ORIE6300 **FA24**

Mathematical Programming I

Notes

Authors: Shaleen Baral

Contents

1. Introduction: LP and Duality

1.1. Why should you take this class?

Mathematical programming is a very well-studied field. Many optimization problems can be recast as or at least well-approximated by mathematical programs. Doing so not only allows us to use algorithms but also use the theory of mathematical programming to uncover structural insights. The course will roughly cover

a. Linear Programming
   - Geometry
   - Duality Theory
   - Algorithms

b. Convex Programming

c. First-Order Methods

1.2. Logistics

Weekly assignments due every Friday night (11:59pm). Grading scheme is as follows:

- 40% HW assignments (~10)
- 20% in-person final
- 15% take-home midterm
- 15% take-home final
- 10% participation/scribing

1.3. Linear Programming

Definition 1.3.1: A *linear program (LP)* is an optimization problem of the following shape

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

where

- $x \in \mathbb{R}^n$ is the *decision variable*;
- $c \in \mathbb{R}^n$ is fixed;
- $c^T x$ is called the *objective function*;
- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are fixed;
- $Ax \leq b$ is called the *constraint.*

Remark: The inequality $Ax \leq b$ is read elementwise i.e. it is equivalent to the system $a_i x \leq b_i, \ \forall i \in [m]$, where a_i denotes the i-th row of A.

Definition 1.3.2: A $x \in \mathbb{R}^n$ is a *feasible solution* if $Ax \leq b$.

Definition 1.3.3: The *feasible region* is the set of all feasible solutions,

$$Q = \{x \in \mathbb{R}^n \mid Ax \leq b\}.$$

Definition 1.3.4: A feasible solution $x \in \mathbb{R}^n$ is *optimal* if its *value*, $c^T x$, is at least as large as the value of any other feasible solution.

Suppose we are given a feasible solution $x \in \mathbb{R}^n$. Is there any way we can certify its optimality? The essential idea is that we want to multiply each constraint $a_i x \leq b_i$ with a $y_i \geq 0$ so that,

$$
\begin{array}{rccl}
 & y_1 & \times & (a_1 x \le b_1) \\
 & y_2 & \times & (a_2 x \le b_2) \\
 & & & \vdots \\
+ & y_m & \times & (a_m x \le b_m) \\
\hline
 & & & c^T x \le y^T b
\end{array}
$$

That is, this problem of certification is equivalent to generating upper bounds for $y^T b$ subject to the constraint $A^T y = c, y \ge 0$. Searching for the best upper bound then motivates the notion of a *dual* in the next definition.

Definition 1.3.5: An LP of the following form is known as the *primal*

$$
\begin{aligned}
\max \quad & c^T x \\
\text{s.t.} \quad & Ax \le b.
\end{aligned}
$$

The *dual* to the *primal* above is defined to be

$$
\begin{aligned}
\min \quad & b^T y \\
\text{s.t.} \quad & A^T y = c \\
& y \ge \mathbf{0}.
\end{aligned}
$$

Proposition 1.3.1 (weak duality): If $x \in \mathbb{R}^n$ feasible in primal and $y \in \mathbb{R}^m$ feasible in dual then $c^T x \le b^T y$.

Proof: We know that $A^T y = c$. Then,

$$c^T x = y^T A x \le y^T b.$$

The last inequality is justified by the fact that $y \ge \mathbf{0}$.

□

In fact, a stronger result holds.

Proposition 1.3.2 (strong duality): If the primal LP or the dual LP has an optimal feasible solution then both have an optimal feasible solutions and their values are equal.

2. Strong Duality and Dual of the Dual

2.1. Strong Duality

We present a not-quite rigorous argument for why strong duality holds. The purpose of this is to just develop our intuition and preview what we will be spending much of the upcoming lectures building up to.

Recall the primal and dual programs from the previous lecture

$$\begin{aligned} \text{primal:} \quad & \max c^T x \qquad \text{s.t.} \qquad Ax \leq b. \\ \text{dual:} \quad & \min b^T y \qquad \text{s.t.} \qquad A^T y = c, \;\; y \geq \mathbf{0}. \end{aligned}$$

First, we interpret our linear program in a "physical" sense. Suppose the decision variable represents the position of a ball that is always acted on by a force, c. We make three observations,

- if we place ball at an optimal feasible solution, it doesn't accelerate
- if the ball doesn't accelerate, the forces acting on it sum to 0
- each *wall*, $a_i x_i \leq b_i$, may exert a force on ball along $-a_i^T$; if $a_i x < b_i$, then this force is 0.

The following lemma formalizes this notion.

Lemma 2.1.1: If $x^\star$ is an optimal feasible solution to the primal, then there exists $y \in \mathbb{R}^n$ such that

a. $y \geq \mathbf{0}$
b. $c + \sum_{i=1}^m (-y_i) a_i^T = 0$
c. $\forall i \in [m], \;\; y_i (a_i x^\star - b_i) = 0,$

The last condition is equivalent to saying $y_i = 0$ whenever $a_i x^\star < b_i$.

Proposition 2.1.1 (strong duality): If $x^\star$ is an optimal feasible solution to the primal, there exists an optimal feasible solution of the dual, $y^\star$, such that $c^T x^\star = b^T y^\star$.

Proof: Assuming $c^T x^\star = b^T y^\star$, we are guaranteed the optimality of $y^\star$ by weak duality.

Now, we actually show that such a feasibly solution must exist. Fix $x^\star$ and let $y^\star$ be the dual solution given by the prior lemma. Then $y^\star$ is dual feasible as the previous lemma guarantees that $y^\star \geq \mathbf{0}$ and $A^T y^\star = c$. By weak duality, we can show that $y^\star$ is optimal by showing that $c^T x^\star = b^T y^\star$. Consider,

$$\begin{aligned} c^T x^\star &= \left(A^T y^\star\right)^T x^\star \\ &= \left(y^\star\right)^T A x^\star \\ &= \left(y^\star\right)^T (b + A x^\star - b) \\ &= \left(y^\star\right)^T b + \left(y^\star\right)^T (A x^\star - b) \\ &= b^T y^\star \end{aligned}$$

The last inequality follows from that fact that if $(a_i x^\star - b_i)$ is non-zero for some $i \in [m]$, then by the prior lemma $y_i = 0$.

□

2.2. Dual of the Dual

Now, we focus on taking the dual of the dual. Particularly, we want to prove the following result.

Proposition 2.2.1: The dual of the dual is the primal.

First, we note that there are two forms in which we can specify a linear program

$$\begin{array}{rlll} \text{basic form:} & \max c^T x & \text{s.t.} & Ax \le b. \\ \text{standard form:} & \min b^T y & \text{s.t.} & A^T y = c,\ y \ge \mathbf{0}. \end{array}$$

Now, we provide two ways of proving this result.

Proof (transform LPs from standard form into basic form, then take the dual):

Consider an LP in standard form

$$\begin{aligned} \min \quad & \overline{c}^T x \\ \text{s.t.} \quad & \overline{A}x = \overline{b} \\ & x \ge \mathbf{0}. \end{aligned}$$

This corresponds to the dual of the primal

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \le b. \end{aligned}$$

under the subtitution $\overline{A} = A^T, \overline{b} = c$ and $\overline{c} = b$.

We may rewrite the standard form LP as a maximization and use inequality constraints.

$$\begin{aligned} \max \quad & -\overline{c}^T x \\ \text{s.t.} \quad & -\overline{A}x \le -\overline{b} \\ & \overline{A}x \le \overline{b} \\ & -Ix \le \mathbf{0}. \end{aligned}$$

Now, we take the dual of this program. We have

$$\begin{aligned} \min \quad & \begin{pmatrix} -\overline{b}^T & \overline{b}^T & 0 & \cdots & 0 \end{pmatrix}^T y \\ \text{s.t.} \quad & \begin{pmatrix} -\overline{A}^T & \overline{A}^T & -I \end{pmatrix} y = -\overline{c}. \end{aligned}$$

For convenience, let $y = \begin{pmatrix} s \\ t \\ w \end{pmatrix}$ such that

$$\begin{pmatrix} -\overline{A}^T & \overline{A}^T & -I \end{pmatrix} y = -\overline{A}^T s + \overline{A}^T t - w.$$

This lets us rewrite our program as

$$\begin{aligned} \max \quad & \overline{b}^T (s - t) \\ \text{s.t.} \quad & \overline{A}^T (s - t) + w = \overline{c} \\ & s, t, w \ge \mathbf{0}. \end{aligned}$$

Making the substitution $z = s - t$, we simplify this expression.

$$\begin{aligned} \max \quad & \overline{b}^T z \\ \text{s.t.} \quad & \overline{A}^T z + w = \overline{c} \\ & w \ge \mathbf{0}. \end{aligned}$$

Finally the requirement that $w \ge \mathbf{0}$ is equivalent to simply saying $\overline{A}^T z \le \overline{c}$. Thus, we have recovered the following program

$$\begin{aligned} \max \quad & \overline{b}^T z \\ \text{s.t.} \quad & \overline{A}^T z \le \overline{c}. \end{aligned}$$

Making the subtitutions $\overline{A} = A^T, \overline{b} = c$ and $\overline{c} = b$, we recover the primal,

$$
\begin{aligned}
\max \quad & c^T x \\
\text{s.t.} \quad & Ax \leq b.
\end{aligned}
$$

□

Proof (directly derive the dual):

In the previous lecture, we derived the dual of an LP in basic form by wanting to bound its objective value. Again, we will derive the dual of the dual by trying to bounds its objective value.

As before, an LP in standard form looks like

$$
\begin{aligned}
\min \quad & \overline{c}^T x \\
\text{s.t.} \quad & \overline{A}x = \overline{b} \\
& x \geq \mathbf{0}.
\end{aligned}
$$

We want to take the linear combinations

$$
\begin{aligned}
& y_1 \times \left(\overline{a}_1 x = \overline{b}_1\right) \\
& y_2 \times \left(\overline{a}_2 x = \overline{b}_2\right) \\
& \vdots \\
+ \quad & y_n \times \left(\overline{a}_n x = \overline{b}_n\right)
\end{aligned}
$$

such that for every $j \in [n]$,

$$
\sum_{j=1}^{n} y_j \overline{a}_{1j} \leq c_j.
$$

This then gives us the lower bound $\overline{c}^T x \geq y^T \overline{b}$. The best bound is then given by the linear program,

$$
\begin{aligned}
\max \quad & y^T \overline{b} \\
\text{s.t.} \quad & \overline{A}^T y \leq \overline{c}.
\end{aligned}
$$

□

The following table succinctly summarizes the relationship between variables in the primal and the dual.

min LP constraints		**variables**		**max LP constraints**
equality	$\longleftrightarrow$	unbounded variables	$\longleftrightarrow$	equality
$\geq$	$\longleftrightarrow$	nonnegative variables	$\longleftrightarrow$	$\leq$
$\leq$	$\longleftrightarrow$	nonpositive variables	$\longleftrightarrow$	$\geq$

3. Polyhedron

3.1. Polyhedrons

Definition 3.1.1: A set $S \subseteq \mathbb{R}^n$ is called a *polyhedron* if $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ for some $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Definition 3.1.2: A set $S \subseteq \mathbb{R}^n$ is *convex* if, for all $x, y \in S$ and all $0 \leq \theta \leq 1$, we have $\theta x + (1 - \theta)x \in S$.

Lemma 3.1.1: Any polyhedron P is convex.

Proof: Suppose $P = \{x \mid Ax \leq b\}$ and let $x, y \in P$. Then for $\theta \in [0, 1]$, consider

$$A \cdot [\theta x + (1 - \theta)x] = \theta Ax + (1 - \theta)Ax \leq \theta b + (1 - \theta)b = b.$$

Thus, $\theta x + (1 - \theta)x \in P$.

□

3.2. Vertices

For the remaining definitions, let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polyhedron.

Definition 3.2.1: A point $x \in P$ is a *vertex* if there exists $c \in \mathbb{R}^n$ such that $c^T x > c^T y$, for all $y \in P \setminus \{x\}$.

Definition 3.2.2: A point $x \in P$ is an *extreme point* if there aren't any $y, z \in P \setminus \{x\}$ and $\theta \in [0, 1]$ such that $x = \theta y + (1 - \theta)z$.

Definition 3.2.3: Consider $x \in P$. Then we call $a_i x \leq b_i$ a *binding constraint* if $a_i x = b$. Otherwise, we call it a *non-binding constraint*.

Definition 3.2.4: For any $x \in P$, we may define the following matrices

$$\begin{aligned} A_= &= \text{the submatrix of } A \text{ containing the rows of binding constraints}, \\ b_= &= \text{the subvector of } b \text{ for the binding rows}, \\ A_< &= \text{the submatrix of } A \text{ containing the rows of non-binding constraints}, \\ b_< &= \text{the subvector of } b \text{ for the non-binding rows}. \end{aligned}$$

Definition 3.2.5: A point $x \in P$ is a *basic feasible solution (BFS)* if $\operatorname{rank}(A_=) = n$.

Proposition 3.2.1: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. The following are equivalient:

a. x is a vertex,
b. x is an extreme point,
c. x is a BFS.

Proof:

- *(a) ⇒ (b)*
  We have $c \in \mathbb{R}^n$ such that $c^T x > c^T y$, for all $y \in P \setminus \{x\}$. Let $y, z \in P \setminus \{x\}$. WLOG assume that $c^T y \geq c^T z$. Then consider,

$$\begin{aligned} c^T(\theta y + (1-\theta)z) &= \theta c^T y + (1-\theta)c^T z \\ &\leq \theta c^T y + (1-\theta)c^T y \\ &= c^T y \\ &< c^T x \end{aligned}$$

Thus, we have that $\theta y + (1-\theta)z \neq x$.

- *(b)* $\Rightarrow$ *(c)*
  We prove the contrapositive, $\neg$ *(c)* $\Rightarrow$ $\neg$ *(b)*.
  Assume $\operatorname{rank}(A_=) < n$. Then there exists $y \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $A_= y = \mathbf{0}$. Note that for any $\varepsilon > 0$ we can conclude that $x \pm \varepsilon y$ satisfies the binding constraints as verified below

  $$A_=(x \pm \varepsilon y) = A_= x \pm \varepsilon A y = A_= x = b.$$

  Now, we show that for an appropriate choice of $\varepsilon > 0$, these vectors also satisfy the non-binding constraints.
  Note that $A_< x < b_<$ and hence, $b_< - A_< x > \mathbf{0}$. Thus, for small enough $\varepsilon > 0$ we have

  $$\pm \varepsilon A_< y < b_< - A_< x.$$

  Choosing ε
  For example, take $\varepsilon = \frac{\zeta}{\|A_< y\|_\infty + 1}$ where $\zeta = \min_{i \in [m]} (b_< - A_< x)_i$.

  Then, this immediately gives us what we want,

  $$A_<(x \pm \varepsilon y) = A_< x \pm \varepsilon A_< y \leq b_<.$$

  So, x can be written as a convex combination of $x \pm \varepsilon y \in P$.

- *(c)* $\Rightarrow$ *(a)*
  Define $I = \{i \in [m] \mid a_i x = b_i\}$. Then, let $c = \sum_{i \in I} a_i^T$. Then,

  $$c^T x = \left(\sum_{i \in I} a_i\right) x = \sum_{i \in I} a_i x = \sum_{i \in I} b_i.$$

  Let $y \in P$. Then,

  $$c^T y = \sum_{i \in I} a_i y \leq \sum_{i \in I} b_i.$$

  If this inequality is tight, that is $c^T y = \sum_{i \in I} b$ then, $a_i y = b_i$. As $\operatorname{rank}(A_=) = n$, $A_= y = b$ has a unique solution, $y = x$. Thus, x is a vertex.

$\square$

4. More Polyhedrons

4.1. Existence of Vertices

Definition 4.1.1: A polyhedron P contains a line if there exists $x \in P, y \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $\{x + \lambda y \mid \lambda \in \mathbb{R}\} \subseteq P$. Otherwise, P is *pointed*.

Remark: Without loss of genarality, we may just consider the half line, $\{x + \lambda y \mid \lambda \geq \mathbf{0}, \lambda \in \mathbb{R}\}$.

Proposition 4.1.1: Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a non-empty polyhedron. The following three are equivalent

a. P has at least one vertex,
b. P is pointed,
c. $\operatorname{rank}(A) = n$.

Proof:

- *(b) ⇒ (a)*
  Consider any point $x \in P$. If $\operatorname{rank}(A_=) = n$, then x is a vertex and we are done. Otherwise, we have $\operatorname{rank}(A_=) < b$. Then there exists $y \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $A_= y = \mathbf{0}$. As P is pointed, we know that the line $\{x + \lambda y \mid \lambda \geq \mathbf{0}, \lambda \in \mathbb{R}\}$ isn't entirely contained in P. That is to say, there exists $\lambda^\star$ such that the line exits P at $x + \lambda^\star y$. Particularly, at this specific $\lambda^\star$ a new constraint $j \in [m]$ becomes tight: $a_j x < b_j$ but $a_j(x + \lambda^\star y) = b_j$.

  A More Rigorous Argument for $\lambda^\star$
  There exists λ' is such that $x + \lambda' y \notin P$. Let $I \subseteq [m]$ such that for $i \in I$,

  $$a_i(x + \lambda' y) > b_i.$$

  Since $x + \lambda' y \notin P$, I must be non-empty. Note that for every $i \in I$, $\lambda \mapsto a_i(x + \lambda y)$ is continuous and by the Intermediate Value Theorem, there exists $\lambda_i \in (0, \lambda]$ such that $a_i(x + \lambda_i y) = b_i$. Then take $\lambda^\star = \min_{i \in I} \lambda_i$.

  Let $x' = x + \lambda^\star y$. Note that the prior equality constraints are still tight,

  $$A_= x' = A(x + \lambda^\star y) = A_= x + \lambda^\star A_= y = b_=.$$

  Let $A'_=$ be the submatrix of A containing the rows of binding constraints for x'. We know that $A'_=$ contains all the rows of $A_=$ and at least an addition row a_j corresponding to a previous non-binding constraint, $a_j x < b_j$, becoming tight, $a_j(x + \lambda^\star y) = b_j$.
  Furthermore, we claim that a_j is linearly independent of the rows of $A_=$. If a is a row of $A_=$, then $ay = 0$. Note that

  $$a_j y > \frac{b_j - a_j x}{\lambda^\star} > 0.$$

  That is, a_j is not a linear combination of the rows of $A_=$. Thus, $\operatorname{rank}({A_=}') > \operatorname{rank}(A_=)$. Repeating this argument, by induction, we can find a basic feasible solution.

- *(a) ⇒ (c)*
  Suppose $x \in P$ is a vertex. Then,

  $$n = \operatorname{rank}(A_=) \leq \operatorname{rank}(A) \leq n.$$

  Thus, $\operatorname{rank}(A) = n$.

- *(c) ⇒ (b)*
  Assume $\operatorname{rank}(A) = n$. Suppose P contains a line $\{x + \lambda y \mid \lambda \in \mathbb{R}\} \subseteq P$ for $y \in \mathbb{R}^n$. We will show that we must have $y = 0$.
  Let $j \in [m]$. Then we know that for all $\lambda \in \mathbb{R}$, $a_j(x + \lambda y) \leq b_j$ holds. Taking the limit as $\lambda \to \operatorname{sgn}(a_j y) \cdot \infty$, we get that $a_j(x + \lambda y) \to \infty$ if $a_j y \neq 0$. Thus, it must be the case that $a_j y = 0$. By injectivity of A, we note that $y = \mathbf{0}$. Thus, it must be the case that P is pointed.

□

Definition 4.1.2: Let B be a normed space. A set $X \subseteq B$ is called bounded whenever there exists $M \in \mathbb{R}$, such that $\|x\| \leq M$ for all $x \in X$.

Corollary 4.1.1.1: If a polyhedron P is nonempty and bounded then P has at least one vertex.

Proof: A bounded bolyhedron cannot contain a line. Consider a line $\lambda \mapsto x + \lambda y$, for $y \neq \mathbf{0}$. Then, for some $M \in \mathbb{R}$, points of the line contained in P must satisfy

$$\begin{aligned} |\lambda|\|y\| - \|x\| &\leq \|x + \lambda y\| \leq M \\ \Longrightarrow |\lambda| &\leq \frac{M + \|x\|}{\|y\|}. \end{aligned}$$

□

Corollary 4.1.1.2: The feasible region of an LP in standard form, if it is nonempty, has at least one vertex.

Proof: The feasible region of a standard form LP is a subset of the nonnegative orthant $\mathbb{R}^n_{\geq 0}$ which does not contain a line.

□

An important result is that we can achieve optimality at a vertex.

Proposition 4.1.2 (fundamental theorem of linear programming):
Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a nonempty, pointed polyhedron. Suppose we are also given an LP

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & x \in P. \end{aligned}$$

If an optimal feasible solution exists, there is an optimal feasible solution that is a vertex of P.

Proof: Define $Q = P \cap \{x \in \mathbb{R}^n \mid c^T x = c^T x^\star\}$. Note that Q is a nonempty polyhedron as $x^\star \in Q$. Furthermore, Q is pointed because $Q \subseteq P$ and P is pointed.

By Proposition 4.1.1, there exists a vertex v of Q. Note that v is feasible and $c^T v = c^T x^\star$, so v is optimal. Suppose that v is not an extreme point of P. That is to say, there exists $y, z \in P \setminus \{v\}$ and $\theta \in [0, 1]$ such that

$$v = \theta y + (1 - \theta) z.$$

Then,

$$c^T x^\star = c^T v = c^T(\theta y) + (1 - \theta) z = \theta c^T y + (1 - \theta) c^T z \leq c^T x^\star.$$

Note that $c^T y = c^T z = c^T x^\star$ as otherwise $c^T x^\star < c^T x^\star$, a clear contradiction. Thus, we have $y, z \in Q \setminus \{v\}$. However, this would imply that v is not an extreme point of Q— a contradiction! Thus, v is a vertex of P too.

□

4.2. Convex Hull

Definition 4.2.1: Given $v_1, ..., v_k \in \mathbb{R}^n$, a *convex combination* is any $\sum_{i=1}^k \lambda_i v_i$ where $\lambda_i \geq 0$ for all $i \in [k]$ and $\sum_{i=1}^k \lambda_i = 1$.

Definition 4.2.2: Given $v_1, ..., v_k \in \mathbb{R}^n$, their *convex hull* is the set of all their convex combinations,

$$\operatorname{conv}(\{v_1, ..., v_k\}) = \left\{ \sum_{i=1}^k \lambda_i v_i \mid \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1 \right\}.$$

Definition 4.2.3: Any $S \subseteq \mathbb{R}^n$ that can be written as $S = \operatorname{conv}(\{v_1, ..., v_k\})$ for some $v_1, ..., v_k$ is called a *polytope*.

Proposition 4.2.1 (Carathéodory's Theorem):
Suppose $v_1, ..., v_k \in \mathbb{R}^n$ and $y \in \operatorname{conv}(\{v_1, ..., v_k\})$. Then, there exists $S \subseteq \{v_1, ..., v_k\}, |S| \leq n + 1$ such that $y \in \operatorname{conv}(S)$.

Proof: Assume, without loss of generality, let $k > n + 1$ as otherwise the proposition holds trivially due to linear dependence.

Define

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n \times k} \quad \text{and} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix} \in \mathbb{R}^n.$$

Then the following polyhedron represents all the ways in which we may write y as a convex combination of $v_1, ..., v_k$,

$$A\lambda = y, \tag{1}$$

$$\mathbf{1} \cdot \lambda = 1 \tag{2}$$

$$y \geq \mathbf{0}. \tag{3}$$

Note that this polyhedron is nonempty as $y \in \operatorname{conv}(\{v_1, ..., v_k\})$. By Corollary 4.1.1.2, there exists some vertex $\lambda^\star$ of this polyhedron. Furthermore, $\lambda^\star$ is a BFS. Thus, $\operatorname{rank}(A_=) = k$. The constraints corresponding to equation (1) and (2) have rank at most $n + 1$. So, at least $k - (n + 1)$ many constraints from (3) must be binding. That is, at least $k - (n + 1)$ entries of $\lambda^\star$ are 0. Let $S = \{i \in [k] \mid \lambda_i^\star > 0\}$.

Note that $|S| = k - (k - (n + 1)) = n + 1$ and

$$y = \sum_{i \in S} \lambda_i^\star v_i,$$

as desired.

□

5. Bounded Polyhedra and Polytopes I

5.1. Bounded Polyhedra $\subseteq$ Polytope

Proposition 5.1.1: If P is a bounded polyhedron, then $P = \operatorname{conv}(\{v_1, ..., v_k\})$, where $v_1, ..., v_k$ are the vertices of P.

Proof:

- $\operatorname{conv}(\{v_1, ..., v_k\}) \subseteq P$
  Note that P is convex and contains $v_1, ..., v_k$. Thus, $\operatorname{conv}(\{v_1, ..., v_k\}) \subseteq P$.
- $P \subseteq \operatorname{conv}(\{v_1, ..., v_k\})$
  Let $x \in P$, we want to show that

  $$\exists \lambda_1, ..., \lambda_k \geq 0, \sum_{i=1}^{k} \lambda_i = 1.\ \ x = \sum_{i=1}^{k} \lambda_i v_i.$$

  We proceed by induction on $n - \operatorname{rank}(A_=)$.
  Base Case. When $\operatorname{rank}(A_=) = n$, x is a basic feasible solution of P. Thus, $x = v_i$ for some $i \in [k]$. Then the claim follows from setting $\lambda_i = 1$ and $\lambda_j = 0$ for all $j \neq i$.
  Induction Hypothesis. We assume that any $x' \in P$ for which the binding constraints have a rank strictly higher than $\operatorname{rank}(A_=)$ can be written as a convex combination of $\{v_i\}_{i \in [k]}$.
  Inductive Step. We know that $\operatorname{rank}(A_=) < n$. Thus, tehre exists $y \in \mathbb{R}^n \setminus \{0\}$ such that $A_= y = 0$. Consider the line $\{x + \alpha y \mid \alpha \in \mathbb{R}\}$. Since P is bounded, we know that there exists $\alpha^- < 0$ and $\alpha^+ > 0$ such that that $x + \alpha z \in P$ if and only if $\alpha^- \leq \alpha \leq \alpha^+$. Furthermore, $x + \alpha^+ y, x + \alpha^- y$ have at least one more binding constraints than x. So, their binding constraints have rank strictly greater than $\operatorname{rank}(A_=)$. By the induction hypothesis, there are $\lambda_i^+, \lambda_i^- \geq 0$ with $\sum_{i=1}^{k} \lambda_i^+ = 1 = \sum_{i=1}^{k} \lambda_i^-$ such that

  $$x + \alpha^+ y = \sum_{i=1}^{k} \lambda_i^+ v_i, \quad x + \alpha^- y = \sum_{i=1}^{k} \lambda_i^- v_i.$$

  Consider the following convex combination,

  $$x = \left(\frac{-\alpha^-}{\alpha^+ - \alpha^-}\right)(x + \alpha^+ y) + \left(\frac{\alpha^+}{\alpha^+ - \alpha^-}\right)(x + \alpha^- y).$$

  Thus, x is a convex combination of $x + \alpha^+ y, x + \alpha^- y \in \operatorname{conv}(\{v_1, ..., v_k\})$. By convexity of the convex hull, $x \in \operatorname{conv}(\{v_1, ..., v_k\})$.

$\square$

5.2. Separating Hyperplane Theorem

We now discuss a sufficient condition for separating a set from a point by a hyperplane.

We start by proving the Extreme Value Theorem.

Proposition 5.2.1 (Weierstrass' Exteme Value Theorem):
Let $C \subseteq \mathbb{R}^n$ be a compact. Let $f : C \to \mathbb{R}$ be continuous. Then there exists $x \in C$ such that $f(x) \leq f(x')$ for all $x' \in C$.

Proof: As f is continuous, $f(C)$ compact and hence, closed. Since $\inf_x f(x)$ is a limit point of C, it must be contained in $f(C)$. Hence, there exists $x \in C$ such that $f(x) = \inf_x f(x) \leq f(x')$ for all $x' \in C$.

$\square$

Now, we move to the main result.

Proposition 5.2.2 (Separating Hyperplane Theorem):
Suppose $C \subseteq \mathbb{R}^n$ is nonempty, closed and convex. Suppose $y \notin C$. Then, there exist $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that $a^T y > b > a^T x$.

Proof: Consider $f : C \to \mathbb{R}$ defined by $x \mapsto \|x - y\|^2$. Fix some $q \in C$. Define

$$\hat{C} = \{x \in C \mid \|x - y\| \leq \|q - y\|\}.$$

Note that $\hat{C}$ is nonempty as $q \in \hat{C}$. Furthermore, $\hat{C}$ is closed and bounded being the intersection of C with the closed ball of radius $\|q - y\|$ centered at y.
As $\hat{C}$ is compact, we can choose a $z \in \operatorname{argmin}_{x \in \hat{C}} f(x)$. Furthermore, for $x \in C \setminus \hat{C}$, we have $f(x) > f(q) \geq f(z)$ as

$$\|x - y\| > \|q - y\| \geq \|z - y\|.$$

Thus, $z \in \operatorname{argmin}_{x \in C} f(x)$ too. We then set $a = y - z, b = \frac{1}{2}(a^T y + a^T z)$. Note that as $y \notin C$ and $z \in C$, we have $a \neq 0$. Note then that z, y are separated by a as

$$a^T y - a^T z = a^T (y - z) = a^T a > 0.$$

Furthermore, $a^T y > b$ and $a^T z < b$ as

$$\begin{aligned} a^T y - b &= \frac{1}{2}(a^T y - a^T z) > 0, \\ a^T z - b &= \frac{1}{2}(a^T z - a^T y) < 0. \end{aligned}$$

All we have to show now is that $a^T x < b$ for all $x \in C$. This is equivalent to showing $a^T (z - x) \geq 0$ for all $x \in C$. We prove this now. Fix an $x \in C$ and consider $x_\theta = \theta x + (1 - \theta) z$ for $\theta \in (0, 1]$. Then,

$$f(x_\theta) = \|x_\theta - y\|^2 = \|(z - y) + \theta(x - z)\|^2.$$

By convexity of C, we have $x_\theta \in C$. Thus, $f(x_\theta) \geq f(z) = (z - y)^T (z - y)$. Substituting the expression above into this we get,

$$\begin{aligned} &\theta^2 \|x - z\|^2 + 2\theta(x - z)^T (z - y) \geq 0 \\ \Longrightarrow\ &\theta \|x - z\|^2 + 2(x - z)^T a \geq 0 \\ \Longrightarrow\ &(z - x)^T a \geq -\frac{\theta}{2} \|x - z\|^2. \end{aligned}$$

This inequality hods for all $\theta > 0$. Taking the limit as $\theta \to 0$, we have $a^T (z - x) \geq 0$.

□

6. Bounded Polyhedra and Polytopes II

6.1. Polars

Definition 6.1.1: Let $S \subseteq \mathbb{R}^n$. Then, the *polar of S* is the set

$$S^\circ = \{z \in \mathbb{R}^n \mid z^T x \leq 1,\ \forall x \in S\}.$$

Lemma 6.1.1: If $C \subseteq \mathbb{R}^n$ is closed, convex, and contains $\mathbf{0}$, then $C^{\circ\circ} = C$.

Proof:
- $C \subseteq C^{\circ\circ}$
  Fix $x \in C$. We need to show that $x^T z \leq 1$, for all $z \in C^\circ$. This follows from the fact that, as $z \in C^\circ$, we have $x^T z = z^T x \leq 1$.
- $C^{\circ\circ} \subseteq C$
  Assume for contradiction that $x \in C^{\circ\circ}$ but $x \notin C$. Since $\mathbf{0} \in C$, C is nonempty. By assumption C is closed and convex. By the Separating Hyperplane Theorem, we obtain $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $a^T x > b > a^T y$ for all $y \in C$. Since $\mathbf{0} \in C$, it follows that $b > \mathbf{0}$.
  Let $\tilde{a} = a/b$ so that $\tilde{a}^T x > 1 > \tilde{a}^T y$ for all $y \in C$. The second inequality implies that $\tilde{a} \in C^\circ$. As $x \in C^{oo}$, we must have $x^T \tilde{a} \leq 1$. This contradicts $x^T \tilde{a} > 1$.
  Thus, $x \in C$ and hence $C^{\circ\circ} \subseteq C$.

□

Lemma 6.1.2: The polar of a polytope is a polyhedron.

Proof: Let $P = \operatorname{conv}(\{v_1, ..., v_k\})$. We claim that $P^\circ = \{z \in \mathbb{R}^n \mid z^T v_i \leq 1, \forall i \in [k]\}$.
- ($\subseteq$). Note that $P^\circ \subseteq \{z \in \mathbb{R}^n \mid z^T v_i \leq 1, \forall i \in [k]\}$. This is because an element polar must satisfy all the constraints of the set on the right, and more.
- ($\supseteq$). Fix $z \in \mathbb{R}^n$ such that $z^T v_i \leq 1$ for all $i \in [k]$. Any $x \in P$ can be written as a convex combination $\sum_{i=1}^k \lambda_i v_i$. Then,

$$z^T x = z^T \left(\sum_{i=1}^k \lambda_i v_i \right) = \sum_{i=1}^k \lambda_i \underbrace{z^T v_i}_{\leq 1} \leq \sum_{i=1}^k \lambda_i = 1.$$

  Thus, $z \in P$.

□

Lemma 6.1.3: If $\mathbf{0}$ is in the interior of S, then S° is bounded.

Proof: By assumption, there exists $\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ satisfying $\|x\| \leq \varepsilon$, we have $x \in S$. Let's fix some $z \in S^\circ \setminus \{\mathbf{0}\}$ and we claim that $\|z\| \leq \frac{1}{\varepsilon}$.
Let $x = \frac{\varepsilon}{\|z\|} z$. Since $\|x\| = \varepsilon$, we have $x \in S$. Since $z^T x \leq 1$, we have

$$z^T x = z^T \left(\frac{\varepsilon}{\|z\|} z \right) = \varepsilon \|z\| \leq 1.$$

Thus, $\|z\| \leq \frac{1}{\varepsilon}$ and S° is bounded.

□

6.2. Polytopes $\subseteq$ Bounded Polyhedra

We start by showing that all polytopes are bounded.

Lemma 6.2.1: Any polytope is closed and bounded (equiv. compact).

Proof: First, define the simplex

$$\Delta_n = \left\{ \lambda \in \mathbb{R}^k \mid \lambda \geq \mathbf{0}, \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

Note that this is closed, being the intersection of the nonnegative orthant $\mathbb{R}^n_+$ and the hyperplane $H = \left\{ \lambda \in \mathbb{R}^k \mid \sum_{i=1}^{k} \lambda_i = 1 \right\}$, both of which are closed. Furthermore, Δ_n is bounded as it is a subset of $[0, 1]^n$.

Next, the map $f : \Delta_n \to \mathbb{R}^n$ defined by $\lambda \mapsto \sum_{i=1}^{k} \lambda_i v_i$ is continuous. Thus, the polytope $P = \operatorname{conv}(\{v_1, ..., v_k\})$ is the continuous image $g(\Delta_n)$ of Δ_n. So, P must be compact too.

□

We shall prove the following result in this lecture.

Proposition 6.2.1: Any polytope that contains $\mathbf{0}$ in its interior is a bounded polyhedron.

Proof: Let P be our polytope. It is enough, by Lemma 6.1.1, to show that $P^{\circ\circ}$ is a bounded polyhedron. By Lemma 6.1.2, we note that P° is a polyhedron. Since P contains $\mathbf{0}$, P° is a bounded polyhedron. Thus, P° is a polytope and $P = P^{\circ^{\circ}}$ is a polyhedron. Finally, P is bounded by the convexity of the Euclidean norm.

□

7. Farkas' Lemma and Infeasibility

7.1. Farkas' Lemma

We are interested in certifying the feasibility of LP problems. This naturally leads us to Farkas' Lemma. Before that, we introduce a new construction that will be useful for us.

Definition 7.1.1 (cone): For a set of vectors $\{v_1, ..., v_k\}$, we define their *cone* to be

$$\mathrm{cone}(\{v_1, ..., v_k\}) = \left\{ \sum_{i=1}^{k} \lambda_i v_i \mid \lambda_i \geq 0 \right\}.$$

Lemma 7.1.1: For any $v_1, ..., v_k \in \mathbb{R}^n$, the $\mathrm{cone}(\{v_1, ..., v_k\})$ is convex.

Proof: Suppose $p, p' \in Q$ and $\theta \in [0, 1]$. Then, $p = \sum_{i=1}^{k} \lambda_i v_i$ and $q = \sum_{i=1}^{k} \lambda_{i'} v_i$ for $\lambda_i, \lambda_{i'} \geq 0$. The following then proves the convexity of Q,

$$\theta p + (1 - \theta) q = \sum_{i=1}^{k} \underbrace{[\theta \lambda_i + (1 - \theta) \lambda_i']}_{\geq 0} \cdot v_i \in Q.$$

□

Lemma 7.1.2: For any $v_1, ..., v_k \in \mathbb{R}^n$, the $\mathrm{cone}(\{v_1, ..., v_k\})$ is closed.

Proof: At some point, do Exercise 4.37 in BT97. □

Now, we move to the main result.

Proposition 7.1.1 (Farkas' Lemma I):
Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$. Exactly one of the following holds

a. $\exists x \in \mathbb{R}^n.\ Ax = b, x \geq \mathbf{0}$;
b. $\exists y \in \mathbb{R}^m.\ A^T y \geq \mathbf{0}, b^T y < 0$.

Proof:

- (a) and (b) cannot both hold.
  For contradiction, suppose we have such an x and y. Consider the following linear programs that are duals of one another,

$$\begin{array}{llll} \text{(P)} & \max & \mathbf{0}^T x & \\ & \text{s.t.} & Ax = b & \\ & & x \geq \mathbf{0} & \end{array} \qquad \begin{array}{lll} \text{(D)} & \min & b^T y \\ & \text{s.t.} & A^T y \geq \mathbf{0} \end{array}$$

  Then x is a feasible solution of (P) with value 0 and y is a feasible solution of D with value $b^T y < 0$. This contradicts weak duality.

- $\neg(a) \Rightarrow (b)$.
  Suppose (a) doesn't hold. Then, we define

$$A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \qquad \text{and} \qquad Q = \mathrm{cone}(\{v_1, ..., v_n\}).$$

  By assumption, note that $b \notin Q$. Furthermore, Q is nonempty, convex and closed. Now, the Separating Hyperplane Theorem yields $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}$ such that

$$\alpha^T b > \beta > \alpha^T p. \qquad \forall p \in Q$$

with $\alpha \neq \mathbf{0}$. Furthermore, $\beta > 0$ as $\mathbf{0} \in Q$. Note that for $y = -\alpha$, we have

$$b^T y = -b^T \alpha = -\alpha^T b < 0.$$

Now, we show that $A^T y \geq \mathbf{0}$. That is, for all $i \in [n]$, we want to show that $v_i^T y \geq 0$ (equiv. $v_i^T \alpha \leq 0$). Fix some $i \in [n]$. Then $\lambda v_i \in Q$ for all $\lambda \geq 0$. Thus, for all $\lambda \geq 0$, we also have

$$\alpha^T (\lambda v_i) < \beta.$$

Assume for contradiction that $v_i^T \alpha > 0$. We choose $\lambda = \frac{\beta}{v_i^T \alpha}$, then

$$\beta = \lambda v_i^T \alpha < \beta.$$

This is a contradiction! Thus, $v_i^T \alpha \leq 0$.

□

The following variant of Farkas' Lemma will turn out to be useful.

Proposition 7.1.2 (Farkas' Lemma II):
Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Then exactly one of the following holds,

a'. $\exists x \in \mathbb{R}^n.\ Ax \leq b$,

b'. $\exists y \in \mathbb{R}^m.\ A^T y = \mathbf{0}.\ y \geq \mathbf{0}, b^T y < 0$.

Furthermore, (b') is equivalent to (b''),

b''. $\exists y \in \mathbb{R}^n.\ A^T y = \mathbf{0}, y \geq \mathbf{0}, b^T y = -1$.

Proof:

- $(b') \iff (b'')$
  Note that $(b'') \Rightarrow (b')$ as $b^T y = -1 < 0$.
  For $(b') \Rightarrow (b'')$, let $y \in \mathbb{R}^m$ such that $A^T y = \mathbf{0}, y \geq \mathbf{0}, b^T y < 0$. Then, define

$$y' = -\frac{1}{b^T y} y.$$

  Indeed,

$$\begin{aligned} A^T y' &= -\frac{1}{b^T y} A^T y = \mathbf{0}, \\ y' &= -\frac{1}{b^T y} y \geq \mathbf{0}, \\ b^T y' &= -\frac{1}{b^T y} b^T y = -1. \end{aligned}$$

- (a') and (b') cannot both hold.
  For contradiction, suppose we have such an x and y. Consdier the following linear programs that are duals of one another,

$$\begin{array}{llll} \text{(P)} & \max & \mathbf{0}^T x & \\ & \text{s.t.} & Ax \leq b & \end{array} \qquad \begin{array}{lll} \text{(D)} & \min & b^T y \\ & \text{s.t.} & A^T y = \mathbf{0} \\ & & y \geq \mathbf{0} \end{array}$$

  Then x is a feasible solution to (P) with value 0 whereas y is a feasible solution to (D) with value $b^T y < 0$. This contradicts weak duality.
- $\neg(b') \Rightarrow (a')$.
  Suppose $\neg(b')$ so that the following is infeasible,

$$\begin{pmatrix} A^T \\ b^T \end{pmatrix} y = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}$$
$$y \geq \mathbf{0}.$$

By Farkas' Lemma, there exists $z \in \mathbb{R}^{n+1}$ such that

$$(A \;\; b)\, z \geq \mathbf{0}, \qquad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}^T z < 0.$$

Let $\binom{x}{\lambda} = z$ be such that

$$Ax + b\lambda \geq \mathbf{0}.$$

Furthermore, $-\lambda < 0$ implies that $\lambda > 0$. So, we define $x' = -\frac{x}{\lambda}$. Then,

$$Ax' = -\frac{1}{\lambda}Ax \leq -\frac{1}{\lambda}(-\lambda b) = b.$$

Thus, (a') is true.

□

7.2. Infeasibility

Definition 7.2.1: Let $f(x)$ be a function and consider the mathematical program,

$$\begin{aligned} \max \quad & f(x) \\ \text{s.t.} \quad & x \in X. \end{aligned}$$

Then, the *optimal value* of this program is

$$\sup\{f(x) \mid x \in X\}.$$

Remark: A few notes on this definition,
- If $X = \emptyset$, the program is infeasible and the optimal value is $\sup \emptyset = -\infty$.
- If the program is unbounded, its optimal value is ∞.
- If a mathematical program has finite optimal value, then it may or may not be an optimal soluton. For example, $\sup_{x \in \mathbb{R}} e^{-x} = 0$ but no feasible solution acquires this value.

What can we say about the feasibility of linear programming problems then?

$$\text{(P)} \quad \begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned} \qquad\qquad \text{(D)} \quad \begin{aligned} \min \quad & b^T y \\ \text{s.t.} \quad & A^T y = \mathbf{0} \\ & y \geq \mathbf{0} \end{aligned}$$

Let v_p be the optimal value of (P) and v_d the optimal value of (D). By weak duality, $v_p \leq v_d$. Consider the following table that considers all possibilities for v_p, v_d. We cross out all the combinations eliminated by weak duality.

v_d

v_p	∞ (infeasible)	$\in \mathbb{R}$ (no opt. solution)	$\in \mathbb{R}$ (with opt. solution)	$-\infty$ (unbounded)
$-\infty$ (infeasible)				

$\in \mathbb{R}$ (no opt. solution)				×
$\in \mathbb{R}$ (with opt. solution)				×
∞ (unbounded)		×	×	×

8. Strong Duality I

The central aim is to prove the following,

Proposition 8.1 (Strong Duality): For the following primal and dual linear programs,

$$\begin{array}{llll} \text{(P)} & \max & c^T x \\ & \text{s.t.} & Ax \leq b \end{array} \qquad \begin{array}{llll} \text{(D)} & \min & b^T y \\ & \text{s.t.} & A^T y = c \\ & & y \geq \mathbf{0}, \end{array}$$

exactly one of the following holds

a. The primal is unbounded and the dual is infeasible
b. The primal is infeasible and the dual is unbounded
c. Both primal and dual are infeasible
d. Both primal and dual have optimal feasible solutions with equal value

Proof: We proceed by casework on the optimal value of the dual, v_d

- $v_d = -\infty$
  Then, by weak duality, the primal is infeasible.
- $v_d = \infty$
  Here, the dual is infeasible. That is, $A^T y = c$ with $y \geq \mathbf{0}$ is infeasible. By Farkas' Lemma, we know that there exists $z \in \mathbb{R}^n$ with $Az \geq 0$ and $c^T z < 0$. Now, we claim that if there exists any primal feasible solution, the primal is unbounded. Suppose $x \in \mathbb{R}^n$ is primal feasible. Then consider $x - \lambda z$ for any $\lambda \geq 0$. This is feasible as,

  $$A(x - \lambda z) = Ax - \lambda Az \leq b + \mathbf{0} = b.$$

  Next, consider the objective,

  $$c^T (x - \lambda z) = c^T x - \lambda c^T z.$$

  Note that as $-c^T z > 0$, the primal objective becomes unbounded as $\lambda \to \infty$.
- $v_d \in \mathbb{R}$
  We want to find a primal feasible solution with value at least v_d as that would imply that we have found an optimal solution too. This is equivalent to asking whether the following inequalities have a solution

  $$\begin{aligned} Ax &\leq b \\ -c^T x &\leq -v_d \end{aligned}$$

  By Farkas' Lemma, we have two cases
  - there exists $x \in \mathbb{R}^n$ with $Ax \leq b, -c^T x \leq -v_d$.
    In this case, we have found a feasible solution with value at least v_d and by weak duality, this is an optimal feasible solution.
  - there exists $z \in \mathbb{R}^{m+1}$ with $\begin{pmatrix} A^T & -c \end{pmatrix} z = \mathbf{0}, z \geq \mathbf{0}, \begin{pmatrix} b & -v_d \end{pmatrix}^T y < 0$.
    Let $z = \begin{pmatrix} y \\ \lambda \end{pmatrix}$ such that $y \geq \mathbf{0}, \lambda \geq 0$ and

    $$\begin{aligned} A^T y - c\lambda &= 0 \\ b^T y - v_d \lambda &< 0. \end{aligned} \tag{4}$$

    Now, we show that $\lambda > 0$. Suppose for contradiction, $\lambda = 0$. Then, $A^T y = \mathbf{0}, b^T y < 0$, and $y \geq \mathbf{0}$. Let $\hat{y}$ be a feasible solution of the dual. Consider $\hat{y} + \lambda' y$ for all $\lambda' \geq 0$. This is feasible as $\hat{y} + \lambda' y \geq \mathbf{0}$ (all terms are nonnegative) and

    $$A^T (\hat{y} + \lambda' y) = A^T \hat{y} + \lambda' A^T y = c + \mathbf{0} = c.$$

    Consider the dual objective,

$$b^T(\hat{y} + \lambda' y) = b^T\hat{y} + \lambda' b^T y.$$

As $\lambda \to \infty$, since $b^T y < 0$, the above expression goes to $-\infty$. Therefore, the dual is unbounded. However, this contradicts the fact that $v_d \in \mathbb{R}$!
Hence, $\lambda > 0$. We may then rewrite (4) as

$$A^T\left(\frac{y}{\lambda}\right) = c,$$

with $y/\lambda \geq \mathbf{0}$ and $b^T(y/\lambda) < v_d$. This contradicts the assumption that the optimal value of the dual is v_d. Thus, this case cannot occur!

We have shown that whenever $v_d \in \mathbb{R}$, the primal has an optimal feasible solution and $v_p = v_d$. Note that the dual of the dual is the primal. So, symmetrically, we also have that whenever $v_p \in \mathbb{R}$, the dual also has an optimal feasible solution with $v_d = v_p$.

□

We finally completely fill up the table that considers all the possibilities for v_p, v_d. We cross out all the combinations eliminated by weak duality and strong duality.

v_d

v_p	∞ (infeasible)	$\in \mathbb{R}$ (no opt. solution)	$\in \mathbb{R}$ (with opt. solution)	$-\infty$ (unbounded)
$-\infty$ (infeasible)		×	×	
$\in \mathbb{R}$ (no opt. solution)	×	×	×	×
$\in \mathbb{R}$ (with opt. solution)	×	×		×
∞ (unbounded)		×	×	×

9. Strong Duality II

9.1. A Retrospective and Alternative Proof

We present an alternative proof for the following result that we chalked up to the idempotence of taking duals. In fact, the proof we recover is pretty much a more rigorous look at the *intuitively-motivated* proof of strong duality we gave in Chapter 2.

Lemma 9.1.1: If $x^\star$ is an optimal feasible solution to the primal $\max(c^T x \mid Ax \leq b)$, there is an optimal feasible solution to the dual with matching value.

The following lemma was "physically-motivated" in Chapter 2. Now, equipped with all the appropriate tools, we provide a mathematical proof for it!

Lemma 9.1.2: If $x^\star$ is an optimal feasible solution to the primal, there exists $y \in \mathbb{R}^m$, such that

$$\begin{gathered} y \geq \mathbf{0} \\ \sum_{i=1}^{m} y_i(-a_i^T) + c = 0 \\ y_i(b_i - a_i x^\star) = 0, \ \forall i \in [m]. \end{gathered}$$

Proof: Fix $x^\star$. Consider $A_=$. WIthout loss of generality, let $A_=$ just be the first m' rows of A. Consider the following linear ineqaulities

$$\exists d \in \mathbb{R}^n . A_= d \leq \mathbf{0}, c^T d \geq 1.$$

We apply Farkas' Lemma to get two cases

- such a d exists.
  Then, $x^\star + \varepsilon d$, for sufficiently small $\varepsilon > 0$, is feasible in the primal and whose objective value is

  $$c^T(x^\star + \varepsilon d) = c^T x^\star + \varepsilon c^T d > c^T x^\star.$$

  This contradicts the optimality of $x^\star$.
- such a d does not exist.
  Then there exists $z \in \mathbb{R}^{m'}, \lambda \in \mathbb{R}$ such that

  $$A_=^T z - \lambda c = \mathbf{0}$$

  for $\lambda > 0$ and $z \geq \mathbf{0}$. Thus,

  $$\begin{aligned} & A_=^T \left(\frac{z}{\lambda}\right) - c = \mathbf{0} \\ \Longrightarrow & -A_=^T \frac{z}{\lambda} + c = \mathbf{0} \\ \Longrightarrow & \sum_{j=1}^{m'} \left(\frac{z_j}{\lambda}\right)(-a_j^T) + c = 0 \end{aligned}$$

  Define

  $$y_j = \begin{cases} \frac{z_i}{\lambda} & \text{if } j \leq m', \\ 0 & \text{otherwise.} \end{cases}$$

  Note that $y \geq 0$. By our prior computation,

  $$\sum_{j=1}^{m} y_j(-a_j^T) + c = 0.$$

Finally, for all $j \in [m]$, we also have $y_j(b_j - a_j\lambda^\star) = 0$. First, note that for all $j \leq m'$, the constraint is binding at $x^\star$, i.e. $b_j = a_j x^\star$. Second, for all $m' < j \leq m$, $y_j = 0$ by definition.

□

9.2. Covering and Packing

Definition 9.2.1: We have a collection of entities $[n] = \{1, ..., n\}$ and subsets $S_1, ..., S_m \subseteq [n]$. In the *covering problem*, we want to choose $X \subseteq [n]$ such that

a. $X \cap S_j \neq \emptyset$ for every $j \in [m]$,
b. $|X|$ is minimized.

We can represent this problem as an integer program

$$\begin{aligned} \min \quad & \sum_{i \in [n]} x_i \\ \text{s.t.} \quad & \sum_{i \in S_j} x_i \geq 1, \forall j \in [m], \\ & x_i \in \{0, 1\}, \forall i \in [n]. \end{aligned}$$

We relax this into the following linear program

$$\begin{aligned} \min \quad & \sum_{i \in [n]} x_i \\ \text{s.t.} \quad & \sum_{i \in S_j} x_i \geq 1, \forall j \in [m], \\ & x_i \geq 0, \forall i \in [n]. \end{aligned}$$

Remark: Note that, as this is a minimization problem, the constraint $x_i \leq 1$ would be redundant.

This is known as the *fractional covering* LP.

The dual of the fractional covering LP is known as the *fractional packing* LP,

$$\begin{aligned} \max \quad & \sum_{j \in [m]} y_j \\ \text{s.t.} \quad & \sum_{\substack{j \in [m], \\ i \in S_j}} y_i \leq 1, \forall i \in [n], \\ & y_j \geq 0, \forall j \in [m]. \end{aligned}$$

When we restrict ourselves to $y_i \in \{0, 1\}$, the problem is known as *packing*. For *packing*, we are interested in selecting the largest cardinality collection of disjoint subsets.

Furthermore, note that

$$\min \text{integer covering} \geq \min \text{fractional covering} = \max \text{fractional packing} \geq \max \text{integer packing}.$$

There are two things we will investigate

a. bounds on optimal values can be useful for finding approximately optimal solutions to our problem.
b. are there some specific covering/packing problems where teh inequalities are equalities?

We introduce (a), by consider a greedy algorithm for minimum cover.

Algorithm 1: `COVERING-GREEDY`

$X = \emptyset$
$U = [m]$
while $U \neq \emptyset$ **do**

choose some $i \in [n]$ maximizing $|\{j \in U \mid i \in S\}|$
define $t = |\{j \in U \mid i \in S\}|$
$X = X \cup \{i\}$
for $j \in U$ such that $i \in S_j$ **do**
  $\text{price}(S_j) = \frac{1}{t}$
end
$U = U \setminus \{j \mid i \in S_j\}$
end
return X

Note that $|X| = \sum_j \text{price}(S_j)$. If $y_j := \text{price}(S_j)$ were a fractional packing then that would imply that X is optimal.

Lemma 9.2.1: For all $i \in [n]$,

$$\sum_{\substack{j \in [m] \\ i \in S_j}} \text{price}(S_j) \leq H_m$$

where $H_t = \sum_{t'=1}^{t} \frac{1}{t'}$ is the t-th harmonic number.

Proof: Fix some i, and let k be the number of sets that contain i. Without loss of generality, let these sets be $S_1, ..., S_k$ and let them be ordered in the order in which they were hit by the greedy algorithm.

Consider some S_j, with $j \in [k]$ and the iteration of the algorithm in which j ends up removed from U. At this point, there must be $\left|\{j \in U \mid i \in S_j\}\right| \geq k - j + 1$. That is to say i woud hit at least these many $k - j + 1$ new sets. So, the i-th element chosen by the algorithm must hit at least as many sets to. So, $t \geq k - j + 1$ which implies $\text{price}(S_j) \leq \frac{1}{k-j+1}$. Then,

$$\begin{aligned}
\sum_{\substack{j \in [m] \\ i \in S_j}} \text{price } (S_j) &= \sum_{j=1}^{k} \text{price}(S_j) \\
&= \frac{1}{k} + \frac{1}{k-1} + \frac{1}{k-2} + ... + \frac{1}{2} + 1 \\
&= H_k \leq H_m
\end{aligned}$$

□

Lemma 9.2.2: $H_m \leq \ln m + 1$

The prior lemma then implies that $y_j = \text{price}(S_j)/H_m$ defines a fractional packing. Thus, the maximum fractional packing has size at least

$$\frac{\sum_j \text{price}(S_j)}{H_m} = \frac{|X|}{H_m}.$$

So, the minimum integer cover has size $\geq \frac{|X|}{H_m}$. Thus, the integer covering produced by our algorithm is at most a factor H_m larger than the minimum integer covering.