Optimization Oct 2 - Oct 9, 2024

Lectures 4-5-6: Duality

Dan A. Iancu

1 Motivation

In this class we will discuss duality theory, one of the most important concepts in optimization. To appreciate its importance, consider a linear program (LP) in inequality form:

$$\begin{aligned} &\text{minimize } c^\intercal x \\ &\text{such that } Ax \leq b, \end{aligned}$$

and the following practical questions that you might be faced with:

1. Suppose we have a feasible x, how can we know **"how good"** it is? More formally, assuming the problem has an optimal solution x^* (resulting in a finite optimum cost $c^\intercal x^*$), how can we to quantify the gap $c^\intercal x - c^\intercal x^*$?

2. Suppose we do not yet have a feasible x (and we have been searching for a while...) How can we **certify** that no such feasible x exists, i.e., that $\{x : Ax \leq b\} = \emptyset$?

3. Suppose one of constraints in our problem depends on parameters that are uncertain. For instance, suppose the i-th constraint is $a_i^\intercal x \leq 0$ and the row vector of parameters a_i is only known to reside in a given set, i.e., $a_i \in \mathcal{A}$. This situation occurs often in practice, where optimization problems are affected by uncertainty. How can we ensure this constraint is feasible **for any** $a_i \in \mathcal{A}$? Note that if the set $\mathcal{A}$ contains infinitely many points (for instance, if it's a polyhedral set), this constraint would give rise to **infinitely many** constraints. Could we reformulate such a problem as a **finite-dimensional** optimization problem?

4. Suppose some components of the right-hand-side vector b correspond to some valuable resources. For instance b_i is a capacity in a plant that you are operating or is a monetary budget that constrains your investments. Clearly, more b_i would be nice, because it would enlarge your feasible set and (possibly) reduce your objective. If someone were to offer you a bit more of b_i, **what is a "suitable" price for** b_i, i.e., a price that you are willing to pay that makes the deal worthwhile?

Duality theory will help us answer all of these questions. With duality, we can construct bounds on the optimal values of optimization problems and provide **optimality gaps** (i.e., know how suboptimal a given solution is) and **optimality certificates** (i.e., guarantee when a given solution is optimal). Moreover, duality will provide **feasibility certificates**, i.e., it will allow us to know when a given optimization problem is infeasible. Lastly, duality will provide alternative algorithms to solve optimization problems and will lead to important applications in economics, finance, and engineering, which we will discuss more amply later.

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 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 Setup

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:

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

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:

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

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:

$$\begin{aligned} g(p) := \min\nolimits_x \;& \left[p^\intercal b + (c^\intercal - p^\intercal A)x\right] \\ \text{s.t. } & x_j \geq 0,\; j \in N_1, \\ & x_j \leq 0,\; j \in N_2, \\ & x_j \text{ free},\; j \in N_3. \\ = & \begin{cases} p^\intercal b, & \text{if } \operatorname{sign}(c_j - p^\intercal A_j) = \operatorname{sign}(x_j),\, \forall\, j \in N_1 \cup N_2 \text{ and } c_j = p^\intercal A_j,\, \forall j \in N_3 \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

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:

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

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

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

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**

(9)

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 $p_i \le 0$ for the primal constraint "$\le b_i$". 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.

Subsequently, we will use a notation that parallels the one we used for the primal $(\mathscr{P})$. Specifically, we will denote the feasible set for the dual problem $(\mathscr{D})$ with D; by (8), D is clearly a polyhedral set.

The following implication, which is immediate from table (9) and our earlier derivation, will be useful subsequently:

$$\forall\, x \in P, \forall\, p \in D \quad : \quad \operatorname{sign}(a_i^\intercal x - b_i) = \operatorname{sign}(p_i), \quad \operatorname{sign}(x_j) = \operatorname{sign}(c_j - p^\intercal A_j). \tag{10}$$

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 $p \in D$:

$$\operatorname{sign}(a_i^\intercal x - b_i) = \operatorname{sign}(p_i), \quad \operatorname{sign}(x_j) = \operatorname{sign}(c_j - p^\intercal A_j). \tag{11}$$

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 $c \in \mathbb{R}^n$ and scalar $d \in \mathbb{R}$ such that $S \subset \{x \in \mathbb{R}^n : c^\intercal x < d\}$ and $U \subset \{x \in \mathbb{R}^n : c^\intercal x > d\}$.*

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

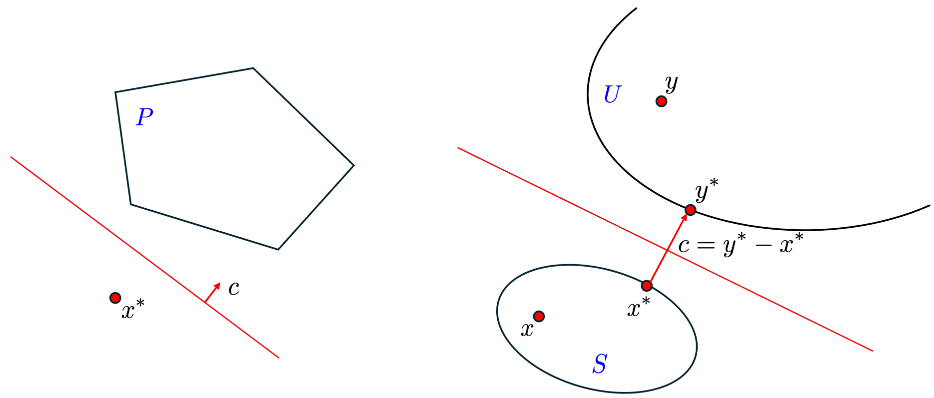

Figure 1: Separating Hyperplane Theorem. Left: separating a point from a polyhedral set. Right: more general result on separating two convex sets.

Proof: Consider the following optimization problem:

$$\begin{aligned} &\text{infimum } \|x - y\| \\ &\text{such that } x \in S,\ y \in U. \end{aligned} \tag{12}$$

We claim that the infimum is achieved, i.e., $\exists (x^*, y^*) \in S \times U$ such that $\|x^* - y^*\| \le \|x - y\|$ 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, 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:

$$\begin{aligned} \|x - y\| &\le \|x - x'\| + \|x' - y\| \\ \|x' - y\| &\le \|x' - x\| + \|x - y\|. \end{aligned}$$

These imply that

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

and therefore

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

which proves that f is continuous and that the minimum is achieved in (12). 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 $c := y^* - x^*$ and the scalar $d := \frac{c^\intercal(x^*+y^*)}{2}$ give a strictly separating hyperplane. Specifically, we prove that for any $x \in S$ and any $y \in U$:

$$c^\intercal x \le c^\intercal x^* < d < c^\intercal y^* \le c^\intercal y.$$

See the right panel of Figure 1 for the corresponding geometric intuition.

First, with $z := (x^* + y^*)/2$, observe that:

$$\begin{aligned} d - c^\intercal x^* &= (y^*-x^*)^\intercal(y^*-x^*)/2 > 0 \\ c^\intercal y^* - d &= (y^*-x^*)^\intercal(y^*-x^*)/2 > 0, \end{aligned}$$

where the inequalities hold because the optimal value in problem (12) must be strictly positive because S and U are closed and have non-empty intersection.

We next argue that $c^\intercal y^* \le c^\intercal y$ 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:

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

which implies that

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

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

It is important to note that the **strict** separation for convex sets requires all the assumptions in Theorem 6, namely (convex) sets that are closed and at least one of which is bounded. To see this, consider the following two examples, also depicted in Figure 2.

Example 1 (Strict Separation Failing) *Consider the following two sets in $\mathbb{R}^2$:*

$$S = [-1,1] \times [-1,0) \cup \{(x,y) : x \in [-1,0], y = 0\}, \quad U = [-1,1]^2 \setminus S.$$

*Both sets are bounded and **not** closed; no hyperplane that strictly separates them exists. Moreover, consider the choice:*

$$S = \{(x,y) : x \le 0\}, \qquad U = \{(x,y) : x \ge 0, y \ge 1/x\}.$$

*Both sets are closed and **unbounded**; no hyperplane that strictly separates them exists.*

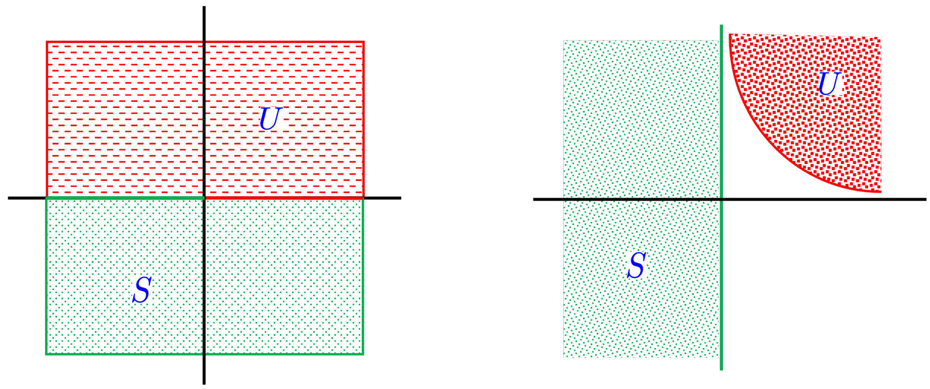

Figure 2: **Strict** Separating Hyperplane Theorem failing when assumptions are not met. **Left:** two convex sets that are not closed (but are both bounded) and that cannot be strictly separated. The sets are $S = [-1, 1] \times [-1, 0) \cup \{(x, y) : x \in [-1, 0], y = 0\}$ and $U = [-1, 1]^2 \setminus S$. **Right:** two convex sets that are closed but are unbounded that cannot be strictly separated. The sets are $S = \{(x, y) : x \leq 0\}$ and $U = \{(x, y) : x \geq 0, y \geq 1/x\}$.

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

Corollary 2 ${}^{\intercal}IfP$ *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 in Theorem 6 is much stronger than what we need, 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* $p^{\intercal}A \geq 0$ *and* $p^{\intercal}b < 0$.

Before proving this result, let's develop a bit of geometric intuition. Figure 3 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.

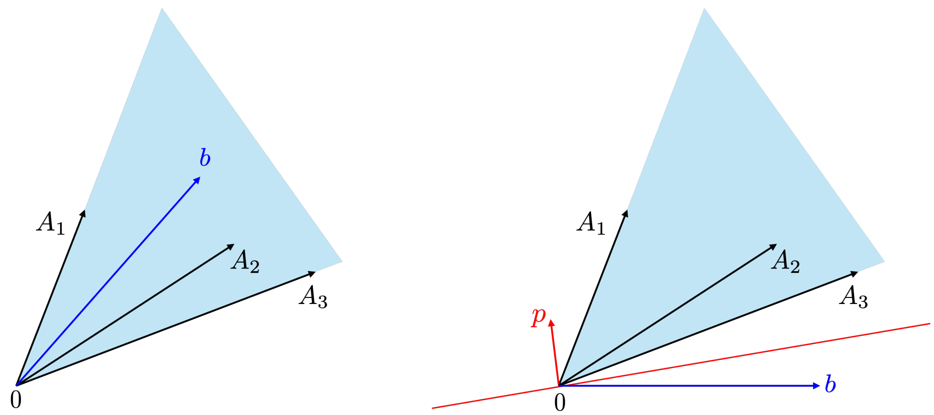

Figure 3: The two alternatives possible in the Farkas Lemma. Left: the vector b belongs to the cone generated by the columns A_i of the matrix A, so there exists $x \geq 0$ such that $Ax = b$. Right: the vector b does not belong to the cone generated by the columns of A, so a separating hyperplane exists.

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 $p^\intercal A \geq 0$, then $p^\intercal b = p^\intercal Ax \geq 0$, 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

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

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

[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.

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

$$\begin{array}{ll} \text{Primal Problem } (P_1): & \text{Dual Problem } (D_1): \\ \text{minimize } c^\intercal x & \text{maximize } p^\intercal b \\ \text{subject to } Ax \geq b, & \text{subject to } p^\intercal A = c^\intercal,\ p \geq 0. \end{array} \tag{13}$$

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(13) 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 4 for a visualization.)

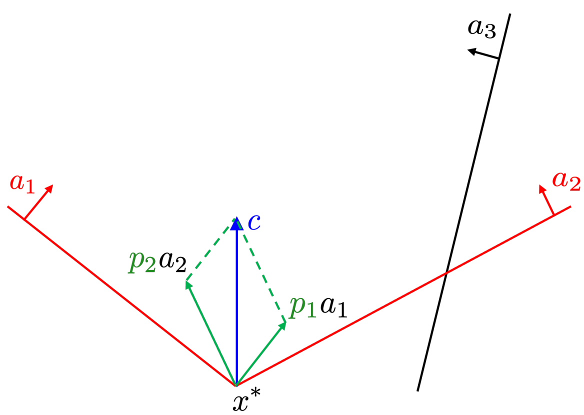

Figure 4: Interpretation of optimality conditions at x^*. The red constraints a_1, a_2 are active at x^*, whereas the black constraint a_3 is not active. The proof shows that in this case, the cost vector c can be generated as a conic combination of the active constraints a_1, a_2, with coefficients 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. Moreover, if $c^\intercal d < 0$, then $c^\intercal(x^* + \epsilon d) < c^\intercal x^*$ would 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 Possibilities for Primal/Dual Pairs

In view of the strong duality result, we can see that the only possibilities for a primal-dual pair are summarized in Table 1.

	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 2 (Infeasible $(\mathscr{P})$ and $(\mathscr{D})$) *Consider the infeasible primal problem:*

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

Its dual is:

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

which is also infeasible.

4 Some Initial Applications of Duality

We can now revisit some of the motivating applications discussed in Section 1. The previous results already showed how duality provides optimality certificates and certificates of infeasibility, so we now focus on more practical aspects.

4.1 Robust Optimization

Consider a setting where we have a linear program where one of the constraints is uncertain. Specifically, suppose one of the constraints defining the feasible set $Ax \le b$ is:[2]

$$a^\intercal x \le b, \tag{14}$$

where a is only known to reside in a polyhedral set, $a \in \mathcal{A}$, and we seek decisions x that are **robustly feasible**, i.e., that satisfy the constraint for any possible value of a:

$$a^\intercal x \le b, \, \forall \, a \in \mathcal{A}, \tag{15}$$

where the set $\mathcal{A}$ is a known polyhedral set

$$a \in \mathcal{A} := \{a \in \mathbb{R}^n \, : \, Ca \le d\} \, ,$$

where the matrices C and D are given. (As a more concrete example, consider the very practical case when we only have bounds on each coefficient, $\underline{a}_i \le a_i \le \bar{a}_i$, so the set $\mathcal{A}$ is a hyper-rectangle.)

Note that the constraint (15) makes our LP a difficult optimization problem – in fact, it is called a "semi-infinite" optimization problem because we would have an **infinite** number of constraints (one for every $a \in \mathcal{A}$).

However, strong duality will enable us to rewrite the constraint (15) as a **finite-dimensional** optimization problem even when $\mathcal{A}$ is specified through inequalities. In that case, constraint (15) is equivalent to:

$$\max_{a \, : \, Ca \le d} (a^\intercal x) \le b. \tag{16}$$

The left-hand side in (16) is a linear program. For the constraint to be feasible, that linear program must have a finite optimal value, in which case (by strong duality) its value will equal the value of its dual, which is given by the following problem:

$$\begin{aligned} \max_a \; & x^\intercal a & = \min_p \; & p^\intercal d \\ & Ca \le d & & p^\intercal C = x^\intercal \\ & & & p \ge 0. \end{aligned}$$

Then, constraint (16) is satisfied for a given x if and only if the following system of constraints is feasible in variable p:

$$\begin{aligned} p^\intercal d &\le b \\ p^\intercal C &= x^\intercal \\ p &\ge 0. \end{aligned}$$

This finite-dimensional system of linear (in)equalities in variables x, p reformulates the semi-infinite constraint (15). Note that the reformulation does require introducing new variables

[2] The case where the right-hand-side is also uncertain can be captured by extending the vector x with one extra component x_{n+1} constrained to equal 1.

and constraints: we introduced a new set of decisions p (as many as there are rows of C) and n additional constraints ($p^\intercal C = x^\intercal$).

As a small side remark, note that if we had access to a representation of the polyhedral set $\mathcal{A}$ through its extreme points and extreme rays, i.e., $\mathcal{A} = \text{conv}(\{a^1, \ldots, a^k\}) + \text{cone}(\{w^1, \ldots, w^r\})$, then we could readily reformulate constraint (15) equivalently as the following **finite-dimensional** system of inequalities:

$$\begin{aligned} &a^\intercal x \le b, \, \forall \, a \in \{a^1, \ldots, a^k\} \\ &a^\intercal x \le 0, \, \forall \, a \in \{w^1, \ldots, w^r\}. \end{aligned}$$

However, this approach may not be practical because there might be an exponentially large number of extreme points. (For instance, if $\mathcal{A} = \{a \in \mathbb{R}^n : \underline{a}_i \le a_i \le \bar{a}_i, \forall i\}$, there would be 2^n extreme points!)

4.2 A Polynomially-Sized Representation for CVaR

Recall the homework problem where we wanted to ensure that the Conditional Value-at-Risk (CVaR) of a portfolio payoff was exceeding some lower limit. CVaR was defined as the average over the k-smallest values among the monetary payoffs (for a suitable integer k). As such, if the values of the monetary payoffs in the different scenarios are given as $v_1, v_2, \ldots, v_n$, the key constraint that needs to be met is of the form:

$$\sum_{i=1}^{k} v_{[i]} \ge b, \tag{17}$$

where $v_{[1]} \le v_{[2]} \le \cdots \le v_{[n]}$ is the sorted vector of payoffs. That constraint can be satisfied by enumerating over all the vectors $x \in \{0,1\}^n$ that contain exactly k values of 1. However, that formulation would require exponentially many constraints.

Here, we leverage strong duality to rewrite the constraint with a "small" (i.e., polynomially-sized) number of variables and constraints. We claim that the sum of the k-smallest values among the values $v_1, \ldots, v_n$ can be obtained as the optimal value of the linear program:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \; & \sum_{i=1}^{n} v_i x_i \\ & x \ge 0 \\ & x \le e \\ & e^\intercal x = k. \end{aligned} \tag{18}$$

In the above formulation, the vector $x \in [0,1]^n$ represents weights placed on each value v_i. The weights are non-negative and there is a budget k that must be exactly distributed for the weights. We claim that LP (18) achieves its optimal value at:

$$x_i = \begin{cases} 1, & \text{if } v_i \in \{v_{[1]}, v_{[2]}, \ldots, v_{[n]}\} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

This solution is clearly feasible in the LP (18) and its optimality can be shown through a simple interchange argument. Assume that some other feasible solution could achieve a strictly lower objective. Such a solution would necessarily be taking some strictly positive weight from at least one of $\{v_{[1]}, v_{[2]}, \ldots, v_{[n]}\}$ to distribute to $\{v_{[i]} : i > n\}$, all of which are (weakly) larger; so that solution cannot strictly decrease the objective.

By strong duality, the optimal value of LP (18) is the same as the optimal value of its dual:

$$\begin{aligned} \max_{p,t}\ & e^\intercal p + k \cdot t \\ & p + t \cdot e \leq v \\ & p \geq 0. \end{aligned}$$

So constraint (17) is satisfied if and only if the following system in variables p, t is feasible:

$$\begin{aligned} & e^\intercal p + k \cdot t \geq b \\ & p + t \cdot e \leq v \\ & p \geq 0. \end{aligned}$$

We challenge the reader to try proving the optimality of the solution (19) (to LP (18)) by producing a feasible solution to the dual with the same optimal value!

5 Other Implications of Duality

We next discuss a few other important results and implications of duality, including optimality conditions (for standard-form LPs and general LPs), computational choices to make when solving LPs, complementary slackness, and the resolution theorem.

5.1 Optimal Solutions for Standard-Form LPs.

Strong duality provides powerful certificates of (in)feasibility and (sub)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 assures us that these optimality bounds are actually **really good** and that the dual optimal solution provides a "certificate of optimality" for the primal (so we would be able to **prove** that a given feasible x is optimal by simply checking whether it satisfies $c^\intercal x = b^\intercal p^*$).

To appreciate these points, it is helpful to reconsider the case of a standard-form LP and examine some properties of the optimal primal and dual solutions. So consider a primal in standard form and its dual:

$$\begin{aligned} (\mathscr{P})\ \ \min\ & c^\intercal x & \qquad (\mathscr{D})\ \ \max\ & p^\intercal b \\ & Ax = b & & p^\intercal A \leq c^\intercal \\ & x \geq 0 & & \end{aligned}$$

Primal $(\mathscr{P})$**.** Recall from our earlier classes that the primal always admits an optimal solution that is a basic feasible solution. 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 then be determined as

$$x = [x_B, 0], \quad x_B = A_B^{-1} b.$$

The feasibility of x therefore translates into the condition:

$$\text{Feasibility-}(\mathscr{P}): \qquad x_B := A_B^{-1} b \geq 0. \tag{20}$$

Moreover, recall from our lecture on the simplex algorithm that the condition for optimality of x is that the **reduced costs** are non-negative, i.e.,

$$\text{Optimality-}(\mathscr{P}): \qquad c^\intercal - c_B^\intercal A_B^{-1} A \geq 0. \tag{21}$$

Dual $(\mathscr{D})$. The same basis B that gave us a primal solution 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, this system has a unique solution, which can be written as:

$$p^\intercal = c_B^\intercal A_B^{-1}, \, \forall \, i \in B. \tag{22}$$

The dual objective value of p is exactly:

$$p^\intercal b = c_B^\intercal A_B^{-1} b = c^\intercal x.$$

Moreover, p is feasible in the dual if and only if:

$$\text{Feasibility-}(\mathscr{D}): \qquad c^\intercal - p^\intercal A \geq 0 \quad \Leftrightarrow \quad c^\intercal - c_B^\intercal A_B^{-1} A \geq 0 \tag{23}$$

So **dual feasibility is exactly the same as primal optimality**! This gives an alternative interpretation to the termination conditions for the simplex method: simplex stops when the current basis B corresponds to a solution p that is feasible for the dual! Because that solution p has the same cost as the current primal solution x by construction, when p is feasible it automatically **proves** the optimality of the primal solution x.

5.2 Solve $(\mathscr{P})$ or $(\mathscr{D})$? Primal and Dual Simplex.

Instead of solving the primal, one can attempt to solve the dual. The **dual simplex** algorithm does exactly that: it tries to solve the **dual** $(\mathscr{D})$ by maintaining dual feasible solutions p and trying to prove their optimality – or equivalently, construct a primal solution x with the same objective that is feasible in $(\mathscr{P})$. Whereas the primal simplex is swapping decisions in and out of the basis, the dual simplex can then be seen as 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. We do not include a detailed discussion here but refer the interested reader to standard textbooks (such as Bertsimas & Tsitsiklis) for more.

Here, we just note that the dual simplex method is **not** a perfect mirror of the primal simplex because dual simplex works entirely with a problem with inequality constraints, whereas the primal simplex is tailored to problems in standard form.

A natural question is: when should one solve ($\mathscr{P}$) (with primal-simplex) rather than solve ($\mathscr{D}$) (with dual-simplex)? This largely depends on specification of the problem and an important consideration is whether it is easier to generate a feasible solution for one formulation rather than the other. Sometimes, optimal solutions are available for a "simpler" version of the problem and the choice of method depends on how the "larger" problem is related to the simpler one. For instance, suppose we solved a primal problem with n decisions and m constraints to optimality and we have both a primal-optimal solution x^* and a dual-optimal solution p^*. Then:

- If the larger problem involves an extra set of decisions x_e, it is natural to consider the primal simplex because we can readily generate a feasible solution for the larger problem as $[x^*, x_e = 0]$.
- If the larger problem involves adding extra constraints in the primal, $A_e x = b_e$, then it is natural to consider the dual simplex because we can readily generate a feasible solution for the larger problem as $[p^*, p_e = 0]$.

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. If you are curious, you can read more at this url: `https://www.gurobi.com/documentation/current/refman/method.html`.

5.3 Optimality Conditions and Complementary Slackness

Sometimes all we want is to **characterize** the optimal solutions to a problem – this is often the case in theoretical work, where we want to show that solutions 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 $u_i = p_i(a_i^\intercal x - b_i)$ and $v_j = (c_j - p^\intercal A_j)x_j$ 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:

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

By the strong duality theorem, if x and p are optimal, then $c^\intercal x = p^\intercal b$, which implies that $u_i = v_j = 0$ for all i, j. Conversely, if $u_i = v_j = 0$ for all i, j, then $c^\intercal x = p^\intercal b$, 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 large-scale problems in practice!

Also in some special cases, we can actually strengthen this result to a **strict complementarity** result. One such instance is due to Goldman and Tucker, and concerns the following primal-dual pair.

Theorem 10 (Goldman and Tucker) *Consider the following primal-dual pair of LPs:*

$$
\begin{aligned}
(\mathscr{P}) \quad & \min\ c^\intercal x & \qquad (\mathscr{D}) \quad & \max\ p^\intercal b \\
& Ax = b & & p^\intercal A \leq c^\intercal \\
& x \geq 0 & &
\end{aligned}
$$

*If ($\mathscr{P}$) and ($\mathscr{D}$) are both feasible, they admit optimal solutions satisfying **strict complementarity**, i.e., there exist x^* and p^* optimal in ($\mathscr{P}$) and ($\mathscr{D}$), respectively, such that*

$$x_j^* > 0 \quad \Leftrightarrow \quad p^\intercal A_j = c_j.$$

We will not prove this here, but the result will be useful when we discuss one of the main applications of duality in finance (for proving a fundamental result in asset pricing).

5.4 Representation of Polyhedra

An important consequence of duality theory is an alternative representation of all polyhedra, in terms of their extreme points and extreme rays. We first need a quick definition.

Definition 1 *Consider a nonempty polyhedron $P = \{x \in \mathbb{R}^n : Ax \geq b\}$. Then:*

1. *The set $\mathcal{C} := \{d \in \mathbb{R}^n : Ad \geq 0\}$ is called the **recession cone** of P, i.e., the set of directions d along which we can move indefinitely without leaving the set P. Any nonempty element of the recession cone is referred to as a **ray** of P.*

2. *Any $d \in \mathcal{C}$ for which there exists $S \subset \{1, \ldots, m\}$ with $|S| = n - 1$ and $a_i^\intercal d = 0$, $\forall i \in S$ (i.e., $n - 1$ linearly independent constraints active) is called an **extreme ray** of P.*

The rays of P are directions in which we can move indefinitely in the set P, and the recession cone is the set of all rays; the extreme rays are extremal directions, and any ray can be

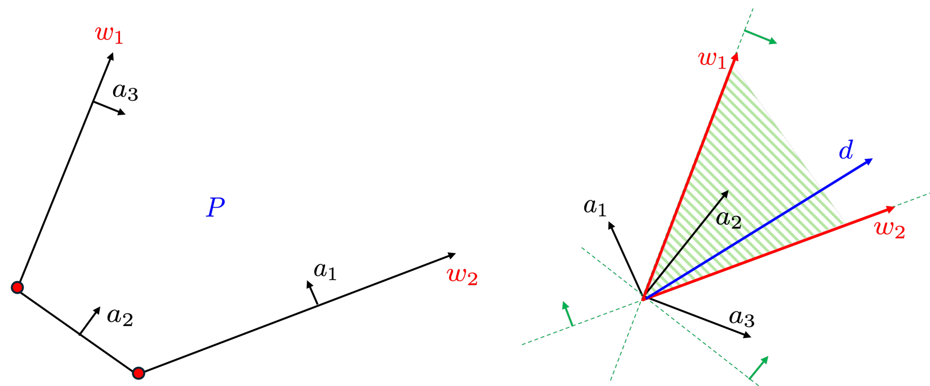

Figure 5: A polyhedron and its recession cone and extreme rays. Left: the polyhedron $P = \{x : Ax \geq b\}$ and the rows of the matrix A. Right: the shaded area denotes the recession cone of P (described by the green normal vectors) with a generic ray d shown in blue; the extreme rays are w_1 and w_2, which can also be visualized in the figure on the left.

written as a conic combination of the extreme rays. A visualization is provided in Figure 5. Clearly, any poltyope (i.e., bounded polyhedron) has $\mathcal{C} = \{0\}$.

Recession cones provide some insight regarding cases where optimization problems are unbounded. In particular, the following result is true.

Proposition 1 *Consider the problem minimize*$\{c^\intercal x : x \in P := \{x \in \mathbb{R}^n : Ax \geq b\}\}$*. Then, the optimal value is $-\infty$ if and only if there exists a ray d of P such that $c^\intercal d < 0$; moreover, if P has at least one extreme point, that ray can be taken as an extreme ray.*

We do not prove this result, but the intuition should be quite immediate from Figure (5). With this result, we can provide the following fundamental representation result for an arbitrary polyhedron P.

Theorem 11 (Resolution Theorem) *Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ be a non-empty polyhedron. Let $x^1, x^2, \ldots, x^k$ be its extreme points (possibly empty) and let $w^1, w^2, \ldots, w^r$ be the set of extreme rays of P (possibly empty). Then, with*

$$Q := conv\big(\{x^1, \ldots, x^k\}\big) + cone\big(\{w^1, \ldots, w^r\}\big) = \left\{\sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j : \lambda \geq 0, \theta \geq 0, e^\intercal \lambda = 1\right\},$$

we have $P = Q$.

Proof: Proving that $Q \subseteq P$ can be done by checking that $Ax \leq b$ holds for any $x \in Q$.

To prove $P \subseteq Q$, assume for purposes of deriving a contradiction that there exists $z \in P$

such that $z \notin Q$. Consider the linear programming problem:

$$
\begin{aligned}
(P_1) \quad \underset{\lambda,\theta}{\text{maximize}} \quad & \sum_{i=1}^{k} 0\lambda_i + \sum_{j=1}^{r} 0\theta_j \\
\text{subject to} \quad & \sum_{i=1}^{k} \lambda_i x^i + \sum_{j=1}^{r} \theta_j w^j = z \\
& \sum_{i=1}^{k} \lambda_i = 1 \\
& \lambda_i \geq 0, \quad i = 1, \ldots, k, \\
& \theta_j \geq 0, \quad j = 1, \ldots, r,
\end{aligned}
$$

which is infeasible because $z \notin Q$. This problem is the dual of the problem

$$
\begin{aligned}
(D_1) \quad \text{minimize}_{p,q} \; & p^\intercal z + q \\
\text{subject to } & p^\intercal x_i + q \geq 0, \quad i = 1, \ldots, k, \\
& p^\intercal w_j \geq 0, \qquad j = 1, \ldots, r,
\end{aligned}
$$

Because problem (P_1) is infeasible whereas (D_1) is trivially feasible with $p = q = 0$, it must be that (D_1) has optimal cost $-\infty$. This implies that (D_1) admits as feasible (p, q) so that $p^\intercal z + q < 0$. Because p is feasible, this implies $p^\intercal z < -q \leq p^\intercal x_i$ for any $i = 1, \ldots, k$, and also that $p^\intercal w_i \geq 0$.

Having fixed p as above, we now consider the linear programming problem

$$
\begin{aligned}
& \text{minimize } p^\intercal x \\
& \text{subject to } Ax \geq b.
\end{aligned}
$$

If the optimal cost is finite, there exists an extreme point x^i which is optimal. Because $z \in P$ is a feasible solution in this problem and $p^\intercal z < p^\intercal x_i$, that would lead to a contradiction. If the cost is $-\infty$, this implies (by Proposition 1) that there exists an extreme ray w^j of P such that $p^\intercal w^j < 0$, which is also a contradiction. $\square$

The resolution theorem is a fundamental result in linear optimization that states that a polyhedron can be represented in two ways: (i) in terms of a finite number of linear constraints, or (ii) in terms of a finite collection of extreme points and extreme rays. However, going from one description to the other is a nontrivial task.

6 Dual Variables as Marginal Costs; Sensitivity Analysis

Next, we develop an important interpretation of the dual variables as **marginal costs** (or **shadow prices**) for the constraints they are associated with. These results will provide a direct economic interpretation of shadow prices and will also lead to simple mnemonic rules for remembering the signs of shadow prices and writing duals.

Because some of the developments are related to the simplex algorithm, we work with a primal in standard form, but we emphasize that the concepts — and particularly the shadow price interpretation – hold for general LPs. Consider the following primal-dual pair:

$$(\mathscr{P}) \ \min c^\intercal x \qquad (\mathscr{D}) \ \max p^\intercal b \tag{24}$$
$$Ax = b \qquad p^\intercal A \leq c^\intercal$$
$$x \geq 0$$

Suppose we solved this LP and obtained an optimal primal solution x^* and an optimal dual solution p^*. Our main goal is to show that p^* is the gradient of the optimal cost with respect to b ("almost everywhere"). To that end, we must examine how the optimal cost changes with b, which is a type of **sensitivity analysis**. (Sensitivity analysis tries to understand how results change when problem parameters change; this is important in their own right because problem parameters are often estimated from noisy/imperfect data, so it is natural to examine how the noise in the estimates might impact the solution.)

6.1 Dependency on b.

Let $P(b) := \{x : Ax = b, x \geq 0\}$ denote the feasible set in (24) as a function of b, and $F(b)$ denote the optimal cost as a function of b. Throughout this section, we assume that the dual feasible set $\{p : p^\intercal A \leq c^\intercal\}$ is nonempty; this set is independent of b, so this assumption guarantees that $F(b) > -\infty$.

We want to prove that the function $F(b)$ is **convex** and piece-wise linear in b.

Definition 2 (Convex/concave functions) *A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is **convex** if its domain X is a convex set and for any $x, y \in X$ and $\lambda \in [0, 1]$, we have:*

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y). \tag{25}$$

*A function is **concave** if $-f$ is convex.*

The geometric intuition is depicted in Figure 6. For a convex function, the line segment between any two points $(x, f(x))$ and $(y, f(y))$ lies above the graph of the function.

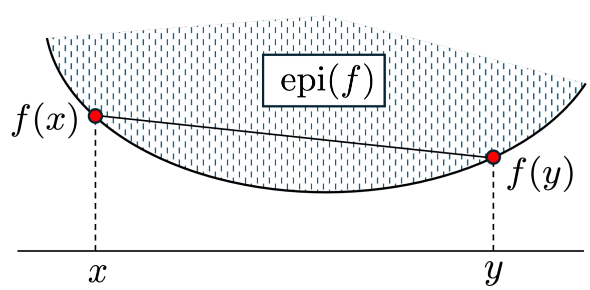

Figure 6: For a convex function, the line segment between two points $(x, f(x))$ and $(y, f(y))$ lies above the graph. The epigraph is the shaded area (extending vertically to $+\infty$).

An equivalent definition of a convex function is in terms of its **epigraph**. The **epigraph** of a function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is the set of points that lie on or above its graph:

$$\text{epi}(f) = \{(x, t) \in X \times \mathbb{R} \, : \, t \geq f(x)\}. \tag{26}$$

Then, f **is convex if and only if epi**(f) **is a convex set**.

With these definitions, we can present the first main result.

Theorem 12 (Global Dependency On b.) *The optimal cost $F(b)$ is a convex and piecewise linear function of b on the set $S := \{b : P(b) \neq \emptyset\}$.*

Proof: First, we show that S is convex. Let $b_1, b_2 \in S$ and for any $\lambda \in [0,1]$, let $b := \lambda b_1 + (1-\lambda) b_2$. Also let $x_i \in \arg\max\{c^\intercal x : x \geq 0, Ax = b_i\}$ and define $x_\lambda := \lambda x_1 + (1-\lambda) x_2$.

To prove that $b \in S$, note that:

$$x_\lambda \geq 0 \text{ and } Ax_\lambda = A(\lambda x_1 + (1-\lambda)x_2) = \lambda b_1 + (1-\lambda) b_2 := b,$$

which implies that $x_\lambda \in P(b)$ and therefore $b \in S$, so S is convex. Although the convexity of F can be proved from base principles, we adopt a different proof that leverages duality. Because we assumed that the dual is feasible, $F(b)$ will be finite and (by strong duality) will equal the optimal value of the dual. Let $p^1, p^2, \ldots, p^r$ be the extreme points of the dual feasible set. (Clearly these exist because the optimal values are finite.) Then, we have:

$$F(b) = \max_{i=1,\ldots,r} b^\intercal p^i, \; \forall\, b \in S. \tag{27}$$

Figure 7 visualizes this result. $F(b)$ is the pointwise maximum of a finite collection of linear functions of b. That this function is convex follows readily from properties of convex functions, but to see an elementary proof, note that the epigraph of $F(b)$ is the intersection of the epigraphs of the functions $b^\intercal p^i$ for $b \in S$. Because S is convex, every such epigraph is convex, and the intersection of convex sets is convex, proving that $F(b)$ is convex. $\square$

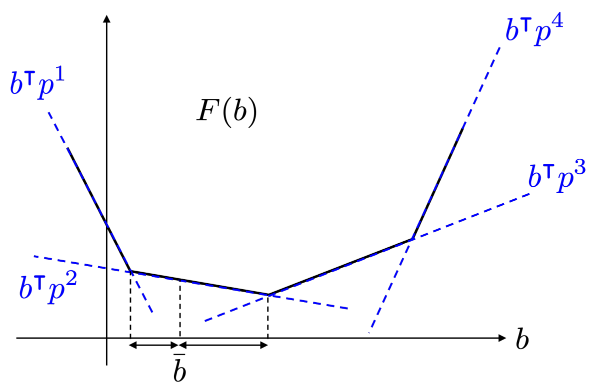

Figure 7: The optimal value $F(b)$ as a function of the right-hand-side b. The dashed linear functions (in blue) correspond to extreme points p_i of the dual feasible set. Note that at any point $\bar{b}$ where $F(b)$ is differentiable, $F(b)$ is linear in a certain range and the optimal dual variable p^* defines its gradient.

This representation gives the interpretation we were seeking for dual variables. Note that at any point $b = \bar{b}$ where $F(b)$ is differentiable, **the optimal dual solution p^* is exactly the gradient of $F(b)$**. The dual variable p_i associated with the i-th constraint thus acts as a "marginal cost" or "shadow price" associated with the constraint. Moreover, because $F(b)$ is linear in a certain range around $\bar{b}$, the dual variable allows estimating the **exact change in $F(b)$ in that range**: increasing b_i by δ increases $F(b)$ by $\delta \cdot p_i$.

Modern solvers give direct access to (i) the optimal value of the dual variable associated with any constraint and (ii) the allowable range described above within which the optimal *objective is linear. The linear optimization tutorial that we circulated describes how that can be done in Gurobipy: for a constraint object* **c***, the optimal dual variable can be obtained from the attribute* **c.Pi***, the left end of the range can be obtained from* **c.SARHSLow***, and the right end of the range can be obtained from* **c.SARHSUp***.* If you are wondering how this is done, read through §6.4.

When the change in a particular right-hand-side component b_i is large enough to put us outside this allowable range, we cannot predict exactly the change in the objective or the new value for the dual-optimal solution (we would need to resolve the problem!), but we can predict **directionally** what could happen. Specifically, a simple mnemonic rule to remember is that **outside the allowable range, things only "get worse"**. For instance, if increasing b_i increased the optimal costs (so p_i was positive) in the range, increasing b_i outside the range would result in an even larger p_i and costs increasing at faster rate. In this case, although decreasing b_i in the range reduced the costs, decreasing b_i beyond the range would reduce costs at a slower rate (so p_i would decrease). A similar interpretation also holds for the case when increasing b_i reduced costs (so p_i was negative) in the range: increasing b_i beyond the range would reduce

6.1.1 Optimal Dual Solutions As Subgradients of $F(b)$

The points b where $F(b)$ is not differentiable actually correspond to cases where **multiple dual optimal solutions exist** and it can be shown that in that case, all such dual solutions are valid **subgradients** of $F(b)$. To formalize this, we introduce one more definition.

Definition 3 (Subgradient.) *Let F be a convex function defined on a (convex) set S. Let $\bar{b}$ be an element of S. We say that a vector p is a* ***subgradient*** *of F at $\bar{b}$ if*

$$F(\bar{b}) + p^\intercal(b - \bar{b}) \leq F(b), \quad \forall b \in S.$$

At points b where $F(b)$ is differentiable, there is a unique subgradient; but at breakpoints, multiple subgradients are valid (which form the so-called "sub-differential"). The following result makes the connection with optimal dual variables clear.

Theorem 13 *Suppose that the linear programming problem $\min\{c^\intercal x : Ax = \bar{b}, x \geq 0\}$ has a finite optimal cost. Then, a vector p is an optimal solution to the dual problem if and only if it is a subgradient of the optimal cost function F at the point $\bar{b}$.*

Proof: We first show that any dual optimal p is a valid subgradient. Suppose that p is an optimal solution to the dual. Then, strong duality implies that $p^\intercal\bar{b} = F(\bar{b})$. Consider now some arbitrary $b \in S$. For any feasible solution $x \in P(b)$, weak duality yields $p^\intercal b \leq c^\intercal x$. Taking the minimum over all $x \in P(b)$, we obtain $p^\intercal b \leq F(b)$. Hence,

$$p^\intercal b - p^\intercal\bar{b} \leq F(b) - F(\bar{b}),$$

and we conclude that p is a subgradient of F at $\bar{b}$.

For the reverse direction, let p be a subgradient of F at $\bar{b}$, that is,

$$F(\bar{b}) + p^\intercal(b - \bar{b}) \leq F(b), \quad \forall b \in S. \tag{28}$$

Pick some $x \geq 0$ and let $b = Ax$. Then, $x \in P(b)$ and $F(b) \leq c^\intercal x$. By (28), we have:

$$p^\intercal Ax = p^\intercal b \leq F(b) - F(\bar{b}) + p^\intercal \bar{b} \leq c^\intercal x - F(\bar{b}) + p^\intercal \bar{b}.$$

Because this is true for all $x \geq 0$, we must have $p^\intercal A \leq c^\intercal$, which shows that p is a dual feasible solution. Also, by letting $x = 0$, we obtain $F(\bar{b}) \leq p^\intercal \bar{b}$. Using weak duality, every dual feasible solution q must satisfy $q^\intercal \bar{b} \leq F(\bar{b}) \leq p^\intercal \bar{b}$, which shows that p is optimal. □

6.2 Dependency on c.

In a similar fashion, let $G(c)$ denote the optimal cost in (24) as a function of c, and define the set $T := \{c \in \mathbb{R}^n : \min_{x \geq 0, Ax=b} c^\intercal x > \infty\}$ as all cost vectors resulting in a finite objective value. The following result summarizes the dependency of the optimal value on c.

Theorem 14 *Consider a feasible linear programming problem in standard form. Then:*
(a) The set T of all c for which the optimal cost is finite is convex.
(b) The optimal cost $G(c)$ is a ***concave*** *function of c on the set T.*
(c) If for some value of c the primal problem has a unique optimal solution x^*, then G is linear in the vicinity of c and its gradient is equal to x^*.

The proof follows by applying an identical line of arguments to the dual.[3]

Importantly, this result also affords an interpretation for the optimal primal solution x^* as a shadow price for the constraints in the dual and implies that x^* remains optimal for a certain range of changes in each objective coefficient c_j.

Modern solvers also allow obtaining this range of changes. For instance, in Gurobipy, every decision variable x_j has two attributes, **SAObjLow** *and* **SAObjUp**. *Once an optimal solution x^* is obtained, these attributes will specify the lower limit and upper limit, respectively, of the values of the objective coefficient c_j (corresponding to variable x_j) within which the solution x^* would remain optimal.*

6.3 Optimal Dependency and Signs of Shadow Prices.

On remark is important. Although our discussion in §6.1 and §6.2 relied on a primal in standard form, the results hold for a primal problem in general form. That is, if we considered a primal in general form and defined $F(b, c)$ as its optimal value:

$$\begin{array}{llllll}
F(b,c) := \min_x & c^\intercal x & & = \max_p & p^\intercal b & \\
& a_i^\intercal x \geq b_i, & i \in M_1, & & p_i \geq 0, & i \in M_1, \\
& a_i^\intercal x \leq b_i, & i \in M_2, & & p_i \leq 0, & i \in M_2, \\
& a_i^\intercal x = b_i, & i \in M_3, & & p_i \text{ free}, & i \in M_3, \\
& x_j \geq 0, & j \in N_1, & & p^\intercal A_j \leq c_j, & j \in N_1, \\
& x_j \leq 0, & j \in N_2, & & p^\intercal A_j \geq c_j, & j \in N_2, \\
& x_j \text{ free}, & j \in N_3. & & p^\intercal A_j = c_j, & j \in N_3.
\end{array}$$

[3]Specifically, we can convert the dual maximization into the minimization problem $-\min(-p^\intercal b)$ and apply arguments similar to those above because c is a right-hand-side in this problem.

we would readily have that $F(b,c)$ is a piece-wise linear, convex function of b and a piece-wise linear, concave function of c. Moreover, the optimal dual solutions p^* are valid supergradients for $F(b,c)$ with respect to b and the optimal primal solutions x^* are valid subgradients for $F(b,c)$ with respect to c.

This discussion also highlights and reemphasizes the direct connection between the signs of shadow prices and the types of optimization problems solved. The only four possibilities, which we already discussed in Table 1, are again depicted in Figure 8. Note that given the type of optimization problem (minimization or maximization) and the type of constraint ($\leq$ or $\geq$), one can precisely infer the sign of the dual variable corresponding to that constraint and also predict the kind of dependency that the optimal objective has with respect to the right-hand-side.

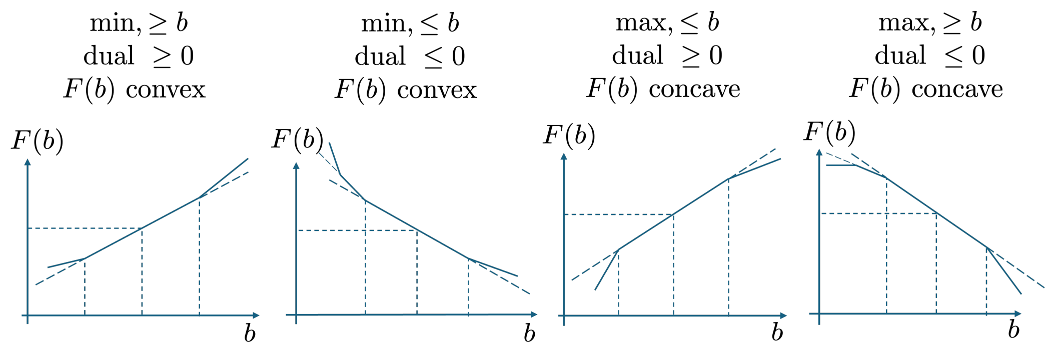

Figure 8: The relationship between the type of optimization problem solved, the type of constraint, and the sign of the dual variable. Given two, one can figure out the third.

6.4 Exact Range for Shadow Prices and Local Sensitivity Analysis

We now consider the question of characterizing the precise range of changes in the right-hand-side of a constraint (discussed in §6.1) or in an objective coefficient (discussed in §6.2) that allow us to precisely infer what happens to the optimal objective or the optimal solution. This will rely on a **local sensitivity analysis** related to the workings of the simplex algorithm, so we again consider the primal LP in standard form and its dual:

$$(\mathscr{P})\ \min\ c^\mathsf{T}x \qquad\qquad (\mathscr{D})\ \max\ p^\mathsf{T}b$$
$$Ax = b \qquad\qquad p^\mathsf{T}A \leq c^\mathsf{T}$$
$$x \geq 0$$

Suppose we solved these problems are recovered an optimal basic feasible solution x for the primal and a corresponding optimal p for the dual. Recall from Section 5.1 that for a given basis B, the feasibility and optimality of the primal basic solution $x := [x_B, 0]$ with $x_B := A_B^{-1}b$ and the corresponding dual choice $p^\mathsf{T} := c_B^\mathsf{T}A_B^{-1}$ translate into the conditions:

$$\text{Feasibility in } (\mathscr{P}): \qquad x_B \equiv A_B^{-1}b \geq 0 \tag{29a}$$
$$\text{Optimality in } (\mathscr{P})/\text{Feasibility in } (\mathscr{D}): \qquad c^\mathsf{T} - p^\mathsf{T}A \equiv c^\mathsf{T} - c_B^\mathsf{T}A_B^{-1}A \geq 0. \tag{29b}$$

Subsequently, we examine what happens to the optimal solution when we make changes in b or c. (For other changes, see Chapter 5 of the Bertsimas and Tsitsiklis text.) More specifically, we will try to understand the conditions under which the changes in b and c would **preserve the same optimal basis** B, which will then allow us to characterize how the optimal primal or dual solutions would change.

6.4.1 Local Changes in the Right-Hand-Side b

Suppose that we **change** b **to** $b+\theta e_i$, where e_i is the i-th unit vector. We want to determine the range of values of θ under which the current basis B remains optimal. Note that the optimality conditions (29b) are not affected by a change in b, so we only need to examine the feasibility condition (29a), which now becomes:

$$A_B^{-1}(b+\theta e_i) \geq 0 \;\;\Leftrightarrow\;\; x_B + \theta g \geq 0,$$

where $g := [g_1, g_2, \ldots, g_m]^\intercal$ denotes the i-th column of the matrix A_B^{-1}. Clearly, the condition above holds provided that the change θ is sufficiently small; more precisely, we need

$$\max_{i:g_i>0}\left(-\frac{x_{Bi}}{g_i}\right) \leq \theta \leq \min_{i:g_i<0}\left(-\frac{x_{Bi}}{g_i}\right). \tag{30}$$

For θ in this range, we have the following observations:

- the optimal primal solution is $x^* := [x_B + \theta g,\, 0]$.
- the optimal dual solution remains unchanged.
- the optimal value of the primal changes by:

$$c_B^\intercal A_B^{-1}(b+\theta e_i) = p^\intercal b + \theta p_i. \tag{31}$$

Equation (31) again restates our interpretation of the dual variable p_i as the derivative of the optimal objective value with respect to the right-hand-side b_i. Moreover, the discussion above confirms that this dependency is **exact** for any changes of the right-hand-side coefficient b_i in the range satisfying (30). Modern solvers are thus able to construct the interval (30) based on the optimal simplex tableau.

If θ is outside the allowed range in (30), the current basis satisfies the primal optimality/dual feasibility condition (29b), but is primal infeasible. The dual variables (i.e., shadow prices) will change in a predictable direction, but the magnitude of the change cannot be precisely known, as we saw in §6.1.

6.5 Local Changes In the Cost Vector c

Suppose now that some cost coefficient c_j becomes $c_j + \theta$. The primal feasibility condition (29a) is not affected. We therefore need to focus on the optimality condition (29b):

$$c^\intercal \geq p^\intercal A \text{ where } p = c_B^\intercal A_B^{-1}.$$

If c_j is the cost coefficient of a nonbasic variable x_j, then c_B does not change and the only inequality above that is affected is the one for the j-th component. We need:

$$c_j + \theta \geq p^\intercal A_j \quad \Leftrightarrow \quad \theta \geq -\underline{c}_j, \tag{32}$$

where $\bar{c}_j := c_j - p^\intercal A_j$ is the j-th reduced cost.

If c_j is the cost coefficient of a basic variable, then c_B becomes $c_B+\theta e_j$ and the optimality conditions in the new problem would become:

$$c_i \geq (c_B + \theta e_j)^\intercal A_B^{-1} A_i,\ i \neq j.$$

(We are omitting the constraint for j because it will be automatically satisfied.) The constraints above amount to

$$\theta\big(e_j^\intercal A_B^{-1} A_i\big) \leq \bar{c}_i := c_i - p^\intercal A_i. \tag{33}$$

The lower bounds (32) and the upper bounds (33) establish the range of changes for each cost coefficient c_j within which the optimal basis remains unchanged, which here implies that **the primal solution x would remain optimal**. In these ranges, the optimal costs would increase by a suitable amount, θx_j.

7 More Complex Applications of Duality

We discuss a few more advanced applications of duality in finance and in airline network revenue management.

7.1 Asset Pricing and No-Arbitrage

Let us assume than we are in an investment world where there are $n+1$ securities indexed by $i = 0, \ldots, n$, where $i = 0$ denotes cash and the other securities could be anything (stocks, bonds, complex derivatives).[4] Moreover, we have two periods: the current period "c" and the future period "f". In the current period, the prices of the securities are S_i^c for $i = 1, \ldots, n$ are for cash $S_0^c = 1$. In the future period, the prices are uncertain. Specifically, there is a finite number of states of the world denoted by

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\},$$

each occurring with positive probability, and the prices of the securities are as follows:

- cash is assumed to be riskless, so the price of cash will be $S_0^f = R = 1 + r$, where r is the risk-free rate of return during the current period and the future
- for $i > 1$, security i will have a price that depends on the state of the world, $S_i^f(\omega_j)$.

To state the main result here, we first need a few definitions.

[4]Our treatment here follows a simplified version of Chapter 4 of the book "Optimization Methods in Finance" by Tütüncü and Cornuejols, to which we direct the interested reader for generalizations.

Definition 4 (Arbitrage) *An* ***arbitrage*** *is a trading strategy that either has a positive initial cashflow and has no risk of a loss later (type A) or that requires no initial cash input, has no risk of loss, and has a positive probability of making profits in the future (type B).*

Intuitively, the arbitrage amounts to a sort of "free lunch" where an investor can get something for nothing. To appreciate what this means in our investment model, note that if we purchase an amount x_i of each security $i \in \{0, \ldots, n\}$, we would incur an immediate cost $\sum_{i=0}^{n} S_i^c x_i$ (for the purchase in the current period) and we would have a future cashflow of $\sum_{i=0}^{n} S_i^f(\omega) \cdot x_i$ if the state of the world turns out to be $\omega \in \Omega$. Thus,

- a type-A arbitrage would mean that we can find an investment strategy x_i such that:

$$\begin{aligned} \sum_{i=0}^{n} S_i^c \cdot x_i &< 0 && \text{(positive initial cashflow)} \\ \sum_{i=0}^{n} S_i^f(\omega) \cdot x_i &\geq 0, \, \forall \, \omega \in \Omega && \text{(no risk of loss)} \end{aligned} \tag{34}$$

- a type-B arbitrage would mean that we can find an investment strategy x_i such that:

$$\begin{aligned} \sum_{i=0}^{n} S_i^c \cdot x_i &= 0 && \text{(no initial cash input)} \\ \sum_{i=0}^{n} S_i^f(\omega) \cdot x_i &\geq 0, \, \forall \, \omega \in \Omega && \text{(no risk of loss)} \\ \exists \omega \in \Omega \, : \, \sum_{i=0}^{n} S_i^f(\omega) \cdot x_i &> 0, && \text{(positive probability of profit).} \end{aligned} \tag{35}$$

We also need one more definition.

Definition 5 (Risk-neutral probability measure) *A* ***risk-neutral probability measure*** *on the set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ is a vector $p \in \mathbb{R}^m$ so that $p > 0$ and $\sum_{j=1}^{m} p_j = 1$ and for every security $S_i, i = 0, \ldots, n$,*

$$S_i^c = \frac{1}{R} \left(\sum_{j=1}^{m} p_j S_i^f(\omega_j) \right) = \frac{1}{R} \mathbb{E}_p[S_i^f].$$

Above, $\mathbb{E}_p[S]$ denotes the expected value of the random variable S under the probability distribution $p := (p_1, p_2, \ldots, p_m)$. So the definition states that the current price/value of every asset in the market, S_i^c, exactly equals the discounted expected price/value in the future, where the expectation is taken with respect to the risk-neutral measure (and the discounting is done at the risk-free interest rate).

Theorem 15 (Asset Pricing Theorem) *A risk-neutral probability measure exists if and only if there is no arbitrage.*

Proof: Consider the following linear program with variables x_i, for $i = 0, \dots, n$:

$$\begin{aligned} \min_x \ & \sum_{i=0}^{n} S_i^c \cdot x_i \\ \text{s.t. } & \sum_{i=0}^{n} S_i^f(\omega_j) \cdot x_i \geq 0, j = 1, \dots, m. \end{aligned} \tag{36}$$

Note that type-A arbitrage corresponds to a feasible solution to this LP with a negative objective value. Since $x = 0$ is a feasible solution in (36), the optimal objective value is always non-positive. Furthermore, since all the constraints are homogeneous, if there exists a feasible solution such that

$$\sum S_i^0 x_i < 0$$

(this corresponds to type-A arbitrage), the problem is unbounded. In other words, there is no type-A arbitrage if and only if the optimal objective value of this LP is 0.

Suppose that there is no type-A arbitrage. Then, there is no type-B arbitrage if and only if all constraints are tight for all optimal solutions of (36) because every state has a positive probability of occurring. Note that any such solution must have objective value 0.

Then, consider the dual of this linear program. Let $p_j, j = 1, \dots, m$, be the dual variables corresponding to the constraints in (36). The dual problem is:

$$\begin{aligned} \max_p \ & 0 \\ \text{s.t. } & \sum_{j=1}^{m} p_j \cdot S_i^f(\omega_j) = S_i^c, \ i = 0, \dots, n, \\ & p_j \geq 0. \end{aligned}$$

When there is no type-A arbitrage, the optimal value in the primal and the dual must be 0, which implies that the dual has a feasible solution p^* (that is also optimal). Moreover, if there is no type-B arbitrage, all the constraints in the primal must be binding, i.e., $\sum_{i=0}^{n} S_i^f(\omega_j) \cdot x_i^* = 0$, for $j = 1, \dots, m$. Because the dual problem satisfies the conditions of the Goldman-Tucker Theorem (Theorem 10), this implies that there exists an optimal dual solution p^* such that $p^* > 0$. From the dual constraint corresponding to $i = 0$, we have that $\sum_{j=1}^{m} p_j^* = \frac{1}{R}$. Multiplying p^* by R, we obtain a risk-neutral probability measure (RNPM), proving that the "no arbitrage" assumption implies the existence of such a measure.

The converse direction is proved in an identical manner. The existence of a RNPM implies that the dual is feasible, and therefore its dual, which is the primal (36), must be bounded, which implies that there is no type-A arbitrage. Furthermore, because we have a strictly feasible (and optimal) dual solution, any optimal solution of the primal must have tight constraints, indicating that there is no type-B arbitrage. □

This seemingly simple result is surprisingly powerful and has some profound implications in the theory of finance, including in proving the Value Additivity Theorem and the Modigliani-Miller Theorem regarding the value of a firm.

7.2 Network Revenue Management

In the airline industry, revenue management (or "yield management", as it is sometimes known) entails setting **booking limits** to control how many tickets of each type are sold.

To fix ideas, consider an airline that is planning operations for a specific day in the future. On that day, the airline operates a set F of direct flights in its (hub-and-spoke) network. For each flight leg (or segment) $f \in F$, you know the capacity of the aircraft flying the leg, c_f. Based on the flights, the airline can offer a large number of "products" (i.e., itineraries) I, where an itinerary refers to an origin-destination-fare class combination. Each itinerary i has a price r_i that is fixed,[5] and requires a seat on several flight legs operated by the airline. We can represent the requirements of all itineraries with a matrix A having one row for each flight leg and one column for each itinerary, with $A_{f,i} = 1$ if itinerary i needs a seat on flight leg f and $A_{f,i} = 0$ otherwise.

Resource matrix A :

	Itinerary 1	Itinerary 2	...	Itinerary $\|I\|$
Flight leg 1	1	0	...	1
Flight leg 2	0	1	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Flight leg $\|F\|$	1	1	...	0

For each itinerary, the airline also has estimates of the demand, given by $d_i \in \mathbb{R}_+$. The goal is to decide how many itineraries of each type to sell to maximize the revenue.

The problem can be formulated as follows. Let x_i denote the number of itineraries of type i that the airline plans to sell, and let x be the vector with components x_i. Similarly, let c denote the vector with components $\{c_f\}_{f \in F}$, r the vector with components $\{r_i\}_{t \in I}$, and d the vector with components $\{d_i\}_{i \in I}$. The revenue maximization problem is:

$$
\begin{aligned}
\max_{x \in \mathbb{R}^I} \; & r^\intercal x \\
& Ax \leq c \\
& x \leq d,
\end{aligned}
$$

where $Ax \leq c$ capture the constraints on plane capacity and $x \leq d$ captures that the planned sales cannot exceed the demand for each itinerary type.

In practice, such an approach that includes **all possible itineraries** would encounter several challenges. First, when considering all the possible origin-destination-fare class combinations, the airline would end up with a gargantuan LP, which might be challenging to solve to optimality. More concerning however, the airline might have very poor estimates for certain itineraries (e.g., for very exotic combinations that might only be flown once or twice in a year...) However, the airline would not want to tell its customers that it is unable to fly them along a specific itinerary, so the airline would still like to sell that, but not include it in its optimization. In this case, the most common approach is to rely on the **shadow prices for the capacity constraints** to construct a valid range of prices.

Specifically, if we let $p \in \mathbb{R}^F$ denote the dual variables corresponding to the capacity constraints $Ax \leq c$, note that:

[5] You can think of having multiple price points even for exactly the same origin-destination-fare class.

- $p \geq 0$;
- at optimality, p_f would be the marginal revenue that the airline would lose if it lost one seat on flight f (equivalently, the revenue it would gain if it gained a seat).

Therefore, consider now an "exotic" itinerary that entails offering a customer a seat of several flights $f \in E$. If the airline were to sell this bundle, the marginal impact on its revenue would be **at least** $\sum_{f \in E} p_f$, because the airline would be losing one seat on each flight $f \in E$. Therefore, the **minimum price** that it should charge a customer for the exotic itinerary is given by the sum of the shadow prices, $\sum_{f \in E} p_f$. (Naturally, the airline would likely add a margin on top of this price. Note that an upper bound on how much it can charge for the exotic itinerary is given by the sum of prices r_f for all the one-leg itineraries the constitute the route — such itineraries are almost always included in the optimization explicitly.)

This heuristic is called the **bid-price heuristic** in network revenue management, but the example showcases a broader principle of how a firm could price its "products." This views products as bundles of resources, and the dual variables/shadow prices allow valuing/pricing these resources when the firm optimally allocates its resources.