AM 221: Advanced Optimization	Spring 2016
Prof. Yaron Singer	*Lecture 4 — February 3rd*

1 Overview

In the previous lecture we introduced linear optimization, and saw a few examples showing the modeling power of linear optimization. Our goals for understanding optimization problems and linear optimization in particular can be summarized through the following questions:

1. When is an LP feasible / infeasible?
2. When is an LP bounded / unbounded?
3. What is a characterization of optimality?
4. How do we design algorithms that find optimal solutions to (bounded and feasible) LPs (and optimization problems in general)?

In this lecture we'll introduce duality theory which is a beautiful and fundamental building block in optimization. To do so, we will first prove Farkas lemma which tells us that if a linear program is not feasible there exists a certificate showing this. As we discuss duality we will see that Farkas lemma can also be used to tell us when an LP is bounded or unbounded. Duality is in fact a characterization of optimality and we will use it to develop algorithms for finding optimal solutions of linear programs. Let's start.

2 Farkas Lemma: Certificate of Feasibility

As a first step, we would like to understand when an LP is feasible. That is, when is the set of constraints non-empty. The answer to this question was answered by Gauss somewhere around 1800 for the case in which the constraints of the LP can be represented by equalities (i.e. $Ax = b$). Note that an equivalent question to whether a region is feasible or not, is whether a system of linear equations has a solution. That is, we can encode the constrains as a matrix A and vector b and solve $Ax = b$ by *Gaussian elimination.*

Example. Consider the following linear program:

$$\begin{aligned} \max \quad & x_1 + x_2 + x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 6 \\ & 2x_1 + 3x_2 + x_3 = 8 \\ & 2x_1 + x_2 + 3x_3 = 0 \end{aligned}$$

Does the above LP have a feasible region? Multiplying the first, second, and third row by 4, -1, -1 respectively, and then adding the equations we get that $0x_1 + 0x_2 + 0x_3 = 16$. Therefore we conclude that the feasible region is empty, and the LP is infeasible. The *certificate of infeasibility* is $(4, -1, -1)$. For a program with a feasible region, a *certificate of feasibility* on the other hand, is any point in the feasible region. That is, a solution to the system of equations.

We can write the set of constraints as the matrix A and vector b below:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 6 \\ 8 \\ 0 \end{pmatrix} \tag{1}$$

Stated in these terms, we are interested in knowing whether there exists an $\mathbf{x} \in \mathbb{R}^3$ s.t. $A\mathbf{x} = \mathbf{b}$. By performing raw operations on the matrix (Gaussian elimination), we saw that there exists a vector $\mathbf{y} = (4, -1, -1)$ s.t. $\mathbf{y}^\intercal A\mathbf{x} = \mathbf{0}$ and $\mathbf{y}^\intercal \mathbf{b} = 16 \neq 0$. The vector $\mathbf{y}$ is a *certificate of infeasibility*. We know this procedure as the fundamental theorem in linear algebra.

Theorem. *[Gauss] Let A be a $m \times n$ matrix, and let $\mathbf{b} \in \mathbb{R}^m$ vector. Then exactly one of the following statements hold, but not both:*

(I) $\exists \mathbf{x} \in \mathbb{R}^n \ : A\mathbf{x} = \mathbf{b}$ ***or***

(II) $\exists \mathbf{y} \in \mathbb{R}^m$ *s.t.* $\mathbf{y}^\intercal A = \mathbf{0}$ *and* $\mathbf{y}^\intercal \mathbf{b} \neq 0$.

So whenever the constraints can be encoded as $A\mathbf{x} = b$, we can use the Gaussian elimination process to tell us whether an LP is feasible; if it is not feasible we are able to easily produce the certificate $\mathbf{y}$ which tells us that no solutions that satisfy the constraints exist. In general, we would like to be able to perform such tests for the inequality $A\mathbf{x} \leq \mathbf{b}$. This is Farkas Lemma[1].

Theorem 1 (Farkas)**.** *Let A be a $m \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^m$. Then exactly one of the following statements hold, but not both:*

(I) $\exists \mathbf{x} \in \mathbb{R}^n \ : A\mathbf{x} = \mathbf{b}, \ \ \mathbf{x} \geq \mathbf{0}$ ***or***

(II) $\exists \mathbf{y} \in \mathbb{R}^m$ *s.t.* $\mathbf{y}^\intercal A \leq \mathbf{0}$ *and* $\mathbf{y}^\intercal \mathbf{b} > 0$.

Proof. We will first show that if (I) is true then (II) is necessarily false. Assume $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{x} \geq \mathbf{0}$. If $\mathbf{y}^\intercal A \leq \mathbf{0}$ then for $\mathbf{x} \geq \mathbf{0}$ we have that $\mathbf{y}^\intercal A\mathbf{x} \leq \mathbf{0}$. Since $A\mathbf{x} = \mathbf{b}$ this implies that $\mathbf{y}^\intercal \mathbf{b} \leq \mathbf{0}$, and thus it cannot be that both $\mathbf{y}^\intercal A\mathbf{x} \leq 0$ and $\mathbf{y}^\intercal \mathbf{b} > 0$.

Now we'll prove that if (I) is false then (II) is necessarily true. Define:

$$C = \{\mathbf{q} \in \mathbb{R}^m \ : \ \exists \mathbf{x} \geq \mathbf{0} \ A\mathbf{x} = \mathbf{q}\};$$

Notice that C is a convex set: for $\mathbf{q}_1, \mathbf{q}_2 \in C$ there exist $\mathbf{x}_1, \mathbf{x}_2$ s.t $\mathbf{q}_1 = A\mathbf{x}_1$ and $\mathbf{q}_2 = A\mathbf{q}_2$, and for any $\lambda \in [0, 1]$ we have that $\lambda\mathbf{q}_2 + (1-\lambda)\mathbf{q}_2 = \lambda A\mathbf{q}_1 + (1-\lambda)A\mathbf{q}_2 = A(\lambda\mathbf{q}_1 + (1-\lambda)\mathbf{q}_2)$ and hence $\lambda\mathbf{q}_2 + (1-\lambda)\mathbf{q}_2 \in C$. Since (I) is false $\mathbf{b} \notin C$. From the separating hyperplane theorem, we know there exists $\mathbf{y} \in \mathbb{R}^m \setminus \mathbf{0}$ s.t. $\mathbf{y}^\intercal \mathbf{q} \leq \mathbf{0}$ and $\mathbf{y}^\intercal \mathbf{b} > 0$, for all $\mathbf{q} \in C$. Since $\mathbf{q} = A\mathbf{x}$ that implies that $\forall \mathbf{x} \geq \mathbf{0}$ we have that $\mathbf{y}^\intercal A\mathbf{x} \leq \mathbf{0}$ and $\mathbf{y}^\intercal \mathbf{b} > 0$, as required. □

[1] The result is known as Farkas Lemma, but we write it here as a theorem.

As we will soon see, we also have a constructive manner to find this vector $\mathbf{y}$ which serves as an infeasibility certificate in cases where the set is indeed infeasible.

3 Duality Theory

Consider the following optimization problem:

$$\begin{aligned} \min\ & x_1 + 2x_2 + 4x_3 \\ s.t.\ & x_1 + x_2 + 2x_3 = 5 \\ & 2x_1 + x_2 + 3x_3 = 8 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Our main goal is to understand how to solve linear optimization problems like the one above. That is, we seek to find an optimal solution $\mathbf{x} \in \mathbb{R}^n$ for which the above program in minimal and respects the constraints. To do so we can try to find upper and lower bound on the optimal solution:

$$\text{lower bound} \leq \text{optimal solution} \leq \text{upper bound}$$

The primal-dual method which we now introduce seeks to find the smallest upper bound and the largest lower bound and in doing so produce an optimal solution.

Finding upper bounds on optimal solution. For the optimization problem, we can try to guess solutions, and evaluate their quality. For example, the solution $(2,1,1)$, i.e. $x_1 = 2, x_2 = 1, x_3 = 1$ respects the constraints of the program above, and its value is $1 \cdot 2 + 2 \cdot 1 + 4 \cdot 1 = 8$. So, $\alpha_1 = 8$ is an *upper bound* on the optimal solution – since $\mathbf{x}_1 = (2,1,1)$ is feasible, and its value is 8, we know that the optimal solution is at most 8. Similarly, we can try $(3,2,0)$, see that it also respects the constraints, and the value of this solution would be $\alpha_2 = 1 \cdot 3 + 2 \cdot 2 + 4 \cdot 0 = 7$. Therefore, $\alpha_2 = 7$ is also a lower bound on the optimal solution.

At this point it is not clear how to "guess" candidates for an upper bound, but assuming someone gives us a point, we can easily check whether that point is feasible, and if it is, we know its value is an upper bound on the optimal solution. So what about lower bounds?

Finding lower bounds on optimal solution. Let's consider the performing the following operation on our constraint matrix $\mathbf{A}$ and $\mathbf{b}$: multiply the first row by 2 and subtract the second row from the first row. Or in other words, $(2,-1)^\intercal A$ and $(2,-1)^\intercal \mathbf{b}$. This operation gives us the following equation:

$$0 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 = 2$$

If we look closely at the above equation we see that all multipliers of x_1, x_2, x_3 are *not greater* than the multipliers in the objective. Since all variables in the solution need to be non-negative we have: $0 \cdot x_1 < 1 \cdot x_1, 1 \cdot x_2 \leq 1 \cdot x_2, 1 \cdot x_3 \leq 4x_3$. We therefore have that $\alpha^3 = 2$ is a *lower bound* on the optimal solution, i.e. the optimal solution must be at least 2.

So ideally we want the lower bounds to be as large as possible. If we check the vector $(3,-1)$ we will get a lower bound of $\alpha_4 = 7$. But this lower bound matches our upper bound. Since the value of the optimal solution must be no greater than the value of the upper bound and no smaller than the value of the lower bound, it therefore must be that $\alpha^\star = 7$ is the value of optimal solution.

4 The Primal Dual Theory

Let's formulate the above discussion. For a given LP of the form:

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

For any $\mathbf{y} \in \mathbb{R}^m$ s.t. $\mathbf{y}^\intercal A_j \leq \mathbf{c}_j\ \forall j = 1, \ldots, n$ it must be the case that $\mathbf{y}^\intercal \mathbf{b} \leq \alpha = \mathbf{c}^\intercal \mathbf{x}$, for any feasible solution $\mathbf{x}$. We can find the best lower bound by solving:

$$\begin{aligned} &\max \mathbf{y}^\intercal \mathbf{b} &&(\textsc{Dual})\\ &s.t.\ \mathbf{y}^\intercal A \leq \mathbf{c}^\intercal \end{aligned}$$

Notation: Throughout the rest of this section we will use Primal to denote the primal problem above with $A, \mathbf{c}, \mathbf{b}$ and Dual to denote the dual problem. We will also use $\alpha^\star, \beta^\star$ to denote the optimal solutions of the primal and dual problems, respectively.

4.1 Weak duality theorem

The following theorem is a formulation of a rather straightforward idea that we brought up in our discussion.

Theorem 2 (weak duality)**.** *Let* $\mathbf{x}$ *be a feasible solution to* Primal *and* $\mathbf{y}$ *a feasible solution to* Dual. *Then* $\mathbf{c}^\intercal \mathbf{x} \geq \mathbf{y}^\intercal \mathbf{b}$.

Proof. Since $\mathbf{x}$ is feasible, $A\mathbf{x} = \mathbf{b}$, and since $\mathbf{y}$ is feasible $\mathbf{y}^\intercal A \leq \mathbf{c}^\intercal$. Thus $\mathbf{y}^\intercal \mathbf{b} = \mathbf{y}^\intercal A\mathbf{x} \leq \mathbf{c}^\intercal \mathbf{x}$. □

Remark 3. *Let* $\alpha^\star$, $\beta^\star$, *denote the values of the optimal primal and dual solutions, respectively.*

1. *If the primal is* ***unbounded*** *($\alpha^\star = -\infty$), then the dual is* ***infeasible*** *(notation: $\beta^\star = -\infty$).*
2. *If the dual is* ***unbounded*** *($\beta^\star = \infty$) then it must be that the primal is* ***infeasible*** *($\alpha^\star = \infty$).*
3. *The dual of the dual is the primal.*

4.2 Strong duality theorem

Theorem 4. *If either* Primal *or* Dual *are feasible, then* $\alpha^\star = \beta^\star$.

Proof. W.l.o.g suppose that Primal is feasible (the dual of the Dual is the primal). If Primal is unbounded, then $\alpha^\star = -\infty = \beta^\star$ and we are done. Otherwise, let $\mathbf{x}^\star$ be the optimal solution for Primal, i.e. $\alpha^\star = \mathbf{c}^\intercal \mathbf{x}^\star$. We will show that there exists a $\mathbf{y} \in \mathbb{R}^m$ s.t.:

$$\mathbf{y}^\intercal A \leq \mathbf{c}^\intercal \wedge \mathbf{y}^\intercal \mathbf{b} \geq \alpha^\star \tag{2}$$

If we show this, then this implies that the optimal solution to Dual is at least $\alpha^\star$. But since the weak duality theorem says that solutions to Dual are bounded from above by solutions to Primal it must mean that $\beta^\star = \alpha^\star$.

To show there exists a $\mathbf{y} \in \mathbb{R}^m$ as stated in (2) we will use Farkas Lemma. Notice that to show that such a $\mathbf{y}$ exists we can rule out the existence of the existence of a vector $\mathbf{x} \in \mathbb{R}^n$ that respects:

$$A\mathbf{x} = \mathbf{b}$$
$$\mathbf{c}^\intercal \mathbf{x} < \alpha^\star$$

Notice however that finding such a vector $\mathbf{x}$ contradicts the minimality of $\alpha^\star$. Thus, by Farkas Lemma, such a vector $\mathbf{y}$ exists, and by weak duality $\beta^\star = \alpha^\star$ as we stated above. □