AM 221: Advanced Optimization Spring 2016

Prof. Yaron Singer *Lecture 7 — February 17th*

1 Overview

In the previous lectures we saw applications of duality to game theory and later to learning theory. In this lecture we'll see how to use duality to characterize optimal solutions of linear programs. We'll first discuss extreme points – informally corners of the polytope – and see that there are optimal solutions of LPs which are extreme points. We will later discuss a different characterization of extreme points called *basic feasible solutions*. Finally, we'll describe the simplex method which finds optimal solutions for linear program.

2 Extreme Points

Intuitively, extreme points are the corners of a polytope. More formally:

Definition. *Let $C \subseteq \mathbb{R}^n$ be a non-empty, closed convex set. Then $\bar{\mathbf{x}}$ is an* ***extreme point*** *of C if there are no two points $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\lambda \in (0,1)$ s.t. $\bar{\mathbf{x}} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$.*

Does every polyhedron have extreme points?

No. For example, half spaces. *So when does a set have an extreme point?*

Theorem 1. *Let $C \subseteq \mathbb{R}^n$ be a non-empty, closed, convex, set. Then, C has an extreme point if and only if C does not contain a line.*

Proof. [$\Longrightarrow$] Assume $C \subseteq \mathbb{R}^n$ is non-empty, closed, convex and let $\bar{\mathbf{x}}$ be an extreme in C. We will show that C does not contain a line. Assume, for purpose of contradiction that C contains a line, i.e. $\exists \mathbf{x} \in C$ s.t. $\{\mathbf{x} + \alpha\mathbf{d} \ : \ \alpha \in \mathbb{R}, \ \mathbf{d} \in \mathbb{R}^n\} \subseteq C$.

For any positive integer n, define:

$$\mathbf{x}^{(n)} = (1 - \frac{1}{n})\bar{\mathbf{x}} - \frac{1}{n}(\mathbf{x} + n\mathbf{d}).$$

By convexity of C each point $\mathbf{x}^{(i)}$ is also in C. Since C is convex, $\mathbf{x}^{(n)} \in C$, for all $n \in \mathbb{N}$. Since C is closed the limit of $\{\mathbf{x}^{(i)}\}_{i=1}^{\infty}$ is in C as well, thus:

$$\lim_{n\to\infty} \mathbf{x}^{(n)} = \lim_{n\to\infty} \bar{\mathbf{x}} + \mathbf{d} - \frac{1}{n}(\mathbf{x} - \bar{\mathbf{x}}) = \bar{\mathbf{x}} + \mathbf{d} \in C.$$

Similarly, $\bar{\mathbf{x}} - \mathbf{d} \in C$. But then, $\bar{\mathbf{x}}$ is a convex combination of two points in C since $\bar{\mathbf{x}} = \frac{1}{2}(\mathbf{x} + \mathbf{d}) + \frac{1}{2}(\mathbf{x} - \mathbf{d}) \in C$. Contradiction to $\bar{\mathbf{x}}$ being an extreme point.

[$\Longleftarrow$] In the other direction, let's assume that C does not contain a line, and we'll show that C has an extreme point. We will show this by induction on the dimension of C. If $C \subseteq \mathbb{R}$, the claim trivially holds since the boundary point is the extreme point. Let's assume this claim holds for $C \subseteq \mathbb{R}^{n-1}$ and we will show this holds for $C \subseteq \mathbb{R}^n$.

In $\mathbb{R}^n$, since C does not contain the line, we know it has a boundary point, $\mathbf{x}$. Let $H_{\mathbf{x}}$ be the supporting hyperplane at $\mathbf{x}$, that is: $H_{\mathbf{x}} = \{\mathbf{z} \in \mathbb{R}^n \ : \ \mathbf{a}^\intercal \mathbf{z} = \alpha\}$ for some $\mathbf{a} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, and $\mathbf{a}^\intercal \mathbf{y} \leq \mathbf{a}^\intercal \mathbf{x}$ for every point $\mathbf{y} \in C$. Since $C \cap H_{\mathbf{x}} \subseteq \mathbb{R}^{n-1}$ and by the inductive hypothesis, $C \cap H_{\mathbf{x}}$ has an extreme point, $\bar{\mathbf{x}}$.

We will now show that $\bar{\mathbf{x}}$ is also an extreme point in C. Note that although that it is obvious that $\bar{x} \in C$, it is not obvious that it is an extreme point in C, even though it is an extreme point in $C \cap H_{\mathbf{x}}$. Let $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\lambda \in (0, 1)$ s.t. $\bar{\mathbf{x}} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$. Then:

$$\mathbf{a}^\intercal \bar{\mathbf{x}} = \lambda \mathbf{a}^\intercal \mathbf{x}_1 + (1 - \lambda)\mathbf{a}^\intercal \mathbf{x}_2.$$

Since $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $H_{\mathbf{x}}$ is the supporting hyperplane, it is necessarily the case that $\mathbf{a}^\intercal \mathbf{x}_1 \leq \mathbf{a}^\intercal \bar{\mathbf{x}}$ and $\mathbf{a}^\intercal \mathbf{x}_2 \leq \mathbf{a}^\intercal \bar{\mathbf{x}}$. But since $\mathbf{c}^\intercal \bar{\mathbf{x}} = \lambda \mathbf{a}^\intercal \mathbf{x}_1 + (1 - \lambda)\mathbf{a}^\intercal \mathbf{x}_2$ (as otherwise $\mathbf{x}_1, \mathbf{x}_2 \in C \cap H_{\mathbf{x}}$ which contradicts $\bar{\mathbf{x}}$ being an extreme point in $C \cap H_{\mathbf{x}}$), this necessarily implies that $\mathbf{x}_1 = \mathbf{x}_2 = \bar{\mathbf{x}}$, or in other words that $\bar{x}$ is an extreme point in C. □

Theorem 2. *Let $P = \{\mathbf{x} \in \mathbb{R}^n \ : \ A\mathbf{x} \leq \mathbf{b}\}$, and consider the LP: $\max\{\mathbf{c}^\intercal \mathbf{x} \ : \ \mathbf{x} \in P\}$. If P has an extreme point, and the LP has an optimal solution, then the LP has an optimal solution which is an extreme point in P.*

Proof. Let $\alpha^\star$ be the value of the optimal solution and let O be the set of optimal solutions, i.e. $O = \{\mathbf{x} \in P \ : \ \mathbf{c}^\intercal \mathbf{x} = \alpha^\star\}$. Since P has an extreme point, it necessarily means that it does not contain a line. Since $O \subseteq P$ it doesn't contain a line either, hence, O contains an extreme point $\bar{\mathbf{x}}$. Similar to the previous proof, we will now show that $\bar{\mathbf{x}}$ is also an extreme point in P.

Let $\mathbf{x}_1, \mathbf{x}_2 \in P$ and $\lambda \in (0, 1)$ s.t. $\bar{\mathbf{x}} = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$. Then:

$$c^\intercal \bar{x} = \lambda c^\intercal \mathbf{x}_1 + (1 - \lambda) c^\intercal \mathbf{x}_2.$$

Since $\mathbf{x}_1, \mathbf{x}_2 \in P$ and $\alpha^\star$ is the optimal solution in P, it is necessarily the case that $\mathbf{c}^\intercal \mathbf{x}_1 \leq \mathbf{c}^\intercal \bar{\mathbf{x}} = \alpha^\star$ and $\mathbf{c}^\intercal \mathbf{x}_2 \leq \mathbf{a}^\intercal \bar{\mathbf{x}} = \alpha^\star$. But since $\mathbf{c}^\intercal \bar{\mathbf{x}} = \lambda \mathbf{a}^\intercal \mathbf{x}_1 + (1 - \lambda) a^\intercal \mathbf{x}_2$, this necessarily implies that $\mathbf{x}_1 = \mathbf{x}_2 = \bar{\mathbf{x}}$ (as otherwise $\mathbf{x}_1, \mathbf{x}_2 \in O$ contradicting $\bar{\mathbf{x}}$ being an extreme point in O), implying that $\bar{\mathbf{x}}$ is an extreme point in P. □

3 Basic Feasible Solutions

We will now give a different characterization of extreme points. We will first define basic feasible solutions, and see that the definition of extreme points and basic feasible solutions is equivalent.

Definition. *We say that a constraint $\mathbf{a}^\intercal \mathbf{x} \leq \mathbf{b}$ is **active** (or **binding**) at point $\bar{\mathbf{x}}$ if $\mathbf{a}\bar{\mathbf{x}} = \mathbf{b}$.*

Definition. *A solution in $P = \{\mathbf{x} \ : \ A\mathbf{x} \leq b\}$ is called* ***basic feasible*** *if it has n linearly independent active constraints.*

Definition. *A solution in $P = \{\mathbf{x} \ : \ A\mathbf{x} \leq b\}$ is called* ***degenerate*** *if it has more than n linearly independent active constraints.*

Example: Degeneracy does not imply redundancy. Consider the pyramid in $\mathbb{R}^3$. Any feasible solution in the pyramid only has 3 linearly independent active constraints, but we need at least 4 constraints to represent the pyramid.

3.1 Basic solutions in standard form

We say that an LP is in standard form if we express it as:

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

Let us assume that A is a $m \times n$ matrix. Any linear program can be written in the standard form with $m \leq n$. Without loss of generality we can assume that $rank(A) = m$ (if $rank(A) < m$, then the system has redundant constraints that can be identified and removed). Pick a set of indices $B \subseteq [n]$ that correspond to m linearly independent columns of the matrix A. Now, we can think of the matrix A as the concatenation of two matrices A_B and A_N where A_B is the $m \times m$ matrix whose columns are indexed by the indices in B, and A_N is the $m \times (n-m)$ matrix whose columns are indexed by the indices in $[n] \setminus B$. Similarly we can think of $\mathbf{x}$ as $[\mathbf{x}_B, \mathbf{x}_N]$ in a natural manner.

$$A = [A_B \mid A_N],$$
$$\mathbf{x} = [\mathbf{x}_B \mid \mathbf{x}_N].$$

Remark 3. For any basic feasible solution $\mathbf{x}$, we have a set $B \subseteq [n]$ of m indices that correspond to a linearly independent set of columns of A such that:

1. $\mathbf{x}_N = 0$
2. $\mathbf{x}_B = A_B^{-1}\mathbf{b}$.

In addition, for any set $B \subseteq [n]$ of m indices that correspond to a linearly independent set of columns, if $\mathbf{x}_B = A_B^{-1}\mathbf{b} \geq 0$ then $(\mathbf{x}_B, \mathbf{x}_N)$ is basic feasible.

We'll conclude this discussion with an important theorem which will be used in the simplex method.

Theorem 4. *Let $P = \{\mathbf{x} \in \mathbb{R}^n \ : A\mathbf{x} \leq \mathbf{b}\ \}$ then $\mathbf{x}$ is an extreme point of P if and only if $\mathbf{x}$ is a basic feasible solution of P.*

The proof follows the same principles as the proofs for extreme points and is left as an exercise in your next problem set.

4 The Simplex Algorithm

From the above discussion, it is clear that in order to find an optimal solution, it is sufficient to search over the basic feasible solutions to find the optimal one. The Simplex Algorithm, given by Dantzig, does this search in an organized fashion. There are many different kinds of implementations of Simplex. The different implementations vary by the way in which they search through the basic feasible solutions. The method for iterating through basic feasible solutions is called a *pivot rule*. Below we give an implementation of simplex.

Algorithm 1 Simplex

1: Let $(\mathbf{x}_B, \mathbf{x}_N)$ be a basic feasible solution.
2: $\bar{\mathbf{c}}^\intercal \leftarrow \mathbf{c}^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A$
3: $\mathbf{x}_B \leftarrow \bar{\mathbf{b}} := A_B^{-1}\mathbf{b}$
4: **if** $\bar{\mathbf{c}} \geq 0$ **then**
5: STOP and return $(\mathbf{x}_B, \mathbf{x}_N)$ as optimal solution.
6: **end if**
7: Select $j \in N$ such that $\bar{\mathbf{c}}_j < 0$
8: $\mathbf{d}_j \leftarrow A_B^{-1} A_j$
9: **if** $\mathbf{d}_j \leq 0$ **then**
10: STOP; return "LP is unbounded".
11: **end if**
12: $k \leftarrow \text{argmin}_{\{i \in B: d_{ji} > 0\}} (\bar{\mathbf{b}}_i / d_{ji})$
13: $B \leftarrow (B \setminus \{k\}) \cup \{j\}, N \leftarrow (N \setminus \{j\}) \cup \{k\}$
14: go to step 1.

Proposition 5. *The simplex method returns the optimal solution, if that solution exists. That is, if* $\bar{\mathbf{c}}^\intercal = \mathbf{c}^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A \geq 0$ *then the solution is optimal.*

Proof. Consider the dual of the problem:

$$\begin{aligned} &\max \mathbf{y}^\intercal \mathbf{b} \\ &\text{s.t. } \mathbf{y}^\intercal A \leq \mathbf{c}^\intercal \end{aligned} \tag{1}$$

Observe that $\mathbf{c}^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A \geq 0$ implies that $\mathbf{y}^\intercal = \mathbf{c}_B^\intercal A_B^{-1}$ is a feasible solution to the dual problem. The value of this solution in the dual objective is:

$$\mathbf{c}_B^\intercal A_B^{-1} \mathbf{b}.$$

Now observe that for a feasible solution $\mathbf{x}$ we have that $A\mathbf{x} = \mathbf{b}$, which in the language of basic feasible solutions, is:

$$A\mathbf{x} = A_B \mathbf{x}_B + A_N \mathbf{x}_N = A_B \mathbf{x}_B$$

since $\mathbf{x}_N = 0$. So a feasible solution respects $\mathbf{x}_B = A_B^{-1}\mathbf{b}$ and therefore the value in the primal objective for such a solution is:

$$\mathbf{c}_B^\intercal A_B^{-1} \mathbf{b}.$$

And therefore when $\mathbf{c}^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A \geq 0$ we have that the primal and dual objectives have the same value. From the strong duality theorem this implies that the solution must be optimal. □

4.1 A Running example of simplex

Consider the following linear program.

$$\begin{aligned} \min\ & -x_1 - 3x_2 \\ \text{s.t.}\ & 2x_1 + 3x_2 \leq 6 \\ & -x_1 + x_2 \leq 1 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Let us first write this in standard form:

$$\begin{aligned} \min\ & -x_1 - 3x_2 \\ \text{s.t.}\ & 2x_1 + 3x_2 + x_3 = 6 \\ & -x_1 + x_2 + x_4 = 1 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Say we start with $B = \{3, 4\}$ and $N = \{1, 2\}$. It should be clear that the resulting solution $(x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 1)$ is a basic feasible solution (Verify that the corresponding columns of A are linearly independent). Then,

$$A_B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_N = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix}.$$

Iteration 1:

$$\begin{aligned} \bar{\mathbf{c}}_N^\intercal &= \mathbf{c}_N^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A_N \\ &= (-1, -3) - (0, 0) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix} = (-1, -3) \end{aligned}$$

Next,

$$\mathbf{x}_B = \mathbf{b} = A_B^{-1}\mathbf{b}$$
$$\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

Say we select $j = 2$ to enter the basis:

$$\mathbf{d}_2 = A_B^{-1} A_2$$
$$\begin{bmatrix} d_{23} \\ d_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Thus,

$$k = \operatorname{argmin}\{b_3/d_{23}, b_4/d_{24}\} = \operatorname{argmin}\{6/3, 1/1\} = 4$$

leaves the basis. So, the next $B = \{3, 2\}$ and $N = \{1, 4\}$ and

$$A_B = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}, \quad A_N = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}.$$

Iteration 2:

$$\begin{aligned}\bar{\mathbf{c}}_N^\intercal &= \mathbf{c}_N^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A_N \\ &= (-1, 0) - (0, -3) \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix} = (-4, 3)\end{aligned}$$

Next,

$$\begin{aligned}\mathbf{x}_B = \mathbf{b} &= A_B^{-1}\mathbf{b} \\ \begin{bmatrix} x_3 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_3 \\ b_2 \end{bmatrix} &= \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}\end{aligned}$$

We select $j = 1$ to enter the basis:

$$\begin{aligned}\mathbf{d}_1 &= A_B^{-1} A_1 \\ \begin{bmatrix} d_{13} \\ d_{12} \end{bmatrix} &= \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}.\end{aligned}$$

Thus,

$$k = \operatorname{argmin}\{b_3/d_{13}\} = 3$$

leaves the basis. So, the next $B = \{1, 2\}$ and $N = \{3, 4\}$ and

$$A_B = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix}, \quad A_N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Iteration 3:

$$\begin{aligned}\bar{\mathbf{c}}_N^\intercal &= \mathbf{c}_N^\intercal - \mathbf{c}_B^\intercal A_B^{-1} A_N \\ &= (0, 0) - (-1, -3) \begin{bmatrix} 1/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = (4/5, 3/5)\end{aligned}$$

and hence the optimal solution is:

$$\begin{aligned}\mathbf{x}_B = \mathbf{b} &= A_B^{-1}\mathbf{b} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 1/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 8/5 \end{bmatrix}\end{aligned}$$

Remarks. What is the running time of the Simplex algorithm? Is it even finite? What if the algorithm repeatedly cycles over the same set B of basis indices? Observe that the algorithm is well-defined only after we specify the tie-breaking rules to be used for selecting the indices that leave and enter the basis (Steps 7 and 12). One way to break ties is by picking the least indices among the possible choices. This rule, due to Robert Bland, provably avoids cycling thereby ensuring that the algorithm is finite. It remains open to design variations of the tie-breaking rules so that the total number of iterations performed by the Simplex algorithm is polynomial in the number of variables.