(\mathscr{D})$. Moreover,

$$p^\intercal b = \sum_{i \in \mathcal{F}} p_i b_i = \sum_{i \in \mathcal{F}} p_i a_i^\intercal x^* = c^\intercal x^*,$$

which shows that the objective of the dual $(\mathscr{D})$ under the feasible solution p is the same as the optimal primal objective. The strong duality result follows from Corollary 1. $\square$

3.5 Implications

In view of the strong duality result, we can see that the only possibilities for a primal-dual pair are summarized in the following table:

	Finite Optimum	Unbounded	Infeasible
Finite Optimum	Possible	Impossible	Impossible
Unbounded	Impossible	Impossible	Possible
Infeasible	Impossible	Possible	Possible

Table 1: The different possibilities for the primal and the dual.

The result that is perhaps not immediately obvious is that both the primal and the dual may be infeasible. This can be seen with the following two-dimensional example:

Example.[Infeasible primal-dual pair] Consider the infeasible primal problem:

$$\text{minimize } x_1 + 2x_2 \quad \text{subject to } x_1 + x_2 = 1, \quad 2x_1 + 2x_2 = 3.$$

Its dual is:

$$\text{maximize } p_1 + 3p_2 \quad \text{subject to } p_1 + 2p_2 = 1, \quad p_1 + 2p_2 = 2,$$

which is also infeasible.

Strong duality is very powerful because it provides certificates of (in)feasibility and optimality. We already knew from weak duality that any dual-feasible solution p provides a bound on the cost of a feasible primal solution x. However, strong duality certifies that such bounds are actually **really good**: in fact, the best bound (corresponding to the dual-optimal solution p^*) will provide a certificate of **optimality** for any primal problem.

3.5.1 Optimality Conditions and Complementary Slackness

Sometimes all we want is to **characterize** the optimal solutions to a problem (this is often the case in economic modeling, where we want to show they possess certain properties). We will discuss such optimality conditions more extensively in nonlinear optimization problems, so it is important to appreciate them in the simplest possible case, namely for linear optimization. The following theorem states **necessary and sufficient optimality conditions** in the context of the general primal-dual problem considered in (8).

Theorem 9 (Complementary Slackness) *Consider the primal-dual pair in* (8)*. Let x and p be feasible solutions to the primal and dual problem, respectively. Then x and p are optimal solutions for the primal and the dual* ***if and only if****:*

$$p_i(a_i^\intercal x - b_i) = 0,\ \forall i$$
$$(c_j - p^\intercal A_j)x_j = 0,\ \forall j.$$

Proof: Recall the definitions $u_i = p_i(a_i^T x - b_i)$ and $v_j = (c_j - p^T A_j)x_j$ in the proof of the weak duality result in Theorem 3. We noted that for x primal feasible and p dual feasible, we have $u_i \geq 0$ and $v_j \geq 0$ for all i and j and that:

$$c^T x - p^T b = \sum_i u_i + \sum_j v_j.$$

By the strong duality theorem, if x and p are optimal, then $c^T x = p^T b$, which implies that $u_i = v_j = 0$ for all i, j. Conversely, if $u_i = v_j = 0$ for all i, j, then $c^T x = p^T b$, which implies that x and p are optimal. $\square$

Intuitively, the first set of optimality conditions are always satisfied if the primal ($\mathscr{P}$) is in standard form. If the primal has a constraint like $a_i^\intercal x \geq b_i$, the complementary slackness condition implies that if $a_i^\intercal x > b_i$ (so the constraint is not active), then $p_i = 0$. Put differently, constraints that are not active are "uninteresting" and have zero price: these can be removed from the primal (and the corresponding dual variable can also be removed) without affecting optimality.

In some cases – typically, for smaller-scale problems or more stylized models – these optimality conditions can actually be solved analytically to recover the optimal solutions for the primal and the dual. However, this is almost never the most efficient approach to solving the problems!

3.6 Computational Implications

We conclude by remarking that strong duality also provides us with an alternative approach towards solving a linear optimization problem: we could solve either the primal or the dual! The natural question that arises is when is one better than the other – and more specifically, restriction attention to the simplex method, when should one use the primal simplex method (to solve the primal) or use the dual simplex method?

To appreciate this, consider again a primal in standard form and recall from our previous class that the simplex algorithm works with basic feasible solutions. Let $B \subseteq \{1, \ldots, n\}$ denote a basis and A_B the submatrix of A with columns from B. The basic feasible solution (for the primal) can be determined as

$$x_B = A_B^{-1} b,$$

and the same basis can also be used to determine a dual vector p through the equations:

$$p^\intercal A_i = c_i,\ \forall\, i \in B.$$

Because A_B is invertible, the system has a unique solution, which can be written as $p^\intercal = c_B(A_B)^{-1}$. This also means that there are exactly m linearly independent active dual constraints (among the n constraints $p^\intercal A \leq c^\intercal$). So if p were feasible, it would actually be a basic feasible solution for the dual problem!

This shows that a basis B is associated with a basic point (i.e., a point with at most $n - m$ nonzeros) for the primal and is also associated with a basic point for the dual. The **dual simplex** method can then be seen as iterating and swapping one binding constraint for another: the basic feasible solutions obtained by the dual simplex at consecutive iterations have $m - 1$ active inequality constraints in common, so these solutions are either adjacent or they coincide.

What is interesting – and makes the dual simplex different than the primal simplex – is that the dual simplex is essentially working entirely with a problem with inequality constraints, whereas the primal was tailored to problems in standard form.

Modern solvers such as Gurobi include both "primal simplex" and "dual simplex", and even allow concurrent methods that run in parallel and pick the solution from whichever process terminates first (with an optimality certificate). If you are curious, you can read more at this url: `https://www.gurobi.com/documentation/current/refman/method.html`.