Mathematical Tools in Computer Science, Winter 2006

Lecture 7

Lecturer: Nati Linial *Scribe: Ofer Dekel*

1 Separating Hyperplane Theorems

In the previous lecture, we discussed separating hyperplane theorems. Let S_1 and S_2 be two subsets of $\mathbb{R}^n$, let $\mathbf{u}$ be a vector in $\mathbb{R}^m$, and let μ be a real scalar. We say that the hyperplane $H_{\mathbf{u},\mu} = \{\mathbf{x} \in \mathbb{R}^m \ : \ \langle \mathbf{u}, \mathbf{x}\rangle = \mu\}$ separates S_1 from S_2 if $\langle \mathbf{u}, \mathbf{x}\rangle \geq \mu$ for all $\mathbf{x} \in S_1$ and $\langle \mathbf{u}, \mathbf{x}\rangle \leq \mu$ for all $\mathbf{x} \in S_2$. We say that the separation is *strong* if $\langle \mathbf{u}, \mathbf{x}\rangle > \mu$ for all $\mathbf{x} \in S_1$ and $\langle \mathbf{u}, \mathbf{x}\rangle < \mu$ for all $\mathbf{x} \in S_2$. We say that the separation is *proper* if $H_{\mathbf{u},\mu}$ does not contain $S_1 \cup S_2$.

The following theorems guarantee the existence of a separating hyperplane under certain conditions.

Theorem 1 *Let K and C be disjoint nonempty convex subsets of $\mathbb{R}^m$. Suppose K is compact and C is closed. Then there exist $\mathbf{u} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that the hyperplane defined by $H_{\mathbf{u},\mu} = \{\mathbf{x} \in \mathbb{R}^m \ : \ \langle \mathbf{u}, \mathbf{x}\rangle = \mu\}$ strongly separates K from C. In other words, $\langle \mathbf{u}, \mathbf{x}\rangle > \mu$ for all $\mathbf{x} \in K$ and $\langle \mathbf{u}, \mathbf{x}\rangle < \mu$ for all $\mathbf{x} \in C$.*

Theorem 2 *Let C_1 and C_2 be two non-empty convex subsets of $\mathbb{R}^m$ such that their relative interiors are disjoint. Then there exist $\mathbf{u} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that the hyperplane defined by $H = \{\mathbf{x} \in \mathbb{R}^m \ : \ \langle \mathbf{u}, \mathbf{x}\rangle = \mu\}$ properly separates C_1 from C_2. In other words, $\langle \mathbf{u}, \mathbf{x}\rangle \geq \mu$ for all $\mathbf{x} \in C_1$ and $\langle \mathbf{u}, \mathbf{x}\rangle \leq \mu$ for all $\mathbf{x} \in C_2$, and it is not the case that $C_1 \cup C_2 \subseteq H$.*

Detailed proofs of both theorems can be found in [1].

2 The Farkas Lemma

Assume we are given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $\mathbf{b} \in \mathbb{R}^m$. We are asked to determine whether there exists a vector $\mathbf{x} \in \mathbb{R}^n$ which satisfies the following system of constraints,

$$A\mathbf{x} \leq \mathbf{b} \quad . \tag{1}$$

If the answer is yes, then we have a short *certificate* which proves so, namely, we can present a vector $\mathbf{x} \in \mathbb{R}^n$ which solves the system as our proof. However, if the answer is no, is there also a short certificate which proves this ? The answer lies in the Farkas lemma.

Lemma 3 (Farkas 1902 [3]) *The system $A\mathbf{x} \leq \mathbf{b}$ has no solution iff there exists $\mathbf{u} \in \mathbb{R}^m_+$ which satisfies the constraints $\mathbf{u}^\top A = 0$, $\langle u, \mathbf{b}\rangle < 0$.*

Proof

Let $\mathbf{u} \in \mathbb{R}^m_+$ be a vector which satisfies the constraints $\mathbf{u}^\top A = 0$, $\langle \mathbf{u}, \mathbf{b}\rangle < 0$. Assume that there exists $\mathbf{x} \in \mathbb{R}^n$ which satisfies the system $A\mathbf{x} \leq \mathbf{b}$. Combining our assumptions that $\mathbf{u} \geq 0$ and $A\mathbf{x} \leq \mathbf{b}$, we get, $\mathbf{u}^\top A\mathbf{x} \leq \mathbf{u}^\top \mathbf{b}$. On the other hand, we also assumed that $\mathbf{u}^\top A = 0$ and $\mathbf{u}^\top \mathbf{b} < 0$, which leads to,

$$0 = \mathbf{u}^\top A\mathbf{x} \leq \mathbf{u}^\top \mathbf{b} < 0 .$$

We have arrived at a contradiction, and we conclude that there is no feasible solution to $A\mathbf{x} \leq \mathbf{b}$.

Moving on to the opposite direction of our claim, assume that the system $A\mathbf{x} \leq \mathbf{b}$ has no feasible solution, and consider the set $C = \{\mathbf{b} - A\mathbf{x} \ : \ \mathbf{x} \in \mathbb{R}^n\}$. For any $\mathbf{x} \in \mathbb{R}^n$, we know that at least one element of the vector $A\mathbf{x}$ is greater than the respective element of $\mathbf{b}$, therefore, every vector in C has at least one negative element. Therefore, the sets C and $\mathbb{R}^m_+$ are disjoint. Moreover, both C and $\mathbb{R}^m_+$ are non-empty and convex. Thm. 2 states that there exist $\mathbf{u} \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that the hyperplane $H = \{\mathbf{z} \in \mathbb{R}^m \ : \ \langle \mathbf{u}, \mathbf{z} \rangle = \mu\}$ separates C from $\mathbb{R}^m_+$. In other words, for every $\mathbf{v} \in C$ and every $\mathbf{w} \in \mathbb{R}^m_+$ we have,

$$\langle \mathbf{u}, \mathbf{v} \rangle \leq \mu \leq \langle \mathbf{u}, \mathbf{w} \rangle ,$$

and we can assume w.l.o.g. that, **Nati - why is this ?**

$$\langle \mathbf{u}, \mathbf{v} \rangle < \mu \leq \langle \mathbf{u}, \mathbf{w} \rangle .$$

Our first conclusion is that $\mathbf{u} \geq 0$, since otherwise, we could make $\langle \mathbf{u}, \mathbf{w} \rangle$ arbitrarily small with the appropriate choice of $\mathbf{w}$. Since we also know that $\mathbf{w} \geq 0$ then $\langle \mathbf{u}, \mathbf{w} \rangle \geq 0$. Furthermore, we know that $\mu \leq 0$ since $\mathbf{w}$ can be chosen to be the zero vector. We now have,

$$\forall \mathbf{v} \in C \quad \forall \mathbf{w} \in \mathbb{R}^m_+ \qquad \langle \mathbf{u}, \mathbf{v} \rangle < \mu \leq 0 \leq \langle \mathbf{u}, \mathbf{w} \rangle .$$

Since every $\mathbf{v} \in C$ equals $\mathbf{b} - A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, we know that $\mathbf{u}^\top(\mathbf{b} - A\mathbf{x}) < 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Rearranging the terms in this inequality gives $\mathbf{u}^\top \mathbf{b} < \mathbf{u}^\top A\mathbf{x}$. It must be the case that $\mathbf{u}^\top A = 0$, otherwise the right-hand side of this inequality could be made arbitrarily small with an appropriate choice of $\mathbf{x}$, while the left-hand side remains constant. Additionally, we conclude that $\langle \mathbf{u}, \mathbf{b} \rangle < 0$. Overall, we have shown that $\mathbf{u}$ is a non-negative vector which satisfies the constraints $\mathbf{u}^\top A = 0$ and $\langle \mathbf{u}, \mathbf{b} \rangle < 0$. ■

3 Linear Programs in Canonical Form and Their Dual

A linear program is said to be in canonical form if it has the form,

$$\text{maximize } \mathbf{c}^\top \mathbf{x} \quad \text{s.t.} \quad A\mathbf{x} \leq \mathbf{b} \quad \mathbf{x} \geq 0, \tag{2}$$

where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^n$. Every linear program can be stated in canonical form, by applying the following rules:

- If the LP is a minimization problem, negate $\mathbf{c}$ and maximize.
- If a variable x_j is not required to be positive, replace it with two new variables, x_j^+ and x_j^-. Replace every appearance of x_j with $(x_j^+ - x_j^-)$.

- Replace every constraint of the form $\mathbf{a}^\top \mathbf{x} = b$ with two inequality constraints: $\mathbf{a}^\top \mathbf{x} \le b$ and $-\mathbf{a}^\top \mathbf{x} \le -b$.
- Replace every constraint of the form $\mathbf{a}^\top \mathbf{x} \ge b$ with the constraint $-\mathbf{a}^\top \mathbf{x} \le -b$.

Now consider a linear program given in canonical form, as in Eq. (2). Assume that the constraints of this problem have a feasible solution. If we know an upper bound and a lower bound on the maximum of this problem, then we can approximate its solution by solving multiple linear feasibility problems. Namely, we conduct a binary search for the greatest λ for which the following system is solvable:

$$\langle \mathbf{c}, \mathbf{x} \rangle \ \ge \ \lambda, \ A\mathbf{x} \ \le \ \mathbf{b}, \ \mathbf{x} \ge 0 \tag{3}$$

The above consists of $m + n + 1$ linear inequality constraints which can be rewritten in matrix form as $\bar{A}\mathbf{x} \le \bar{\mathbf{b}}$, where $\bar{A} \in \mathbb{R}^{(m+n+1)\times n}$ and $\bar{\mathbf{b}} \in \mathbb{R}^{m+n+1}$. If λ is smaller than the maximum of the original LP, then clearly there exists $\mathbf{x}$ which satisfies these constraints. On the other hand, if λ is greater than the maximum of the LP, the Farkas lemma tells us that there exists a vector $\bar{\mathbf{y}} \in \mathbb{R}_+^{m+n+1}$ which can prove this. Namely,

$$\bar{\mathbf{y}}^\top \bar{A} = 0, \ \langle \bar{\mathbf{y}}, \bar{\mathbf{b}} \rangle < 0, \ \bar{\mathbf{y}} \ge 0 \ \ . \tag{4}$$

We can now make the following observations:

- Coordinates $1, \ldots, m$ of $\bar{\mathbf{b}}$ are simply the respective components of $\mathbf{b}$. Coordinates $m+1, \ldots, m+n$ are all zeros, due to the n constraints in $-\mathbf{x} \le 0$. The last coordinate in $\bar{\mathbf{b}}$ equals $-\lambda$, from the constraint $\langle -\mathbf{c}, \mathbf{x} \rangle \le -\lambda$. The last coordinate of $\bar{\mathbf{y}}$ is positive, since otherwise the constraints of the original LP (without the constraint $\langle \mathbf{c}, \mathbf{x} \rangle \ge \lambda$) are themselves unsolvable. Therefore, we can assume w.l.o.g. that the last coordinate of $\bar{\mathbf{y}}$ equals 1 (otherwise divide $\bar{\mathbf{y}}$ by $\bar{y}_{m+n+1}$ and Eq. (4) still holds). Denoting $\mathbf{y} = (\bar{y}_1, \ldots, \bar{y}_m)$, we have that $\langle \bar{\mathbf{y}}, \bar{\mathbf{b}} \rangle = \langle \mathbf{y}, \mathbf{b} \rangle - \lambda$. Thus, the constraint $\langle \bar{\mathbf{y}}, \bar{\mathbf{b}} \rangle < 0$ can be written as $\langle \mathbf{y}, \mathbf{b} \rangle < \lambda$.
- $\bar{A}$ is a $(m+n+1) \times n$ matrix. The first m rows in $\bar{A}$ are the respective rows of A. The next n rows contain $-I$ (the negative of the identity matrix) from the n constraints in $\mathbf{x} \ge 0$. The last row in $\bar{A}$ is $-c$, from the constraint $\langle -\mathbf{c}, \mathbf{x} \rangle \le \lambda$. Therefore, coordinate j of the vector $\bar{\mathbf{y}}^\top \bar{A}$ can be written as $(\mathbf{y}^\top A)_j - y_{m+j} - c_j$. We can rewrite the constraint $(\bar{\mathbf{y}}^\top \bar{A})_j = 0$ as $(\mathbf{y}^\top A)_j = y_{m+j} + c_j$. Since $y_{m+j} \ge 0$ then $(\mathbf{y}^\top A)_j \ge c_j$. Written in matrix form, we have, $\mathbf{y}^\top A \ge \mathbf{c}^\top$.

In summary, to prove that λ is greater than the maximum of the original LP, we can find a vector $\mathbf{y} \in \mathbb{R}^m$ which satisfies the following constraints:

$$\mathbf{y}^\top A \ge \mathbf{c}^\top, \ \langle \mathbf{y}, \mathbf{b} \rangle < \lambda, \ \mathbf{y} \ge 0 \ \ . \tag{5}$$

A direct consequence from the Farkas lemma is that for every choice of λ, either Eq. (3) is feasible or Eq. (5) is feasible, but not both. If we relax the constraint $\langle \mathbf{y}, \mathbf{b} \rangle < \lambda$ to be $\langle \mathbf{y}, \mathbf{b} \rangle \le \lambda$ then there will be a single critical value of λ for which both systems have solutions. This critical lambda is the maximum of the original LP, but it is also the minimum of the problem,

$$\text{minimize} \;\; \langle \mathbf{y}, \mathbf{b} \rangle \qquad \text{s.t.} \qquad \mathbf{y}^\top A \geq \mathbf{c}^\top, \; \mathbf{y} \geq 0 \;\; .$$

The above is also a linear program, which we call the *dual* of the original program. We have basically proven the following theorem,

Theorem 4 *Define* $P = \{\mathbf{x} \in \mathbb{R}^n_+ \;\; : \;\; A\mathbf{x} \leq b\}$*. If* $P \neq \emptyset$ *and the linear functional* $\mathbf{c}^\top \mathbf{x}$ *is bounded over the domain* P *then the dual linear program, defined by,*

$$\text{minimize} \;\; \langle \mathbf{y}, \mathbf{b} \rangle \qquad s.t. \qquad \mathbf{y}^\top A \geq \mathbf{c}, \; \mathbf{y} \geq 0 \;\; ,$$

is also satisfiable and bounded, and its optimum equals that of the original problem.

A formal proof of this theorem can be found in [2].

References

[1] KC. Border. Separating hyperplane theorems. 2002. http://www.hss.caltech.edu/∼kcb/Notes/SeparatingHyperplane.pdf.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[3] J. Farkas. Uber die theorie der einfachen ungleichungen. *Journal fur die Reine und Angewandte Mathematik*, 124:1–24, 1902.

[4] M. Grotschel, L. Lovasz, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.