1 Convex Sets

Definition 1. A set $S$ in $\mathbb{R}^n$ is said to be convex if for every two points $x_1$ and $x_2$ in $S$ and every scalar $\alpha \in (0, 1)$, the point $\alpha x_1 + (1 - \alpha)x_2 \in S$.

This definition can be interpreted geometrically as stating that a set is convex if, given two points in the set, every point on the line segment joining these two points is also a member of the set. The point $x = \alpha x_1 + (1 - \alpha)x_2$ is said to be a convex combination of $x_1$ and $x_2$. More generally, we have the following definition.

Definition 2. Given a finite collection of points $x_1, x_2, \ldots, x_k$ in $\mathbb{R}^n$, a point $x$ in $\mathbb{R}^n$ is called a (strict) convex combination of these points if

$$ x = \sum_{i=1}^{k} \alpha_i x_i $$

for some $\alpha_i \geq 0$ ($\alpha_i > 0$), $i = 1, 2, \ldots, k$ with $\sum_{i=1}^{k} \alpha_i = 1$.

Theorem 1. A set $S$ is convex if and only if it contains all convex combinations of points in $S$.

Proof. First, if a set contains all convex combinations of its points, then especially all convex combinations of pairs so that it is convex, in particular. We prove the opposite direction by induction on the number of points, so let $S$ be convex and thus contain all convex combinations of pairs ($n = 2$). Then assume that $S$ contains all convex combinations of at most $n$ points, so $\sum_{i=1}^{k} \alpha_i x_i \in S$ for any $\alpha_i \geq 0$ with $\sum_{i=1}^{k} \alpha_i = 1$ and $x_i \in S$, and let $\sum_{i=1}^{k+1} \tilde{\alpha}_i = 1$ with $\tilde{\alpha}_i \geq 0$ and $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{k+1} \in S$. To show that $x = \sum_{i=1}^{k+1} \tilde{\alpha}_i \tilde{x}_i \in S$, pick any $\tilde{\alpha}_j < 1$ so that $\sum_{i \neq j} \tilde{\alpha}_i = 1 - \tilde{\alpha}_j > 0$ and write

$$ z = \sum_{i=1}^{k+1} \tilde{\alpha}_i \tilde{x}_i = \tilde{\alpha}_j \tilde{x}_j + \sum_{i \neq j} \tilde{\alpha}_i \tilde{x}_i = \tilde{\alpha}_j \tilde{x}_j + (1 - \tilde{\alpha}_j) \sum_{i \neq j} \frac{\tilde{\alpha}_i}{\tilde{\alpha}_j} \tilde{x}_i = \tilde{\alpha}_j \tilde{x}_j + (1 - \tilde{\alpha}_j) \sum_{i \neq j} \frac{\tilde{\alpha}_i}{\sum_{i \neq j} \tilde{\alpha}_i} \tilde{x}_i $$

which belongs to $S$ as the second term corresponds to a point in $S$ as convex combination of $k$ points, so that $x$ belongs to $S$ as convex combination of a pair. By induction, the proof is complete. \qed
1.1 Basic Properties (Luenberger-Ye Appendix B.1)

Some additional properties of convex sets are summarized in the following proposition.

Proposition 1. Let $S$ and $T$ be convex sets in $\mathbb{R}^n$, and $\alpha$ and $\beta$ be two real numbers.

(a) The set $\alpha S + \beta T = \{x \in \mathbb{R}^n : x = \alpha x_1 + \beta x_2, \ x_1 \in S, \ x_2 \in T\}$ is convex.

(b) The set $S \cap T$ is convex.

Exercise 1 (HW). Prove Proposition 1.

More generally, intersections of arbitrary (finite or infinite) collections of convex sets are convex. Prove it!

Definition 3. Given any set $S$ in $\mathbb{R}^n$, the intersection of all convex sets containing $S$ is called the convex hull of $S$, denote by $\text{conv}(S)$ or $\text{co}(S)$:

$$\text{conv}(S) = \bigcap \{C \supseteq S : C \text{ is convex}\}$$

(3)

Theorem 2. The convex hull of a set $S$ is the (smallest) set $C$ that contains all convex combinations of points in $S$:

$$C = \text{co}(S) = \{z \in \mathbb{R}^n : z = \sum_{i=1}^{k} \alpha_i z_i, \ \alpha_i \geq 0, \ \sum_{i=1}^{k} \alpha_i = 1, \ z_i \in S\}.$$  \hspace{1cm} (4)


1.2 Hyperplanes, Halfspaces, and Cones (Luenberger-Ye Appendix B.2)

Definition 4. Given a nonzero vector $a$ in $\mathbb{R}^n$ and a real number $b$, the set

$$H = \{x \in \mathbb{R}^n : a^T x = b\}$$

(5)

is called a hyperplane (or $n-1$-dimensional linear variety or manifold), $a$ is called its normal vector, and the sets

$$H^+ = \{x \in \mathbb{R}^n : a^T x \geq b\} \text{ and } H^- = \{x \in \mathbb{R}^n : a^T x \leq b\}$$

(6)

are called its associated positive and negative (closed) halfspaces, respectively.

It is easy to see (or verify) that hyperplanes and halfspaces are convex sets.

Definition 5. A set $S$ in $\mathbb{R}^n$ is called polyhedral, or a (convex) polyhedron if it can be represented as the intersection of a finite number of (closed) halfspaces:

$$S = \{x \in \mathbb{R}^n : Ax \geq b\}$$

(7)

for some matrix $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. A (nonempty) bounded polyhedron is also called a polytope.

In the above definition, we let the rows of $A$ be denoted by $a_i$, so that $Ax \geq b$ if and only if $a_{i}^T x \geq b_i$ for all $i = 1, \ldots, m$.

Remark 1. The use of polyhedron and polytope is not consistent in the literature. For example, Luenberger and Ye define a polyhedron to be a bounded polytope, and Wikipedia defines a polyhedron to be a bounded polytope in $\mathbb{R}^3$, compared to a polygon in $\mathbb{R}^2$. Use whatever convention makes the most sense to you! (My own research has focused some on polyhedral cones, and as such I tend to use the word polyhedron for a general bounded or unbounded polytope in $\mathbb{R}^n$.)
Definition 6. A set $K$ in $\mathbb{R}^n$ is called a cone if $\alpha K \subseteq K$ for every $\alpha \geq 0$. A convex (polyhedral) cone is a cone that is also convex (polyhedral).

Proposition 2. A cone $K$ is convex if and only if $K + K \subseteq K$.

Exercise 3 (HW). Prove Proposition 2. Also, try to find a cone that is convex but not polyhedral.

Definition 7. Let $S$ be a set in $\mathbb{R}^n$. The two sets

$$S^+ = \{ a \in \mathbb{R}^m : a^T x \geq 0 \text{ for all } x \in S \}$$

and

$$S^- = \{ a \in \mathbb{R}^m : a^T x \leq 0 \text{ for all } x \in S \}$$

are called positive and negative dual cone of $S$, respectively.

Sometimes, the set $S^-$ the is also called the polar cone of $S$. If you don’t believe it, make sure to check yourself that $S^+$ and $S^-$ are in fact cones.

2 Separation Theorems (Luenberger-Ye Appendix B.3)

Definition 8. Given two (nonempty) sets $S$ and $T$ in $\mathbb{R}^n$, a hyperplane $H$ in $\mathbb{R}^n$ is said to separate $S$ and $T$ if the sets are contained in different halfspaces:

$$S \subseteq H^+ \text{ and } T \subseteq H^- \text{ (or vice versa).}$$

The separation is said to be strict if both $S$ and $T$ are contained in the respective open halfspaces.

Theorem 3 (Separating Hyperplane Theorem). Let $S$ and $T$ be two (nonempty) convex sets in $\mathbb{R}^n$ with no common interior point. Then there exists a separating hyperplane for $S$ and $T$, i.e., there exists a nonzero vector $a$ in $\mathbb{R}^n$ such that

$$a^T x \leq a^T y \text{ for all } x \in S \text{ and } y \in T.$$  

Proof. The proof goes in three steps. In each case, $S$ is an arbitrary (nonempty) convex set in $\mathbb{R}^n$.

1. Let $T = \{ y \}$ be a singleton exterior to the closure of $S$, so

$$\delta = \inf_{x \in S} \| y - x \| > 0.$$  

Because $\| \cdot \|$ is a continuous function and $S$ is bounded into direction of $y$, there exists $z$ in the closure of $S$ such that $\| z - y \| = \delta$, and by convexity of $S$ it follows that

$$\alpha x + (1 - \alpha) z = z + \alpha (x - z)$$  

also belongs to $S$ for all $x \in S$ and $0 < \alpha \leq 1$. Because $\| y - z \| \leq \| y - x \|$ for all $x \in S$, we also have that

$$\| y - z \|^2 \leq \| y - z - \alpha (x - z) \|^2 = \| y - z \|^2 + \alpha^2 \| x - z \|^2 - 2\alpha (y - z)^T (x - z)$$  

and letting $\alpha \searrow 0$ yields that $(y - z)^T (x - z) \leq 0$, or equivalently

$$(y - z)^T x \leq (y - z)^T z = (y - z)^T y - (y - z)^T (y - z) = (y - z)^T y - \delta^2 < (y - z)^T y.$$  

Hence, with $a = y - z$ and $b = (y - z)^T y$, the hyperplane $H = \{ x \in \mathbb{R}^n : (y - z)^T x = (y - z)^T y \}$ separates $S \subset H^-$ and $y \in H$. Alternatively, the hyperplane $H = \{ x \in \mathbb{R}^n : (y - z)^T x = 0.5(y - z)^T (y + z) \}$ separates $S \subset H^-$ and $y$ strictly.
(2) Not let \( T = y \) (still a singleton) be a boundary point of \( S \). In this case, \( \delta = \inf_{x \in S} \{ y - x \} = 0 \) so that the minimizer \( z = y \) and \( a = y - z = 0 \) does not provide a nonzero vector as normal vector. Hence, resolve this case using a limiting process of case 1 and let \( \{ y_k \} \) be a sequence of points exterior to the closure of \( S \) that converges to \( y \). Let \( \{ a_k \} \) be the corresponding sequence of normal vectors constructed as in case 1, so \( \inf_{x \in S} a_k^T x < a_k^T y_k \), and without loss of generality let each \( a_k \) be normalized so that \( \| a_k \| = 1 \) for all \( k \) (otherwise \( \| a_k \| \to 0 \) with the same problem as above). Since the sequence \( \{ a_k \} \) is bounded, by the Bolzano-Weierstrass Theorem there exists a convergent subsequence \( \{ a_k \}_{k \in K} \) with \( \lim_{k \in K} a_k = a \), \( \| a \| = 1 \) and

\[
 a^T x = \lim_{k \in K} a_k^T x \leq \lim_{k \in K} a_k^T y_k = a^T y. \tag{15}
\]

Hence, the hyperplane \( H = \{ x \in \mathbb{R}^n : a^T x = a^T y \} \) separates \( S \subseteq H^- \) and \( y \in H \). Note that there is no possible strict separation.

(3) Now let both \( S \) and \( T \) be (nonempty) convex sets without common interior and consider the set difference

\[
 S - T = \{ z \in \mathbb{R}^n : z = x - y \text{ for some } x \in S \text{ and } y \in T \}, \tag{16}
\]

which is convex by Proposition 1 and does not contain the origin in its interior. Hence, from Case 1 or 2, there exists a nonzero vector \( a \in \mathbb{R}^n \) such that

\[
 a^T (x - y) \leq a^T 0 = 0 \iff a^T x \leq a^T y \tag{17}
\]

for all \( x \in S \) and \( y \in T \). The proof is complete. \( \Box \)

Note that for case 2, there is no strict separation between a convex set and one of its boundary point. The following definition and theorem specifies this case formally.

**Definition 9.** Given a set \( S \) in \( \mathbb{R}^n \) and a point \( z \) in the closure of \( S \), a hyperplane \( H \) in \( \mathbb{R}^n \) is said to support \( S \) at \( z \) if \( z \) belongs to \( H \) and \( S \) is contained in one of its associated (closed) halfspaces, \( x \in H \) and \( S \subseteq H^+ \) (or \( S \subseteq H^- \)).

**Theorem 4** (Supporting Hyperplane Theorem). Let \( S \) be a convex set in \( \mathbb{R}^n \) and \( z \) be a boundary point of \( S \). Then there exists a supporting hyperplane for \( S \) at \( z \).

The proof of Theorem 4 follows directly from step 2 in the proof of Theorem 3.

**Corollary 1.** A (closed / open) set is convex if and only if it is the intersection of all its supporting (closed / open) halfspaces.

**Exercise 4** (HW). Prove Corollary 1.

### 3 Representation Theorems

#### 3.1 Bounded Sets and Extreme Points (Luenberger-Ye Appendix B.4)

**Definition 10.** A point \( x \) in a convex set \( S \) is called an extreme point of \( S \) if it cannot be written as (strict) convex combination of any two other points in \( S \):

\[
 x = \alpha y + (1 - \alpha) z \text{ for some } 0 < \alpha < 1, \ y, z \in S \Rightarrow x = y = z. \tag{18}
\]

**Theorem 5.** A closed bounded convex set in \( \mathbb{R}^n \) is equal to the (closed) convex hull of its extreme points.
Proof. The proof goes by induction over the dimension $n$. It can be easily verified (or is obvious) that the theorem is true on a line segment for $n = 1$, so next assume that the theorem is true for $n - 1$, let $S$ be a closed bounded convex set in $\mathbb{R}^n$, let $K$ be the (closed) convex hull of extreme points of $S$, and show that $S \subseteq K$ (because it is clear that $K \subseteq S$ as $K$ is the convex hull of a subset of points of $S$, and $S$ is convex). By contradiction, now assume that there exists $y \in S \setminus K$ and apply Theorem 3 to get a strictly separating hyperplane for $K$ and $y$ with normal vector $a$, so $a^T x < a^T y$ for all $x \in K$. Let $\delta = \inf_{x \in S} \{a^T x\}$, then $a^T z = \delta$ for some $z \in S$ by Weierstrass, and the hyperplane $H = \{x \in \mathbb{R}^n : a^T x = \delta\}$ supports $S$ at $z$ and is disjoint from $K$ because $\delta \leq a^T x < a^T y$ for all $x \in K$ by the above. Now consider $T = S \cap H$ as (closed) bounded convex subset of $H$ of dimension $n - 1$. Clearly, $T$ is nonempty because $z \in T$, and by assumption $T$ is equal to the closed convex hull of its extreme points; in particular, there exists an extreme point $w$ of $T$. Because $T$ is a subset of $H$ which is disjoint from $K$, $w$ cannot be extreme point of $S$ so there exist points $u$ and $v$ in $S$, so $a^T u \geq \delta$ and $a^T v \geq \delta$, such that $w = \alpha u + (1 - \alpha) v$ for some $0 < \alpha < 1$. Because $w \in T = H \cap S$, this implies that $a^T w = \alpha a^T u + (1 - \alpha) a^T v = \delta$ yielding that $a^T u = a^T v = \delta$, and thus $u$ and $v$ also belong to $T = H \cap S$ in contradiction to $w$ being an extreme point of $T$. Hence, the initial assumption $S \setminus K \neq \emptyset$ was wrong, and the proof is done. $\square$

In consequence, the following results are immediate.

**Theorem 6** (Representation Theorem for Bounded Convex Sets). Let $S$ be a (closed) bounded convex set in $\mathbb{R}^n$.

(a) If $S$ is nonempty, then $S$ has at least one extreme point.

(b) A point $x$ belongs to $S$ if and only if $x$ can be expressed as a convex combination of (finitely many) extreme points of $S$.

### 3.2 Unbounded Sets and Extreme Directions

Next, we turn our attention to the case where $S$ is unbounded.

**Proposition 3.** A convex set $S$ in $\mathbb{R}^n$ is unbounded if and only if there exists a halfline in $S$:

$$x = y + \alpha d \in S \text{ for some } y \in S, \ d \in \mathbb{R}^n \setminus \{0\} \text{ and all } \alpha \geq 0$$

**Exercise 5** (HW). Prove Proposition 3 using a general definition of set unboundedness.

A result similar to Theorem 6 also holds for the unbounded case, replacing the notion of extreme points by (extreme) directions.

**Definition 11.** A nonzero vector $d \in \mathbb{R}^n$ is called a (recession) direction of a convex set $S$ if $y + \alpha d \in S$ for some (and thus all) $y \in S$ and $\alpha \geq 0$. It is said to be an extreme direction if it cannot be written as (positive) sum of two other (distinct) directions of $S$:

$$d = u + v \Rightarrow d = \alpha u = \beta v \text{ for some real numbers } \alpha \text{ and } \beta.$$  (20)

**Theorem 7** (Representation Theorem for Unbounded Convex Sets). Let $S$ be a unbounded convex set in $\mathbb{R}^n$ containing no lines.

(a) $S$ has at least one extreme point and one extreme direction.

(b) A point $x$ belongs to $S$ if and only if $x$ can be expressed as sum $x = y + d$ where $y$ is a convex combination of (finitely many) extreme points of $S$ and $d$ is the positive linear combination (or simpler: sum) of extreme directions of $S$.

Note that every positive linear combination of directions can also be written as a sum upon suitable scaling of each direction. For a proof of Theorem 7 (and a lot more about convexity), see the classic monograph “Convex Analysis” by R. Tyrrell Rockafellar (1970).
4 Polyhedral Theory and Linear Programming

We consider the linear program (P) in canonical form

\[
(P) \quad \min \ c^T x \text{ s.t. } Ax \geq b
\]

with cost vector \( c \in \mathbb{R}^n \), right-hand side \( b \in \mathbb{R}^m \), and constraint matrix \( A \in \mathbb{R}^{m \times n} \). We denote the rows of \( A \) by \( a_i^T \) so that \( a_i \) is a column vector in \( \mathbb{R}^n \) for \( i = 1, \ldots, m \). The polyhedron \( S = \{ x \in \mathbb{R}^n : Ax \geq b \} \) is called the feasible set or region of (P), and a point \( x \in \mathbb{R}^n \) is said to be feasible to (P) if \( x \in S \). Each halfspace \( a_i^T x \geq b_i \) is called a constraint of (P), and without loss of generality we assume that \( m \geq n \) so that rank \( A \leq n \) (otherwise we could introduce redundant constraints/duplicate rows).

**Exercise 6** (Math 5593 LP Midterm, UCD, Fall 2011). A constraint is said to be redundant for the system of linear inequalities \( Ax \geq b \) if its addition or removal does not alter the set of feasible solutions. Set up an LP to determine if the constraint \( p^T x \geq q \) is redundant for the system \( Ax \geq b \), and explain.

**Solution.** To check if the constraint \( p^T x \geq q \) is redundant for \( Ax \geq b \), we can solve the problem

\[
\min p^T x \text{ s.t. } Ax \geq b.
\]

Then, if the optimal objective \( p^T x^* \) is greater than or equal to \( q \), it follows that \( p^T x \geq q \) for all feasible \( x \in \{ x : Ax \geq b \} \) and the constraint \( p^T x \geq q \) is redundant for \( Ax \geq b \). Similarly, if \( p^T x^* < q \), then there exists a feasible \( x^* \in \{ x : Ax \geq b \} \) that does not belong to \( \{ x : Ax \geq b, p^T x \geq q \} \), and the constraint \( p^T q \geq x \) is not redundant for \( Ax \geq b \).

4.1 Equivalence between Vertex Solutions and Extreme Points

**Definition 12.** Given a feasible point \( \hat{x} \in S \), a constraint \( a_i^T x \geq b_i \) is said to be active or inactive at \( \hat{x} \) if \( a_i^T \hat{x} = b_i \) or \( a_i^T \hat{x} > b_i \), respectively, the active set \( A(\hat{x}) \) at \( \hat{x} \) is the index set of all active constraints at \( \hat{x} \)

\[
A(\hat{x}) = \{ i = 1, \ldots, m : a_i^T \hat{x} = b_i \},
\]

and the active constraint matrix \( A_{\hat{x}} \) is the matrix of rows of \( A \) with indices in \( A(\hat{x}) \).

**Example 1.** Let \( A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) and \( b = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \). Consider the three feasible points \( x = (1,0)^T, y = (0.5,0.5)^T, \) and \( z = (0.25,0.25)^T \). Then \( A(x) = \{ 2, 3 \}, A(y) = \{ 3 \}, \) and \( A(z) = \emptyset \), so that \( A_x = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \), \( A_y = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \), and \( A_z = \emptyset \).

**Definition 13.** A feasible point \( \hat{x} \in S \) is called a vertex of \( S \) if rank \( A_{\hat{x}} \) is \( n \). It is said to be non-degenerate or degenerate if \( |A(\hat{x})| = n \) or \( |A(\hat{x})| > n \), respectively.

**Theorem 8.** A point \( \hat{x} \) is a vertex of \( S \) if and only if it is an extreme point of \( S \).

**Proof.** For notational convenience, we will denote the matrix of rows of \( A \) with indices not in \( A(\hat{x}) \) by \( A^C_{\hat{x}} \) (inactive constraint matrix), and use the analogous notation \( b^C_{\hat{x}} \) and \( b^C_{\hat{x}} \) for the vectors of components of \( b \) corresponding to rows in \( A_{\hat{x}} \) and \( A^C_{\hat{x}} \), respectively.

Only If: Let \( \hat{x} \) be a vertex, so rank \( A_{\hat{x}} = n \) where \( A_{\hat{x}} \hat{x} = b_{\hat{x}} \), and let \( \hat{x} = \alpha \tilde{y} + (1 - \alpha) \tilde{z} \) for some \( \tilde{y} \) and \( \tilde{z} \) in \( S \) and \( 0 < \alpha < 1 \), so

\[
A_{\hat{x}} = \alpha A_{\hat{x}} \tilde{y} + (1 - \alpha) A_{\hat{x}} \tilde{z} \geq \alpha b_{\hat{x}} + (1 - \alpha) b_{\hat{x}} = b_{\hat{x}}
\]

yielding \( A_{\hat{x}} \hat{x} = A_{\hat{x}} \tilde{y} = A_{\hat{x}} \tilde{z} = b_{\hat{x}} \). Because rank \( A_{\hat{x}} = n \), the solution to this linear system is unique, \( \hat{x} = \tilde{y} = \tilde{z} \), showing that \( \hat{x} \) is an extreme point of \( S \).

If: Now let \( \hat{x} \) be an extreme point and suppose that \( \hat{x} \) is not a vertex, so rank \( A_{\hat{x}} < n \) such that there exists a nonzero vector \( d \) such that \( A_{\hat{x}} d = 0 \). Because \( A^C_{\hat{x}} \hat{x} > b^C_{\hat{x}} \), this means that there exists a (sufficiently small) \( \varepsilon > 0 \) such that \( A^C_{\hat{x}} \hat{x} + \varepsilon d > b^C_{\hat{x}} \) also and, thus, such that the two (distinct) points \( \hat{x} \pm \varepsilon d \) are feasible and \( x = 0.5(x + \varepsilon d) + 0.5(x - \varepsilon d) \) in contradiction to the extremity of \( x \). The proof is complete. \( \square \)
4.2 The Fundamental Theorem of Linear Programming

Definition 14. A point \( x^* \) is said to be optimal for \((P)\) if it is feasible, \( x^* \in S \), and if \( c^T x^* \leq c^T x \) for all other feasible \( x \in S \).

Note that if \( c \neq 0 \), then the above definition defines a point \( x^* \) as optimal if and only if there the set \( H = \{ x \in \mathbb{R}^n : c^T x = c^T x^* \} \) is a supporting hyperplane at \( x^* \) with normal vector \( c \) that contains the feasible set \( S \) in its associated positive halfspace, \( S \subseteq H^+ \). With this intuition (and several results we have proven so far), the following fundamental theorem of linear programming is a piece of cake (and follows independently of the simplex method).

Theorem 9 (Fundamental Theorem of Linear Programming).

(a) If there exists a feasible point for \((P)\), then there exists a feasible vertex.

(b) If there exists an optimal point for \((P)\), then there exists an optimal vertex.


4.3 Optimality Conditions and Farkas Lemma

Definition 15. Given a (nonzero) vector \( c \in \mathbb{R}^n \), a direction \( d \in \mathbb{R}^n \) is called an ascent, descent, or orthogonal direction with respect to \( c \) if \( c^T d > 0 \), \( c^T d < 0 \), or \( c^T d = 0 \), respectively. It is said to be feasible at \( \hat{x} \) for \((P)\) if \( A\hat{x} \geq 0 \).

Theorem 10. A feasible point \( x \) is optimal for \((P)\) if and only if there exists no feasible descent directions at \( x \) with respect to \( c \):

\[
\begin{align*}
  c^T d &\geq 0 \text{ for all } d \in \mathbb{R}^n \text{ s.t. } A_d \geq 0. 
\end{align*}
\]  


The optimality condition in Theorem 10 is highly impractical in practice (why?). To derive an alternative condition, we use the following famous lemma.

Lemma 1 (Farkas Lemma). Let \( A \) be a matrix in \( \mathbb{R}^{k \times n} \) and \( c \) be a vector in \( \mathbb{R}^n \). Then \( c^T d \geq 0 \) for all \( d \in \mathbb{R}^n \) such that \( Ad \geq 0 \) if and only if there exists a nonnegative vector \( y \) in \( \mathbb{R}^k \) such that \( A^T y = c \).

Proof. For the “if” direction, let \( A^T y = c \) for some \( y \geq 0 \), and \( d \) be any vector in \( \mathbb{R}^n \) such that \( Ad \geq 0 \). Then

\[
  c^T d = (A^T y)^T d = y^T (Ad) \geq 0. 
\]  

For the “only if” direction, assume that \( c^T d \geq 0 \) for all \( d \in \mathbb{R}^n \) such that \( Ad \geq 0 \) and define the set \( S = \{ s \in \mathbb{R}^n : A^T y = s \text{ for some } y \geq 0 \} \). It is easy to show that \( S \) is a closed convex cone which contains \( 0 \) (from \( y = 0 \)) and \( a_i \) (from the \( i \)th unit vector \( y = e_i \)) for all \( i = 1, \ldots, k \). By contradiction, now suppose that \( c \notin S \) and apply Theorem 3 to get a separating hyperplane with nonzero normal vector \( d \in \mathbb{R}^n \) and \( z \in S \) such that

\[
  d^T c < d^T z \leq d^T x \text{ for all } x \in S, 
\]  

particularly \( d^T c < d^T 0 = 0 \) because \( 0 \in S \). Moreover, because \( S \) is a convex cone and contains \( z \) and \( a_i \) for all \( i = 1, \ldots, k \), it follows that \( z + a_i \in S \) also and, thus, that

\[
  d^T z \leq d^T (z + a_i) \iff d^T a_i \geq 0 \text{ for all } i = 1, \ldots, k 
\]  

and hence \( Ad \geq 0 \) in contradiction to \( c^T d < 0 \). The proof is complete.

Farkas Lemma comes in many different variants and can equivalently be phrased in the form of a “Theorem of the alternative.”
Theorem 11 (Farkas Lemma). Given a matrix $A \in \mathbb{R}^{k \times n}$ and a vector $c \in \mathbb{R}^n$, one and only one of the following systems admits a solution.

\[(I) \quad A^T y = c, \quad y \geq 0 \quad (II) \quad Ad \geq 0, \quad c^T d < 0.\]  

(29)

Applying Farkas Lemma with $A_x$ replacing $A$ to Theorem 10, we have thus shown the following optimality condition for (P).

Theorem 12. A feasible point $x$ is optimal for (P) if and only if there exists a vector $y \geq 0$ such that

\[A^T x = c.\]

Theorem 12 states the a feasible point $x$ is optimal for (P) if and only its cost vector $c$ can be written as a positive linear combination of the normal vectors $a_i$ of those hyperplanes corresponding to the active constraints at $x$:

\[c = \sum_{i \in A(x)} y_i a_i \text{ for some } y_i \geq 0, i \in A(x).\]  

(30)

4.4 Duality and Complementary Slackness

The well-known duality and complementary slackness properties of linear programs (in canonical form) now follow as an immediate (and hopefully geometrically motivated) consequence of the above.

Theorem 13 (First-Order Optimality (Karush-Kuhn-Tucker) Conditions). A point $x$ is optimal for (P) if and only if there exists a nonzero vector $y \in \mathbb{R}^m$ such that

\[(i) \quad Ax \geq b \quad (ii) \quad A^T y = c \quad (iii) \quad y_i(a_i^T x - b) = 0 \text{ for all } i = 1, \ldots, m.\]  

(31)

Definition 16. Given (P), a nonnegative vector $y \in \mathbb{R}^m$ is said to be dual feasible for (P) if $A^T y = c$, a feasible point $x \in S$ is also said to be primal feasible, and the pair $(x, y)$ is also called a primal-dual feasible solution.

Accordingly, conditions (i) and (ii) in Theorem 13 are also called primal and dual feasibility, and (iii) is called the complementary slackness condition.

Corollary 2 (Weak and Strong Duality). If $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ is a primal-dual feasible solution for (P), then $c^T x \geq b^T y$ with equality if and only if $(x, y)$ satisfy complementarity slackness.

Proof. Let $x \in S$ and $y \geq 0$ be primal and dual feasible, respectively, so $Ax \geq b$ and $A^T y = c$ yielding that

\[c^T x = (A^T y)^T x = y^T (Ax) \geq y^T b = b^T y.\]  

(32)

In particular, the inequality holds with equality if and only if $y_i = 0$ whenever $a_i^T x > b_i$, or equivalently, if and only if $y_i(a_i^T x - b_i) = 0$.  

Hence, we have shown that a (primal) feasible solution $x$ for (P) is optimal if and only if there exists a (dual) feasible solution $y$ that is optimal for the dual linear program (D)

\[(D) \quad \max b^T y \text{ s.t. } A^T y = c, \quad y \geq 0\]  

(33)

if and only if there exists a primal-dual feasible solution $(x, y)$ for (P) and (D) such that $c^T x = b^T y$.

Remark 2. Note that the dual LP derived above is identical to an LP in standard form if we relabel $A^T$ as $A$, $b$ as $c$, $c$ as $b$, and $y$ as $x$:

\[\max c^T x \text{ s.t. } Ax = b, \quad x \geq 0.\]  

(34)

In this case, the Farkas systems in Theorem 11 take the following, most common form:

\[(I) \quad Ax = b, \quad x \geq 0 \quad (II) \quad A^T y \geq 0, \quad b^T y < 0.\]  

(35)