ORF 523	LECTURE 5	Spring 2017, Princeton University
Instructor: A.A. Ahmadi		
Scribe: G. Hall		Thursday, February 23, 2017

When in doubt on the accuracy of these notes, please cross check with the instructor's notes, on aaa.princeton.edu/orf523*. Any typos should be emailed to gh4@princeton.edu.*

In this lecture, we will cover:

- Separation of convex sets with hyperplanes
- The Farkas lemma
- Strong duality of linear programming

1 Separating hyperplane theorems

The following is one of the most fundamental theorems about convex sets:

Theorem 1. *Let C and D be two convex sets in $\mathbb{R}^n$ that do not intersect (i.e., $C \cap D = \emptyset$). Then, there exists $a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$, such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.*

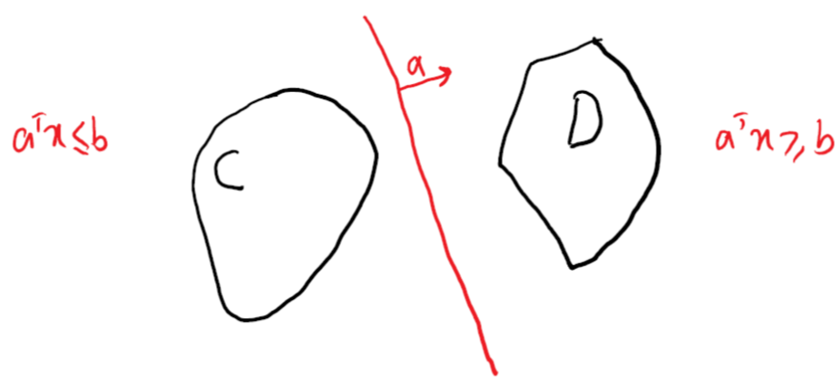
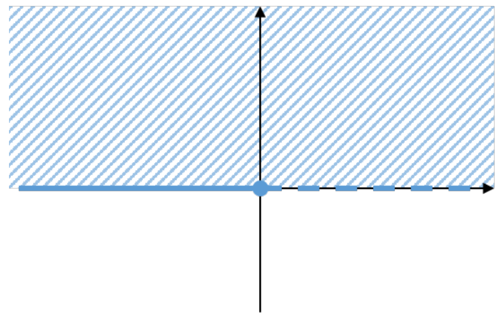


Figure 1: An illustration of Theorem 1.

We remark that neither inequality in the conclusion of Theorem 1 can be made strict. Thanks to Fermi Ma for asking this question and Kaizheng Wang for giving the following nice example:



Consider the set drawn above which we denote by A (the dotted line is not included in the set). We would like to separate it from its complement $\bar{A}$. The two sets are convex and do not intersect. The conclusion of Theorem 1 holds with $a = (1, 0)^T$ and $b = 0$. Nevertheless, there does not exist a, b for which $a^T x \leq b, \forall x \in A$ and $a^T x > b, \forall x \in \bar{A}$.

In the case of the picture in Figure 1, the sets C and D are *strictly separated*. This means that $\exists a, b$ s.t. $a^T x < b, \forall x \in C$ and $a^T x > b,\ \forall x \in D$.
Strict separation may not always be possible, even when both C and D are closed. You can convince yourself of this fact by looking at Figure 2.

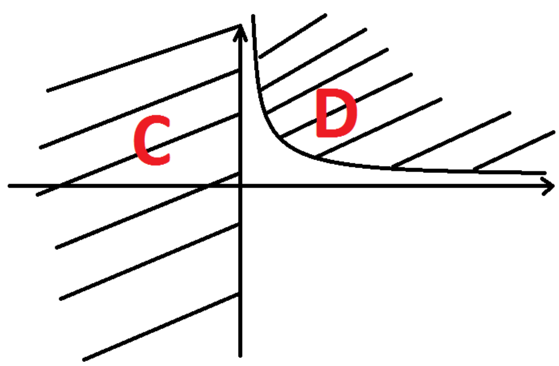


Figure 2: Closed convex sets cannot always be strictly separated.

We will prove a special case of Theorem 1 which will be good enough for our purposes (and we will prove strict separation in this special case).

Theorem 2. *Let C and D be two closed convex sets in $\mathbb{R}^n$ with at least one of them bounded, and assume $C \cap D = \emptyset$. Then $\exists a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$ such that*

$$a^T x > b,\ \forall x \in D \text{ and } a^T x < b,\ \forall x \in C.$$

Proof: Our proof follows [1] with a few minor deviations. Define

$$\begin{aligned} dist(C, D) = &\inf ||u - v|| \\ &\text{s.t. } u \in C, v \in D \end{aligned}$$

The infimum is achieved (why?) and is positive (why?). Let $c \in C$ and let $d \in D$ be the points that achieve it. Let

$$a = d - c, b = \frac{||d||^2 - ||c||^2}{2}.$$

(Note that $a \neq 0$). Our separating hyperplane will be a function $f(x) = a^T x - b$. We claim that

$$f(x) > 0, \ \forall x \in D \text{ and } f(x) < 0, \ \forall x \in C.$$

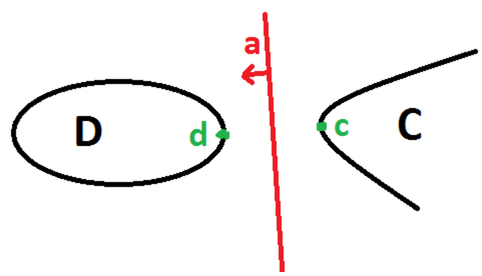


Figure 3: Illustration of the proof of Theorem 2

If you are wondering why b is chosen as above, observe that

$$f\left(\frac{c+d}{2}\right) = (d-c)^T \left(\frac{c+d}{2}\right) - \frac{||d||^2 - ||c||^2}{2} = 0.$$

We show that $f(x) > 0$ for all $x \in D$. The proof that $f(x) < 0$ for all $x \in C$ is identical.

Suppose for the sake of contradiction that $\exists \bar{d} \in D$ with $f(\bar{d}) \leq 0$.

$$\Rightarrow (d-c)^T \bar{d} - \frac{||d||^2 - ||c||^2}{2} \leq 0. \tag{1}$$

Define $g(x) = ||x - c||^2$. We claim that $\bar{d} - d$ is a descent direction for g at d. Indeed,

$$\begin{aligned}\nabla g^T(d)(\bar{d} - d) &= (2d - 2c)^T(\bar{d} - d) \\ &= 2(-||d||^2 + d^T\bar{d} - c^T\bar{d} + c^T d) \\ &= 2(-||d||^2 + (d - c)^T\bar{d} + c^T d) \\ &\leq 2\left(-||d||^2 + \frac{||d||^2 - ||c||^2}{2} + c^T d\right) \\ &= -||d||^2 - ||c||^2 + 2c^T d \\ &= -||d - c||^2 < 0\end{aligned}$$

where the first equality is obtained as

$$g(x) = (x - c)^T(x - c) = x^T x - 2c^T x + c^T c \Rightarrow \nabla g(x) = 2x - 2c,$$

the first inequality is obtained from (1) and the second inequality is implied by the fact that $d \neq c$.

Hence $\exists \bar{a} > 0$ s.t. $\forall \alpha \in (0, \bar{\alpha})$

$$g(d + \alpha(d - \bar{d})) < g(d)$$

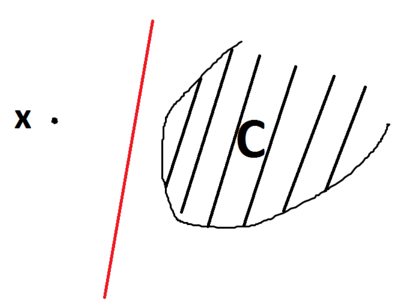
i.e.,

$$||d + \alpha(d - \bar{d}) - c||^2 < ||d - c||^2.$$

But this contradicts that d was the closest point to c. $\square$

The following is an important corollary.

Corollary 1. *Let $C \subseteq \mathbb{R}^n$ be a closed convex set and $x \in \mathbb{R}^n$ a point not in C. Then x and C can be strictly separated by a hyperplane.*



2 Farkas Lemma and strong duality

2.1 Farkas Lemma

Theorem 3 (Farkas Lemma)**.** *Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then exactly one of the following sets must be empty:*

(i) $\{x | \ Ax = b, x \geq 0\}$

(ii) $\{y | \ A^T y \leq 0, b^T y > 0\}$

Remark:

- Systems (i) and (ii) are called *strong alternatives*, meaning that exactly one of them can be feasible. *Weak alternatives* are systems where at most one can be feasible.
- This theorem is particularly useful for proving infeasibility of an LP via an explicit and easily-verifiable certificate. If somebody gives you a y as in (ii), then you are convinced immediately that (i) is infeasible (see proof).

Geometric interpretation of the Farkas lemma:

The geometric interpretation of the Farkas lemma illustrates the connection to the separating hyperplane theorem and makes the proof straightforward. We need a few definitions first.

Definition 1 (Cone)**.** *A set $K \subseteq \mathbb{R}^n$ is a* cone *if $x \in K \Rightarrow \alpha x \in K$ for any scalar $\alpha \geq 0$.*

Definition 2 (Conic hull)**.** *Given a set S, the* conic hull *of S, denoted by $cone(S)$, is the set of all* conic combinations *of the points in S, i.e.,*

$$cone(S) = \left\{ \sum_{i=1}^{n} \alpha_i x_i | \ \alpha_i \geq 0, x_i \in S \right\}.$$

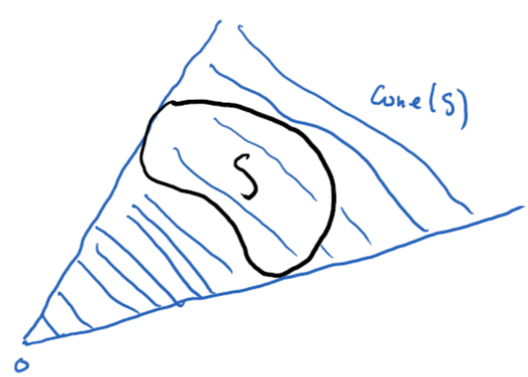


Figure 4: An illustration of the notion of a conic hull

The geometric interpretation of Farkas lemma is then the following. Let $\tilde{a}_1, \ldots, \tilde{a}_n$ denote the columns of A and let $cone\{\tilde{a}_1, \ldots, \tilde{a}_n\}$ be the cone of all their nonnegative combinations. If $b \notin cone\{\tilde{a}_1, \ldots, \tilde{a}_n\}$, then we can separate it from the cone with a hyperplane.

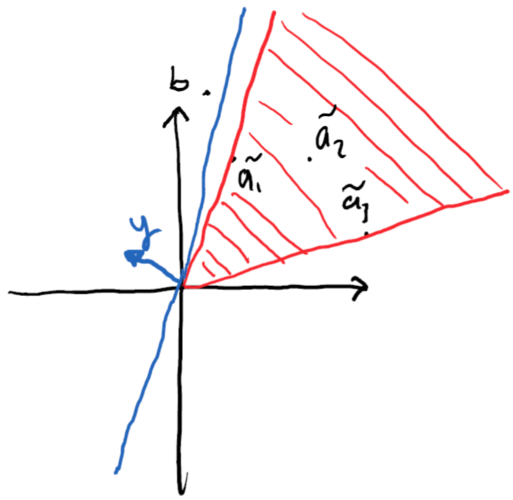
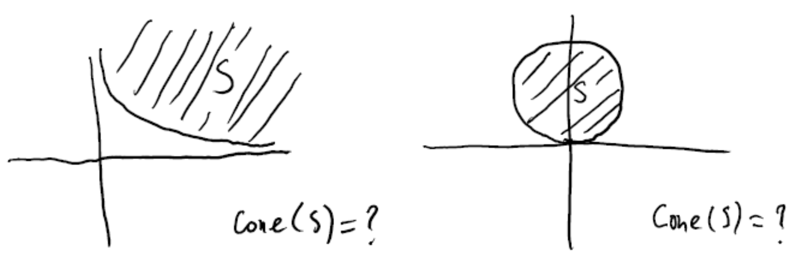


Figure 5: Geometric interpretation of the Farkas lemma

Proof of Farkas Lemma (Theorem 3): $(ii) \Rightarrow (i)$ This is the easy direction. Suppose the contrary: $\exists x \geq 0$ such that $Ax = b$. Then $x^T A^T y = b^T y > 0$. But $x \geq 0$, $A^T y \leq 0 \Rightarrow x^T A^T y \leq 0$. Contradiction.

$(i) \Rightarrow (ii)$ Let $\tilde{a}_1, \ldots, \tilde{a}_n$ be the columns of a matrix A. Let $C := cone\{\tilde{a}_1, \ldots, \tilde{a}_n\}$. Note that C is convex (why?) and closed. The closedness takes some thought. Note that conic hulls of closed (or even compact) sets may not be closed.



We argue that if $S := \{s_1, \ldots, s_n\}$ is a finite set of points, then $cone(S)$ is closed. Hence C is a closed convex set.

Let $\{z_k\}_k$ be a sequence of points in $cone(S)$ converging to a point $\bar{z}$. Consider the following linear program[1]:

$$\begin{aligned}
&\min_{\alpha,z} ||z-\bar{z}||_\infty \\
&\text{s.t. } \sum_{i=1}^{n} \alpha_i s_i = z \\
&\alpha_i \geq 0.
\end{aligned}$$

The optimal value of this problem is greater or equal to zero as the objective is a norm. Furthermore, for each z_k, there exists $\alpha_{(k)}$ that makes the pair $(z_k, \alpha_{(k)})$ feasible to the LP (since $z_k \in cone(S)$). As the z_k's get arbitrarily close to $\bar{z}$, we conclude that the optimal value of the LP is zero. Since LPs achieve their optimal values, it follows that $\bar{z} \in cone(S)$.

We are now ready to use the separating hyperplane theorem. We have $b \notin C$ by the assumption that (i) is infeasible. By Corollary 1, the point b and the set C can be (even strictly) separated; i.e.,

$$\exists y \in \mathbb{R}^m,\ y \neq 0,\ r \in \mathbb{R} \text{ s.t. } y^T z \leq r\ \forall z \in C \text{ and } y^T b > r.$$

Since $0 \in C$, we must have $r \geq 0$. If $r > 0$, we can replace it by $r' = 0$. Indeed, if $\exists z \in C$ s.t. $y^T z > 0$, then $y^T(\alpha z)$ can be arbitrarily large as $\alpha \to \infty$ while we know that $\alpha z \in C$. So

$$y^T z \leq 0,\ \forall z \in C \text{ and } y^T b > 0.$$

Since $\tilde{a}_1, \ldots, \tilde{a}_n \in C$, we see that $A^T y \leq 0$. $\square$

We remark that the Farkas lemma can be directly proven from strong duality of linear programming. The converse is also true! We will show these facts next. Note that there are other proofs of LP strong duality; e.g., based on the simplex method. However the simplex-based proof does not generalize to broader classes of convex programs, while the separating hyperplane based proofs do.

[1] Convince yourself that this can be rewritten as a linear program.

2.2 Farkas lemma from LP strong duality

Consider the primal-dual LP pair:

$$(P)\quad \begin{bmatrix} \min 0 \\ Ax = b \\ x \geq 0 \end{bmatrix} \qquad \text{and} \qquad (D)\quad \begin{bmatrix} \max b^T y \\ A^T y \leq 0 \end{bmatrix}$$

Note that (D) is trivially feasible (set $y = 0$). So if (P) is infeasible, then (D) must be unbounded or else strong duality would imply that the two optimal values should match, which is impossible since (P) by assumption is infeasible.
But (D) unbounded $\Rightarrow \exists y$ s.t. $A^T y \leq 0,\ b^T y > 0$. $\square$

2.3 LP strong duality from Farkas lemma

Theorem 4 (Strong Duality)**.** *Consider a primal-dual LP pair:*

$$(P)\quad \begin{bmatrix} \min\ c^T x \\ Ax = b \\ x \geq 0 \end{bmatrix} \qquad \textit{and} \qquad (D)\quad \begin{bmatrix} \max\ b^T y \\ A^T y \leq c \end{bmatrix}$$

If (P) has a finite optimal value, then so does (D) and the two values match.

Remark: If you don't recall how to write down the dual of an LP, look up the first few pages of Chapter 5 of [1]. The derivation there works more broadly (not just LP).

An alternative way of deriving the dual is the following. Recall that the goal of duality is to provide lower bounds on the primal (if the primal is a minimization problem). Here, we will try to find the largest lower bound on (P). Hence, we aim to solve

$$\begin{aligned} &\max_{\gamma} \gamma \\ &\text{s.t. } \forall x, \begin{bmatrix} Ax = b \\ x \geq 0 \end{bmatrix} \Rightarrow \gamma \leq c^T x \end{aligned}$$

Notice that a sufficient condition[2] for the implication to hold is if $\exists \eta \in \mathbb{R}^m, \mu \in \mathbb{R}^n$ with $\mu \geq 0$ such that

$$\forall x, c^T x - \gamma = \eta^T (Ax - b) + \mu^T x.$$

[2]It turns out that this sufficient condition is also necessary! This is the strong duality theorem.

Indeed, if x is such that $Ax = b, x \geq 0$, then $\eta^T(Ax - b) = 0$ and $\mu^T x \geq 0$, hence $c^T x - \gamma \geq 0$. As a consequence, one can propose a stronger reformulation of the initial problem:

$$\begin{aligned}
&\max_{\gamma,\eta,\mu} \gamma \\
&\text{s.t. } \forall x, c^T x - \gamma = \eta^T(Ax - b) + \mu^T x, \\
&\mu \geq 0.
\end{aligned}$$

To get rid of the quantifier $\forall x$, we notice now that two affine functions of x are equal if and only if their coefficients are equal. Therefore, the previous problem is equivalent to

$$\begin{aligned}
&\max_{\gamma,\eta,\mu} \gamma \\
&\text{s.t. } c = A^T \eta + \mu \\
&\gamma = \eta^T b \\
&\mu \geq 0.
\end{aligned}$$

Simple rewriting gives:

$$\begin{aligned}
&\max_{\eta} b^T \eta \\
&\text{s.t. } c \geq A^T \eta,
\end{aligned}$$

which is indeed our dual problem. □

To prove strong duality from Farkas, it is useful to first prove a variant of the Farkas lemma. This variant comes handy when one wants to prove infeasibility of an LP in inequality form.

Lemma 1 (Farkas Variant)**.** *Let $A \in \mathbb{R}^{m \times n}$.*

$$\{x |\ Ax \leq b\} \text{ is empty } \Leftrightarrow \exists \lambda \geq 0 \text{ s.t. } \lambda^T A = 0,\ \lambda^T b < 0.$$

Proof:

($\Leftarrow$) Easy (why?)

($\Rightarrow$) Rewrite the LP in standard form and apply the (standard) Farkas lemma:

$$\{x \mid Ax \leq b\} \text{ empty} \Leftrightarrow \{(x^+, x^-, s) \mid A(x^+ - x^-) + s = b, s \geq 0, x^+ \geq 0, x^- \geq 0\} \text{ empty}$$

$$\Leftrightarrow \{(x^+, x^-, s) \mid \begin{pmatrix} A & \big| & -A & \big| & I \end{pmatrix} \begin{pmatrix} x^+ \\ x^- \\ s \end{pmatrix} = b,\ x^+, x^-, s \geq 0\} \text{ empty}$$

$$\Rightarrow \exists \lambda \text{ s.t. } b^T \lambda < 0, \begin{pmatrix} A^T \\ -A^T \\ I \end{pmatrix} \lambda \geq 0 \Rightarrow \exists \lambda \text{ s.t. } b^T \lambda < 0,\ A^T \lambda = 0,\ \lambda \geq 0.\ \square$$

Proof of LP strong duality from the Farkas lemma: Consider the primal dual pair:

$$(P) \quad \begin{bmatrix} \min\ c^T x \\ Ax = b \\ x \geq 0 \end{bmatrix} \qquad \text{and} \qquad (D) \quad \begin{bmatrix} \max\ b^T y \\ A^T y \leq c \end{bmatrix}$$

Assume the optimal value of (P) is finite and equal to p^*. We would be done if we prove that the following inequalities are feasible:

$$\begin{bmatrix} y^T b \geq p^* \\ A^T y \leq c \end{bmatrix}. \tag{2}$$

Indeed, any y satisfying $A^T y \leq c$ must also satisfy $y^T b \leq p^*$ by weak duality (whose proof is trivial), so we would get that $y^T b = p^*$. Let's rewrite (2) slightly:

$$\begin{pmatrix} A^T \\ -b^T \end{pmatrix} y \leq \begin{pmatrix} c \\ -p^* \end{pmatrix}.$$

Suppose these inequalities were infeasible. Then, the Farkas lemma variant would imply that

$$\exists \lambda := \begin{pmatrix} \tilde{\lambda} \\ \lambda_0 \end{pmatrix} \geq 0 \text{ s.t. } \tilde{\lambda}^T A^T - \lambda_0 b^T = 0 \text{ and } \tilde{\lambda}^T c - \lambda_0 p^* < 0 \Rightarrow A\tilde{\lambda} = \lambda_0 b,\ c^T\tilde{\lambda} < \lambda_0 p^*.$$

We consider two cases:

- Case 1: $\lambda_0 = 0 \Rightarrow A\tilde{\lambda} = 0,\ c^T\tilde{\lambda} < 0$. Recall that we are assuming that (P) has a finite optimal value. Let x^* be an optimal solution of (P) and let $x = x^* + \tilde{\lambda}$. Then $x \geq 0$ and

$$Ax = Ax^* + A\tilde{\lambda} = Ax^* = b$$

  as x^* is feasible. Furthermore,

$$c^T x = c^T x^* + c^T\tilde{\lambda} = p^* + c^T\tilde{\lambda} < p^*$$

  which contradicts the fact that p^* is the primal optimal value.

- Case 2: $\lambda_0 > 0$. Let $x = \frac{\tilde{\lambda}}{\lambda_0}$. Then $Ax = \frac{1}{\lambda_0} A\tilde{\lambda} = \frac{\lambda_0}{\lambda_0} b = b$, $x \geq 0$ and

$$c^T x = c^T \frac{\tilde{\lambda}}{\lambda_0} < \frac{1}{\lambda_0}\lambda_0 p^* = p^*.$$

  This contradicts p^* being the primal optimal value. □

Notes

Further reading for this lecture can include Chapter 2 of [1].

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, http://stanford.edu/ boyd/cvxbook/, 2004.