X2TE2109: Convex Optimization and Its Applications in Signal Processing

Handout 3: Elements of Linear Programming

Instructor: Anthony Man–Cho So — March 8, 2022

1 Basic Definitions and Properties

We now turn to the study of a very important class of optimization problems, namely that of Linear Programming (LP) problems. To begin, let us recall the following definitions from Handout 2:

Definition 1 *Let $s \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $c \in \mathbb{R}$ be given. Then, the set of solutions to the linear equation $s^T x = c$, namely,*

$$H = \left\{ x \in \mathbb{R}^n : s^T x = c \right\},$$

is called a **hyperplane** *in $\mathbb{R}^n$. Associated with every hyperplane H are the two* **halfspaces**

$$H^- = \left\{ x \in \mathbb{R}^n : s^T x \le c \right\} \quad \textit{and} \quad H^+ = \left\{ x \in \mathbb{R}^n : s^T x \ge c \right\}.$$

Clearly, we have $H = H^+ \cap H^-$ and $\mathbb{R}^n = H^+ \cup H^-$. Also, note that H, H^+, and H^- are all closed convex sets.

Geometrically, a hyperplane is an $(n-1)$–dimensional affine subspace; i.e., it could be put into the form $H = \{\bar{x}\} + V$ for some vector $\bar{x} \in \mathbb{R}^n$ and linear subspace V of dimension $n - 1$. To see this, let $H = \left\{ x \in \mathbb{R}^n : s^T x = c \right\}$ be a hyperplane, where $s \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $c \in \mathbb{R}$ are given. Define

$$V = \left\{ x \in \mathbb{R}^n : s^T x = 0 \right\}, \quad \bar{x} = \frac{c}{s^T s} s.$$

Since V is the set of vectors that are orthogonal to s, it is a linear subspace of dimension $n - 1$. Moreover, a simple calculation shows that $s^T \bar{x} = c$ (i.e., $\bar{x} \in H$) and $\bar{x} + x \in H$ for any $x \in V$. Thus, we have $H \supset \{\bar{x}\} + V$. Conversely, for any $y \in H$, we have $x = y - \bar{x} \in V$, which implies that $H \subset \{\bar{x}\} + V$. It follows that $H = \{\bar{x}\} + V$, as desired. From such a representation of H, we can see that s is a *normal* of H.

Now, let us introduce the main geometric object of interest in the study of LP.

Definition 2 *A* **polyhedron** *is the intersection of a* finite *set of halfspaces. A* bounded *polyhedron is called a* **polytope**.

Recall from Handout 2 that a closed convex set is the intersection of all the halfspaces containing it. However, this does not imply that any closed convex set is a polyhedron, as Definition 2 requires the intersection to be *finite*. In particular, a closed convex set P is a polyhedron if and only if it can be represented as

$$P = \left\{ x \in \mathbb{R}^n : a_i^T x \le b_i \text{ for } i = 1, \ldots, m \right\} \tag{1}$$

for some given $a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$.

2 Extremal Elements of a Polyhedron

Consider now a point $\bar{x} \in P$, where $P \subset \mathbb{R}^n$ is a polyhedron of the form (1). If the index $i \in \{1, \ldots, m\}$ is such that $a_i^T \bar{x} = b_i$, then we say that the corresponding constraint is **active** or **binding** at $\bar{x}$. The following theorem is rather elementary. We leave the proof as an exercise to the reader.

Theorem 1 *Let $P \subset \mathbb{R}^n$ be a polyhedron of the form* (1)*, and consider a point $\bar{x} \in P$. Let $I = \left\{ i : a_i^T \bar{x} = b_i \right\}$ be the set of indices of constraints that are active at $\bar{x}$. Then, the following are equivalent:*

(a) There exist n vectors in the set $\{a_i \in \mathbb{R}^n : i \in I\}$ that are linearly independent.

(b) The point $\bar{x} \in \mathbb{R}^n$ is the unique solution to the following system of linear equations:

$$a_i^T x = b_i \qquad \text{for } i \in I.$$

Since any set of n linearly independent vectors in $\mathbb{R}^n$ forms a basis of $\mathbb{R}^n$, we are led to the following definition:

Definition 3 *Let $P \subset \mathbb{R}^n$ be a polyhedron and $x \in \mathbb{R}^n$ be arbitrary. The vector x is called a* **basic solution** *if there are n linearly independent active constraints at x. If in addition we have $x \in P$, then we say that x is a* **basic feasible solution***.*

Note that the definition of a basic solution is algebraic in nature. However, it seems to coincide with the notion of an extreme point, which is geometric in nature (recall that a point $x \in P$ is called an extreme point if there does not exist two different points $x_1, x_2 \in P$ such that $x = (x_1 + x_2)/2$). It turns out that these two notions are indeed equivalent, as the following theorem shows:

Theorem 2 *Let $P \subset \mathbb{R}^n$ be a polyhedron of the form* (1) *and $x \in P$ be arbitrary. Then, the following are equivalent:*

(a) x is an extreme point.

(b) x is a basic feasible solution.

REMARK: In the sequel we shall also use the term **vertex** to mean an extreme point/basic feasible solution.

Proof Suppose that $x \in P$ is not a basic feasible solution. Let $I = \{i : a_i^T x = b_i\}$. Then, the family $\{a_i \in \mathbb{R}^n : i \in I\}$ does not contain n linearly independent vectors. Hence, there exists a non–zero vector $d \in \mathbb{R}^n$ such that $a_i^T d = 0$ for all $i \in I$. Now, let $\epsilon > 0$ be a parameter to be determined, and set $x_1 = x - \epsilon d \in \mathbb{R}^n$ and $x_2 = x + \epsilon d \in \mathbb{R}^n$. Clearly, for any $i \in I$, we have $a_i^T x_1 = a_i^T x_2 = a_i^T x = b_i$. Moreover, for any $i \notin I$, we have $a_i^T x < b_i$ because $x \in P$. It follows that for sufficiently small $\epsilon > 0$, we have $a_i^T x_1 < b_i$ and $a_i^T x_2 < b_i$ for any $i \notin I$. Hence, $x_1, x_2 \in P$. Since $x = (x_1 + x_2)/2$ and $x_1 \neq x_2$, we conclude that x is not an extreme point.

Conversely, suppose that $x \in P$ is not an extreme point. Let $x_1, x_2 \in P$ be such that $x_1 \neq x_2$ and $x = (x_1 + x_2)/2$, and let $I = \{i : a_i^T x = b_i\}$. Since $x_1, x_2 \in P$, we have $a_i^T x_1 \le b_i$ and $a_i^T x_2 \le b_i$ for $i = 1, \ldots, m$, which yields $a_i^T x_1 = a_i^T x_2 = a_i^T x = b_i$ for all $i \in I$. This implies that the system of linear equations

$$a_i^T z = b_i \qquad \text{for } i \in I$$

has more than one solution in $z \in \mathbb{R}^n$. Hence, by Theorem 1, x is not a basic feasible solution. $\square$

One of the applications of Theorem 2 is to establish the non–polyhedrality of certain closed convex sets:

Example 1 (Non–Polyhedrality of the Euclidean Ball) *Consider the Euclidean ball $B(\mathbf{0},1) \subset \mathbb{R}^n$, which is a closed convex set. For the sake of contradiction, suppose that $B(\mathbf{0},1)$ is a polyhedron. Then, it admits a representation of the form* (1)*. Observe that the maximum number of basic feasible solutions in such a representation is $\binom{m}{n}$, which is finite. By Theorem 2, this is also the maximum number of extreme points of $B(\mathbf{0},1)$. However, this contradicts the result in Example 4 of Handout 2, which shows that the number of extreme points of $B(\mathbf{0},1)$ is infinite. Thus, we conclude that $B(\mathbf{0},1)$ is non–polyhedral. It is worth noting that $B(\mathbf{0},1)$ can be written as the infinite intersection of halfspaces $\bigcap_{d \in \mathbb{R}^n : \|d\|_2 = 1} \{x \in \mathbb{R}^n : d^T x \le 1\}$.*

It is clear that not every polyhedron has a vertex. For instance, a halfspace in $\mathbb{R}^n$, where $n \ge 2$, has no vertex. Thus, it is natural to ask whether one can characterize those polyhedra that have vertices. As it turns out, the answer is affirmative. Let us begin with the following definition:

Definition 4 *A polyhedron $P \subset \mathbb{R}^n$ contains a* **line** *if there exists a point $x \in P$ and a vector $d \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $x + \alpha d \in P$ for all $\alpha \in \mathbb{R}$.*

We then have the following simple characterization:

Theorem 3 *Let $P \subset \mathbb{R}^n$ be a polyhedron of the form* (1)*. Then, the following are equivalent:*

(a) P has at least one vertex.

(b) P does not contain a line.

(c) There exist n linearly independent constraints.

Proof The equivalence of (a) and (c) follows from Theorems 1 and 2. Now, suppose that P does not contain a line. We show that P has at least one vertex. Let $x \in P$ be arbitrary, and set $I = \{i : a_i^T x = b_i\}$. If the family $\mathcal{F} = \{a_i \in \mathbb{R}^n : i \in I\}$ contains n linearly independent vectors, then by Theorem 2 we conclude that x is a vertex of P. Hence, suppose that $\mathcal{F}$ does not contain n linearly independent vectors. Then, there exists a non–zero vector $d \in \mathbb{R}^n$ such that $a_i^T d = 0$ for all $i \in I$. Now, consider the line $\mathcal{L} = \{x + \alpha d : \alpha \in \mathbb{R}\}$. Clearly, for any $y \in \mathcal{L}$ and $i \in I$, we have $a_i^T y = a_i^T x = b_i$. However, since P does not contain a line, there must exist a scalar $\bar{\alpha} \in \mathbb{R}$ and an index $j \notin I$ such that $a_j^T(x + \bar{\alpha} d) = b_j$. We claim that $a_j \in \mathbb{R}^n$ is not a linear combination of the vectors in $\mathcal{F}$. Indeed, since $j \notin I$, we have $a_j^T x < b_j$. Moreover, since $a_j^T(x + \bar{\alpha} d) = b_j$, we must have $a_j^T d \neq 0$. However, if a_j is a linear combination of the vectors in $\mathcal{F}$, then we must have $a_j^T d = 0$, since $a_i^T d = 0$ for all $i \in I$. Thus, we conclude that a_j cannot be a linear combination of the vectors in $\mathcal{F}$. As a corollary, we see that the number of linearly independent active constraints increases by at least one as we move from $x \in P$ to $x + \bar{\alpha} d \in P$. By repeating the above argument, we will eventually end up with a point $\bar{x} \in P$ at which there are n linearly independent active constraints. By Theorem 2, $\bar{x}$ is then a vertex of P.

Now, suppose that there exist n linearly independent constraints, say $a_1, \ldots, a_n \in \mathbb{R}^n$. We show that P cannot contain a line. Suppose to the contrary that P contains the line $\{x + \alpha d : \alpha \in \mathbb{R}\}$, where $d \in \mathbb{R}^n$ is some non–zero vector. Then, for any $\alpha \in \mathbb{R}$ and $i = 1, \ldots, m$, we have $a_i^T(x + \alpha d) \le b_i$. It follows that $a_i^T d = 0$ for $i = 1, \ldots, m$. Since the vectors $a_1, \ldots, a_n$ are linearly independent, we conclude that $d = \mathbf{0}$, which is a contradiction. Hence, P does not contain a line, and the proof is completed. $\square$

3 Existence of Optimal Solutions to Linear Programs

Now, let $P \subset \mathbb{R}^n$ be a non–empty polyhedron of the form (1) and $h \in \mathbb{R}^n$ be a given vector. Consider the LP

$$\min_{x \in P} h^T x. \tag{$*$}$$

A natural first question concerning $(*)$ is whether it always has an optimal solution. Clearly, the answer is negative, as the optimal value of $(*)$ can be $-\infty$. On the other hand, if the optimal value of $(*)$ is finite, then intuitively one should be able to find an optimal solution. Note, however, that since P need not be bounded, we cannot apply Weierstrass' theorem to prove this. Rather, we will first use the techniques developed in the proof of Theorem 3 to prove the following:

Theorem 4 *Consider the LP $(*)$. Suppose that P has at least one vertex. Then, either the optimal value is $-\infty$, or there exists a vertex that is optimal.*

Proof The proof is similar to that of Theorem 3, except we also ensure that as we move towards a vertex, the optimal value does not increase. Before we proceed, let us introduce a definition. We say that $x \in P$ has rank $k \geq 0$ if there are exactly k linearly independent active constraints at x. Now, suppose that the optimal value is finite. Consider an $x \in P$ of rank $k < n$. Our goal is to show that there exists some $y \in P$ of greater rank and satisfies $h^T y \leq h^T x$. We can then repeat the process until we reach an optimal vertex.

As usual, let $I = \{i : a_i^T x = b_i\}$. Since there are only $k < n$ linearly independent vectors in the family $\{a_i \in \mathbb{R}^n : i \in I\}$, there exists a $d \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $a_i^T d = 0$ for all $i \in I$. Without loss of generality, we may assume that $h^T d \leq 0$. Consider the following cases:

<u>Case 1</u>: $h^T d < 0$.
Consider the half–line $\mathcal{L}_+ = \{x + \alpha d : \alpha > 0\}$. Clearly, for any $y \in \mathcal{L}_+$ and $i \in I$, we have $a_i^T y = b_i$. Now, if $\mathcal{L}_+ \subset P$, then the optimal value would be $-\infty$, which we have assumed not to be the case. Hence, there exist a scalar $\bar{\alpha} > 0$ and an index $j \notin I$ such that $a_j^T (x + \bar{\alpha} d) = b_j$. Let $y = x + \bar{\alpha} d$. Then, we have $h^T y < h^T x$. Moreover, upon following the argument in the proof of Theorem 3, we see that the family $\{a_i : i \in I\} \cup \{a_j\}$ is linearly independent, which implies that $y \in P$ has rank at least $k + 1$.

<u>Case 2</u>: $h^T d = 0$.
Consider the line $\mathcal{L} = \{x + \alpha d : \alpha \in \mathbb{R}\}$. Since P does not contain a line, there exist a scalar $\bar{\alpha} \neq 0$ and an index $j \notin I$ such that $a_j^T (x + \bar{\alpha} d) = b_j$. Let $y = x + \bar{\alpha} d$. Then, we have $h^T y = h^T x$, and the rank of $y \in P$ is at least $k + 1$.

In either case, we obtain a $y \in P$ whose rank is greater than that of $x \in P$ and satisfies $h^T y \leq h^T x$. By repeating the above process, we will end up with a $z \in P$ whose rank is n (i.e., z is a vertex of P) and satisfies $h^T z \leq h^T x$. To complete the proof of the theorem, let $z_1, \ldots, z_r$ be the vertices of P and set $i^* = \arg\min_{1 \leq i \leq r} h^T z_i$. Our argument above shows that for every $x \in P$, there exists an $i \in \{1, \ldots, r\}$ such that $h^T z_i \leq h^T x$. It follows that $h^T z_{i^*} \leq h^T x$ for all $x \in P$; i.e., $z_{i^*} \in P$ is an optimal vertex. $\square$

Now, in order to apply Theorem 4, the polyhedron P must have at least one vertex. This is not a serious restriction, however, since we can transform $(*)$ into an equivalent LP problem whose feasible set has at least one vertex. To see this, consider the polyhedron P' given by

$$P' = \left\{ \left(x^+, x^-, s\right) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m : a_i^T \left(x^+ - x^-\right) + s_i = b_i \ \text{ for } i = 1, \ldots, m; \ x^+, x^-, s \geq \mathbf{0} \right\}.$$

Note that if $x \in P$, then by setting

$$
\begin{aligned}
x_i^+ &= \begin{cases} x_i & \text{if } x_i \geq 0, \\ 0 & \text{otherwise} \end{cases} && \text{for } i = 1, \ldots, n, \\
x_i^- &= \begin{cases} 0 & \text{if } x_i \geq 0, \\ -x_i & \text{otherwise} \end{cases} && \text{for } i = 1, \ldots, n, \\
s_j &= b_j - a_j^T x && \text{for } j = 1, \ldots, m,
\end{aligned}
$$

we see that $(x^+, x^-, s) \in P'$. Conversely, if $(x^+, x^-, s) \in P'$, then by setting $x = x^+ - x^-$, we have $x \in P$. It follows that

$$\min_{x \in P} h^T x = \min_{(x^+, x^-, s) \in P'} h^T \left(x^+ - x^- \right); \tag{2}$$

i.e., minimizing $h^T x$ over P is equivalent to minimizing $h^T(x^+ - x^-)$ over P'. Furthermore, note that the polyhedron P' does not contain a line, and thus by Theorem 3, P' has at least one vertex. In particular, we obtain the following corollary to Theorem 4:

Corollary 1 *Consider the LP $(*)$. Suppose that P is non–empty. Then, either the optimal value is $-\infty$, or there exists an optimal solution.*

Corollary 1 is one of the many nice regularity properties enjoyed by the class of LPs. By contrast, the example $\inf_{x \geq 1} x^{-1}$ shows that nonlinear optimization problems need not have such a property.

To simplify notation in the sequel, let us write the problem on the right–hand side of (2) in a more compact fashion. Specifically, let $y = (x^+, x^-, s) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \cong \mathbb{R}^{2n+m}$. Define

$$A = \begin{bmatrix} a_1^T & -a_1^T & 1 & 0 & \cdots & 0 \\ a_2^T & -a_2^T & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_m^T & -a_m^T & 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{m \times (2n+m)}, \; b = (b_1, \ldots, b_m) \in \mathbb{R}^m, \; c = (h, -h, \mathbf{0}) \in \mathbb{R}^{2n+m}.$$

Then, the problem of minimizing $h^T(x^+ - x^-)$ over P' can be written as

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & Ay = b, \\ & y \geq \mathbf{0}. \end{array} \tag{3}$$

We shall call an LP problem of the form (3) a **standard form problem**.

Before we proceed, let us illustrate the above construction by considering a simple example.

Example 2 (Conversion to Standard Form LP) *Let $P = \{x \in \mathbb{R}^2 : e_1^T x \geq 1\} \subset \mathbb{R}^2$ and $h = e_1 \in \mathbb{R}^2$ in the LP $(*)$. It is clear that $(1, x_2)$ is an optimal solution for any $x_2 \in \mathbb{R}$. The polyhedron P' is given by*

$$P' = \{(x_1^+, x_2^+, x_1^-, x_2^-, s) \in \mathbb{R}_+^5 : x_1^+ - x_1^- - s = 1\} \subset \mathbb{R}^5$$

and the LP $(*)$ is equivalent to

$$\min_{(x_1^+, x_2^+, x_1^-, x_2^-, s) \in P'} x_1^+ - x_1^-. \tag{4}$$

Since P' has at least one vertex, by Theorem 4, the LP (4) has a vertex optimal solution. This is given by $y^* = (1, 0, 0, 0, 0)$. To verify that y^* is a vertex of P', it suffices to verify that the five active constraints

$$x_1^+ = 1,\ x_2^+ = 0,\ x_1^- = 0,\ x_2^- = 0,\ s = 0$$

are linearly independent; see Theorem 2.

4 Theorems of Alternatives

Recall that the development in the previous section assumes the non–emptiness of the polyhedron P. This raises the next natural question in the study of LPs: Given a system of linear inequalities, how do we determine whether it is feasible or not? Clearly, it is easy to certify that a given system is feasible, for one can simply exhibit a feasible solution. On the other hand, it is not so obvious how to certify that a given system is infeasible. An answer to this problem is provided by the following theorem of Farkas, which is better known as *Farkas' Lemma* in the literature. The idea is to construct an alternative system of linear inequalities such that the original system is infeasible if and only if the alternative system is feasible. Then, one can certify the infeasibility of the original system by exhibiting a feasible solution to the alternative system.

Theorem 5 (Farkas' Lemma) *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ be given. Then, exactly one of the following systems has a solution:*

$$Ax = b,\ x \geq \mathbf{0}. \tag{5}$$

$$A^T y \leq \mathbf{0},\ b^T y > 0. \tag{6}$$

Proof We first show that it is impossible for both (5) and (6) to have solutions. Indeed, suppose to the contrary that $\bar{x} \in \mathbb{R}^n$ is a solution to (5) and $\bar{y} \in \mathbb{R}^m$ is a solution to (6). We compute

$$0 \geq \left(\bar{y}^T A\right)\bar{x} = \bar{y}^T\left(A\bar{x}\right) = \bar{y}^T b > 0,$$

which is a contradiction. Hence, we conclude that at most one of (5) and (6) has a solution. Now, it remains to show that if (5) has no solution, then (6) has a solution. Consider the set $S = \{Ax : x \geq \mathbf{0}\}$. Since (5) has no solution, we have $b \notin S$. Now, in order to invoke the Separation Theorem (Theorem 7 of Handout 2), we need to first prove the following lemma:

Lemma 1 *S is a non–empty closed convex set.*

Proof It is clear that S is convex. Moreover, since $\mathbf{0} \in S$, we see that S is non–empty. Thus, it remains to show that S is closed. Towards that end, let $\left\{y^k\right\}$ be a sequence in S such that $y^k \to y$. Our goal is to show that $y \in S$. Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \|y - Ax\|_\infty \\ \text{subject to} & x \geq \mathbf{0}. \end{array}$$

Using the techniques introduced in Handout 1, we can rewrite the above problem as follows:

$$\begin{array}{lll} \text{minimize} & t & \\ \text{subject to} & t \geq y_i - (Ax)_i & \text{for } i = 1, \ldots, m, \\ & t \geq -y_i + (Ax)_i & \text{for } i = 1, \ldots, m, \\ & x \geq \mathbf{0}. & \end{array} \tag{7}$$

Note that the feasible set of (7) is a polyhedron with no lines. Hence, by Theorems 3 and 4, there exists an optimal solution (x^*, t^*) whose optimal value is $t^* \geq 0$. We claim that $t^* = 0$. Indeed, suppose that $t^* > 0$. Then, since $y^k \to y$, there exists an index k' such that $\|y - y^{k'}\|_\infty = t' < t^*$. Since $y^{k'} \in S$, there exists an $x' \geq \mathbf{0}$ such that $y^{k'} = Ax'$. It follows that (x', t') is a feasible solution to (7), which contradicts the optimality of (x^*, t^*). Hence, we have $t^* = 0$, from which it follows that $y = Ax^*$; i.e., $y \in S$ as desired. $\square$

Now, by Theorem 7 of Handout 2, there exists a $y \in \mathbb{R}^m$ such that $y^T Ax < y^T b$ for all $x \geq \mathbf{0}$. Since $\mathbf{0} \in S$, we have $b^T y > 0$. We further claim that $A^T y \leq \mathbf{0}$. Suppose that this is not the case. Then, there exists an index $i \in \{1, \ldots, n\}$ such that $\left(A^T y\right)_i > 0$. Consider the vector $x' = \lambda e_i$, where $\lambda > 0$ and e_i is the i–th basis vector. Clearly, we have $x' \geq \mathbf{0}$. Moreover, we have $y^T b > y^T Ax' = \lambda \left(A^T y\right)_i$ for all $\lambda > 0$, which is impossible. Hence, we have $A^T y \leq \mathbf{0}$ as claimed, and this completes the proof. $\square$

Note that a crucial step in our proof of Farkas' lemma is to show that the image of the closed cone $\{x \in \mathbb{R}^n : x \geq \mathbf{0}\}$ under the mapping A is closed. Such a step should not be taken for granted, as the image of a closed cone under an affine map need not be closed. For more information, see [**?**].

Once we have proven Farkas' lemma, we can use it to establish a variety of theorems of alternatives. As an illustration, let us use Farkas' lemma to prove the following theorem of Gordan:

Corollary 2 (Gordan's Theorem) *Let $A \in \mathbb{R}^{m \times n}$ be given. Then, exactly one of the following systems has a solution:*

$$Ax > \mathbf{0}. \tag{8}$$

$$A^T y = \mathbf{0},\ y \geq \mathbf{0},\ y \neq \mathbf{0}. \tag{9}$$

Proof Again, it is clear that (8) and (9) cannot both have solutions, for otherwise there would exist $\bar{x} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$ such that

$$0 = \left(y^T A\right) x = y^T (Ax) > 0,$$

which is a contradiction. Now, note that (8) is equivalent to $Ax \geq e$, since we can scale both sides of (8) by any positive scalar. On the other hand, the system $Ax \geq e$ is equivalent to the system $\tilde{A}z = e$, $z \geq \mathbf{0}$, where $\tilde{A} = [\ A \quad -A \quad -I\ ] \in \mathbb{R}^{m \times (2n+m)}$ and $z = (x^+, x^-, s) \in \mathbb{R}^{2n+m}$. Now, by Farkas' lemma, if the system $\tilde{A}z = e$, $z \geq \mathbf{0}$ has no solution, then there exists a $y \in \mathbb{R}^m$ such that $\tilde{A}^T y \leq \mathbf{0}$ and $e^T y > 0$. From the definition of $\tilde{A}$, we see that $A^T y = \mathbf{0}$ and $y \geq \mathbf{0}$. Moreover, since $e^T y > 0$, we conclude that $y \neq \mathbf{0}$. This completes the proof. $\square$

5 LP Duality Theory

In view of the developments in Sections 3 and 4, it remains to understand when a feasible LP has a finite optimal value. Towards that end, let us recall the standard form LP:

$$\begin{array}{rcll} v_p^* & = & \text{minimize} & c^T x \\ & & \text{subject to} & Ax = b, \\ & & & x \geq \mathbf{0}. \end{array} \tag{P}$$

One obvious idea for assessing whether (P) has finite optimal value is to develop a lower bound on v_p^*. However, this seems more difficult than finding an upper bound on v_p^*, which simply involves taking a feasible solution to (P) and evaluating its corresponding objective value. Nevertheless, we can proceed as follows. Suppose that we can find a vector $y \in \mathbb{R}^m$ such that $A^T y \le c$. Then, for any $x \in \mathbb{R}^n$ that is feasible for (P), we have

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

where the equality is due to $Ax = b$ and the inequality is due to $x \ge \mathbf{0}$ and $A^T y \le c$. Since the above inequality holds for any feasible solution $x \in \mathbb{R}^n$ to (P), it follows that $b^T y$ provides a lower bound on v_p^* for any $y \in \mathbb{R}^m$ satisfying $A^T y \le c$. Naturally, we are interested in finding the largest lower bound on v_p^*. This motivates us to consider the following optimization problem:

$$\begin{array}{rcll} v_d^* & = & \text{maximize} & b^T y \\ & & \text{subject to} & A^T y \le c. \end{array} \tag{D}$$

Note that (D) is also an LP. In the sequel we shall call (P) the **primal problem** and (D) its **dual problem**. Our discussion above concerning the construction of (D) immediately leads to the following result:

Theorem 6 (LP Weak Duality) *Let $\bar{x} \in \mathbb{R}^n$ be feasible for (P) and $\bar{y} \in \mathbb{R}^m$ be feasible for (D). Then, we have $b^T \bar{y} \le c^T \bar{x}$. In particular, $v_p^* \ge v_d^*$.*

Even though the LP weak duality theorem is almost trivial, it has some useful corollaries.

Corollary 3 *The following hold:*

(a) *If the optimal value of (P) is $-\infty$, then (D) must be infeasible.*

(b) *If the optimal value of (D) is $+\infty$, then (P) must be infeasible.*

(c) *Let $\bar{x} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$ be feasible for (P) and (D), respectively. Suppose that the* duality gap *$\Delta(\bar{x}, \bar{y}) = c^T \bar{x} - b^T \bar{y} = 0$. Then, $\bar{x}$ and $\bar{y}$ are optimal solutions to (P) and (D), respectively.*

Motivated by Corollary 3, it is natural to ask whether it is possible to have an optimal solution $\bar{x} \in \mathbb{R}^n$ to (P) and an optimal solution $\bar{y} \in \mathbb{R}^m$ to (D) such that the duality gap is non–zero. As it turns out, the answer is negative. Consequently, our construction of the dual problem (D) gives a tight lower bound on the primal objective value v_p^*.

Theorem 7 (LP Strong Duality) *Suppose that (P) has an optimal solution $x^* \in \mathbb{R}^n$. Then, (D) also has an optimal solution $y^* \in \mathbb{R}^m$, and $c^T x^* = b^T y^*$.*

Proof Suppose that (P) has an optimal solution $x^* \in \mathbb{R}^n$. Then, the system

$$Ax = b,\ x \ge \mathbf{0},\ c^T x < c^T x^* \tag{10}$$

does not have a solution in $x \in \mathbb{R}^n$. To apply Farkas' lemma, we first homogenize the above system to get

$$Ax - bt = \mathbf{0},\ c^T x - \left(c^T x^*\right) t = -1 < 0,\ (x, t) \ge \mathbf{0}. \tag{11}$$

We claim that (11) has no solution in $(x,t) \in \mathbb{R}^n \times \mathbb{R}$. Indeed, if (x',t') is a feasible solution to (11) with $t' > 0$, then x'/t' is a solution to (10), which is a contradiction. On the other hand, if $t' = 0$, then we have $A(x^* + x') = b$, $x^* + x' \geq \mathbf{0}$, and $c^T(x^* + x') = c^T x^* - 1 < c^T x^*$. This shows that $x^* + x'$ is a solution to (10), which again is a contradiction. Thus, the claim is established. Now, by Farkas' lemma, there exists a $y^* \in \mathbb{R}^m$ such that $c - A^T y^* \geq \mathbf{0}$ and $-c^T x^* + b^T y^* \geq 0$. It follows that y^* is feasible for (D). Moreover, by the LP weak duality theorem, we have $c^T x^* = b^T y^*$. This completes the proof. $\square$

Together with Theorem 4, we have the following corollary:

Corollary 4 *Suppose that both (P) and (D) are feasible. Then, both (P) and (D) have optimal solutions, and their respective optimal values are equal.*

One important consequence of the LP strong duality theorem is that the task of finding optimal solutions to (P) and (D) is equivalent to that of finding a feasible solution to the following linear system in $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{aligned} Ax = b,\ x \geq \mathbf{0}, &\qquad \text{(primal feasibility)} \\ A^T y \leq c, &\qquad \text{(dual feasibility)} \\ c^T x = b^T y. &\qquad \text{(zero duality gap)} \end{aligned}$$

In particular, the problem of *linear optimization* is no harder than that of *linear feasibility.*

Using the LP strong duality theorem, one can derive another important relation between the primal and dual optimal solutions.

Theorem 8 (Complementary Slackness) *Let $\bar{x} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$ be feasible for (P) and (D), respectively. Then, the vectors $\bar{x}$ and $\bar{y}$ are optimal for their respective problems iff $\bar{x}_i \left(c - A^T \bar{y}\right)_i = 0$ for $i = 1, \ldots, n$.*

Proof Using the fact that $A\bar{x} = b$, we have

$$c^T \bar{x} - b^T \bar{y} = c^T \bar{x} - \bar{x}^T A^T \bar{y} = \bar{x}^T \left(c - A^T \bar{y}\right) = \sum_{i=1}^{n} \bar{x}_i \left(c - A^T \bar{y}\right)_i . \tag{12}$$

Now, if $\bar{x}_i \left(c - A^T \bar{y}\right)_i = 0$ for $i = 1, \ldots, n$, then we have $c^T \bar{x} = b^T \bar{y}$. By the LP strong duality theorem, we conclude that $\bar{x}$ and $\bar{y}$ are optimal for their respective problems. Conversely, if $\bar{x}$ and $\bar{y}$ are optimal for their respective problems, then by the LP strong duality theorem, we have $c^T \bar{x} - b^T \bar{y} = 0$. Since $\bar{x} \geq \mathbf{0}$ and $c - A^T \bar{y} \geq \mathbf{0}$ by the feasibility of $\bar{x}$ and $\bar{y}$, we conclude by (12) that $\bar{x}_i \left(c - A^T \bar{y}\right)_i = 0$ for $i = 1, \ldots, n$. $\square$

From Theorem 8, we see that another way of solving (P) and (D) is to solve the following (nonlinear) system in $(x,y,s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$:

$$\begin{aligned} Ax = b,\ x \geq \mathbf{0}, &\qquad \text{(primal feasibility)} \\ A^T y + s = c, &\qquad \text{(dual feasibility)} \\ x_i s_i = 0 \ \text{ for } i = 1, \ldots, n. &\qquad \text{(complementarity)} \end{aligned}$$

Although the above system is nonlinear, it plays an important role in the development of efficient interior–point algorithms for LP. We refer the interested reader to [**? ?**] for details.

As an illustration of LP duality, let us consider the following example:

Example 3 (A Simple LP) *Consider the following LP:*

$$
\begin{array}{lrcrcrcl}
\text{minimize} & x_1 & + & 2x_2 & + & x_3 & & \\
\text{subject to} & x_1 & - & 2x_2 & + & x_3 & \geq & 2, \\
& -x_1 & + & & + & x_3 & \geq & 4, \\
& 2x_1 & + & & + & x_3 & \geq & 6, \\
& x_1 & + & x_2 & + & x_3 & \geq & 2, \\
& & & & & x & \geq & \mathbf{0}.
\end{array} \tag{13}
$$

To derive the dual of (13)*, we first put it into standard form by introducing suitable slack variables:*

$$
\begin{array}{ll}
\text{minimize} & (1,2,1,0,0,0,0)^T(x_1,x_2,x_3,s_1,s_2,s_3,s_4) \\
\text{subject to} & \left[\begin{array}{ccc|c} 1 & -2 & 1 & \\ -1 & 0 & 1 & \\ 2 & 0 & 1 & -I \\ 1 & 1 & 1 & \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ s \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 2 \end{bmatrix}, \\
& (x,s) \geq \mathbf{0}.
\end{array} \tag{14}
$$

Using (D)*, we obtain the dual of* (13) *as follows:*

$$
\begin{array}{ll}
\text{maximize} & (2,4,6,2)^T(y_1,y_2,y_3,y_4) \\
\text{subject to} & \left[\begin{array}{cccc} 1 & -1 & 2 & 1 \\ -2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ \hline & \multicolumn{2}{c}{-I} & \end{array}\right] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 2 \\ 1 \\ \mathbf{0} \end{bmatrix}.
\end{array} \tag{15}
$$

Now, consider the point

$$
(\bar{x},\bar{s}) = (\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{s}_1,\bar{s}_2,\bar{s}_3,\bar{s}_4) = \left(\frac{2}{3}, 0, \frac{14}{3}, \frac{10}{3}, 0, 0, \frac{10}{3}\right),
$$

which, as can be easily verified, is feasible for Problem (14)*. By Theorem 8, the point* $(\bar{x},\bar{s})$ *is optimal for Problem* (14) *iff there exists a feasible solution* $\bar{y} \in \mathbb{R}^4$ *to Problem* (15) *such that*

$$
\begin{array}{ll}
\bar{y}_1 = \bar{y}_4 = 0, & (\textit{since } \bar{s}_1, \bar{s}_4 > 0) \\
-\bar{y}_2 + 2\bar{y}_3 = 1 & (\textit{since } \bar{x}_1 > 0) \\
\bar{y}_2 + \bar{y}_3 = 1, & (\textit{since } \bar{x}_3 > 0) \\
\bar{y}_2,\ \bar{y}_3 \geq 0, & (\textit{dual feasibility})
\end{array}
$$

or equivalently, the point

$$
\bar{y} = \left(0, \frac{1}{3}, \frac{2}{3}, 0\right)
$$

is feasible for Problem (15)*. The latter can be easily verified to be the case. Hence, we have certified the optimality of the primal–dual pair of solutions* $(\bar{x},\bar{s},\bar{y})$*. Note that we also have*

$$
(1,2,1,0,0,0,0)^T(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{s}_1,\bar{s}_2,\bar{s}_3,\bar{s}_4) = \frac{16}{3} = (2,4,6,2)^T(\bar{y}_1,\bar{y}_2,\bar{y}_3,\bar{y}_4);
$$

i.e., the duality gap is zero.

By Theorems 4 and 6, we know that Problem (13) *has a vertex optimal solution. Recall from Definition 3 that each vertex of the feasible region of Problem* (13) *should have three linearly independent active constraints. Since the active constraints at* $\bar{x}$ *are*

$$\begin{aligned} (-1,0,1)^T(\bar{x}_1,\bar{x}_2,\bar{x}_3) &= 4, \\ (2,0,1)^T(\bar{x}_1,\bar{x}_2,\bar{x}_3) &= 6, \\ (0,1,0)^T(\bar{x}_1,\bar{x}_2,\bar{x}_3) &= 0, \end{aligned}$$

which are linearly independent (they correspond to the coefficient vectors $(-1,0,1)$*,* $(2,0,1)$*, and* $(0,1,0)$*), we conclude that* $\bar{x}$ *is a vertex optimal solution to Problem* (13)*.*

For further reading on linear programming, we refer the reader to the books [**? ?**].