CSC2411 - Linear Programming and Combinatorial Optimization*
Lecture 2: Different forms of LP. The algebraic objects behind LP. Basic Feasible Solutions

Notes taken by Graham Taylor

January 22, 2005

Summary: We first describe different forms of linear programming, including the standard and canonical forms. The concept of basic feasible solutions is introduced, and we discuss the basic algebraic objects behind LP which will lead to the Simplex method for solving LP.

Overview

In the previous lecture, we introduced the notion of optimization problems. Figure 1 shows several families and examples of optimization problems. In this course, we will focus on the relationship between Linear Programming (a family of continuous optimization problems) and certain finite domain problems. Specifically, we will examine methods of approximating solutions to the latter problems through tools developed for the former.

Forms of Linear Programming

Recall, from the previous lecture, the linear programming problem

$$\begin{aligned}
&\min \sum_{j=1}^{n} c_j x_j \quad \text{subject to} \\
&\sum_{j=1}^{n} a_{ij} x_j \geq b_i \quad \text{for } i = 1 \ldots m \\
&x_j \geq 0.
\end{aligned}$$

* Lecture Notes for a course given by Avner Magen, Dept. of Computer Sciecne, University of Toronto.

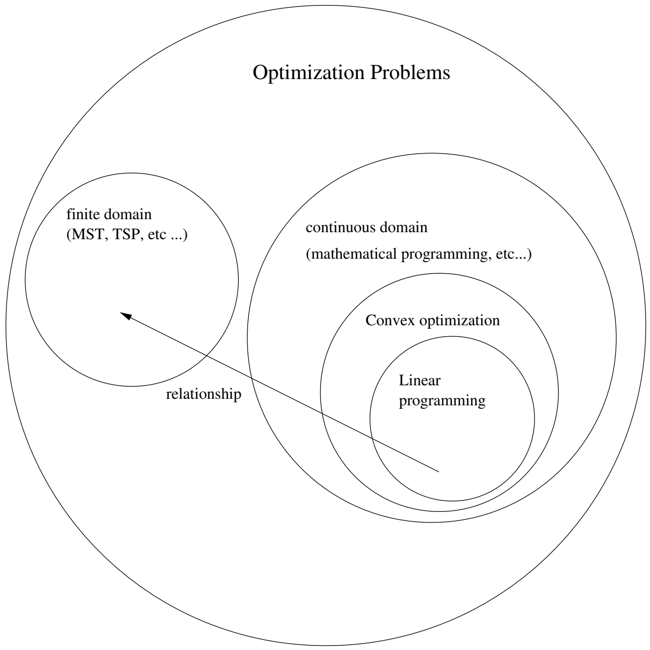

Figure 1: Types of optimization problems

Another way to write this is

$$\begin{aligned}
&\min\langle c, x\rangle \quad \text{subject to} \\
&\langle a_i, x\rangle \geq b_i \quad \text{for } i = 1 \ldots m \\
&a_i \in \mathbb{R}^n \\
&x \geq 0,
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ specifies the dot product. An even more compact form is

$$\begin{aligned}
&\min\langle c, x\rangle \quad \text{subject to} \\
&Ax \geq b \\
&x \geq 0.
\end{aligned}$$

Definition 2.1. An LP is said to be in *standard* form if it is written as

$$\begin{aligned}
&\min\langle c, x\rangle \quad \text{subject to} \\
&Ax = b \\
&x \geq 0.
\end{aligned}$$

An LP is said to be in *canonical* form if it is written as

$$\begin{aligned} &\min\langle c, x\rangle \quad \text{subject to} \\ &Ax \geq b \\ &x \geq 0. \end{aligned}$$

There are many other conventions, but these two will be the ones of interest for this course. The most general form will contain some inequalities, some equalities, some non-negative variables and some unconstrained variables.

$$\begin{aligned} &\langle a_i, x\rangle = b_i, \quad i \in \mathrm{E}, \\ &\langle a_i, x\rangle \geq b_i, \quad i \in \mathrm{I}^+, \\ &\langle a_i, x\rangle \leq b_i, \quad i \in \mathrm{I}^-, \\ &x_j \lesseqgtr 0, \quad j \in U, \text{(uncostrained)} \\ &x_j \geq 0, \quad j \in N. \end{aligned}$$

It is useful to know how to move from an LP in the general form (as above) to standard form. First we need to eliminate inequality constraints. Given an inequality constraint $\langle a_i, x\rangle \leq b_i$, we introduce the *slack* variable y_i and write

$$\langle a_i, x\rangle + y_i = b_i, \quad y_i \geq 0.$$

Since $\langle a_i, x\rangle \geq b_i$ is equivalent to $\langle -a_i, x\rangle \leq -b_i$, this also covers the other type of inequality. To attain standard form, we also must eliminate unconstrained variables of the form

$$x_j \lesseqgtr 0.$$

Notice that any real number can be presented as a difference of two nonnegative numbers, hence we may replace x_j by $x_j^+ - x_j^-$, when $x_j^+, x_j^- \geq 0$. We replace every occurrence of x_j with $x_j^+ - x_j^-$.

Example 2.2. Consider the LP

$$\begin{aligned} &\max x_1 + 3x_2 \quad \text{subject to} \\ &2x_1 - x_2 \geq 10 \\ &x_1 \lesseqgtr 0 \quad x_2 \geq 0. \end{aligned}$$

Convert to standard form.

First we attempt to convert the inequality constraints to equality constraints by introducing the surplus variable, y_1.

$$\begin{aligned} &\max x_1 + 3x_2 \quad \text{subject to} \\ &2x_1 - x_2 - y_1 = 10 \\ &x_1 \lesseqgtr 0 \quad x_2 \geq 0 \quad y_1 \geq 0. \end{aligned}$$

Next we replace the unconstrained variable x_1 by x_1^+ and x_1^-.

$$\begin{aligned}&\max x_1^+ - x_1^- + 3x_2 \quad \text{subject to}\\&2x_1^+ - 2x_1^- - x_2 - y_1 = 10\\&x_1^+, x_1^-, x_2, y_1 \geq 0.\end{aligned}$$

Finally, we convert the maximization problem to a minimization problem as follows

$$\begin{aligned}&\min -x_1^+ + x_1^- - 3x_2 \quad \text{subject to}\\&2x_1^+ - 2x_1^- - x_2 - y_1 = 10\\&x_1^+, x_1^-, x_2, y_1 \geq 0.\end{aligned}$$

since $\max\langle c, x\rangle = -\min\langle -c, x\rangle$.

Basic Feasible Solutions

Let us now consider the linear system of equations $Ax = b$ where A has m rows and n columns. We next show that we may assume that matrix A has full row rank. In particular, $m \leq n$.

The rank of a matrix is the dimension of the linear space spanned by its rows, and also the dimension of the linear space spanned by its columns. We may also say that $m = \text{rank}(A) \leq n$.

Example 2.3. Consider the system of equations

$$\begin{aligned}&x_1 + x_2 = 5\\&2x_2 + x_3 = 8\\&3x_1 + 5x_2 + x_3 = ?.\end{aligned}$$

The missing value can either have value $= 23$ or $\neq 23$. In the former case, the third equation is redundant. In the latter case, the system is inconsistent.

We will now introduce the notation A^i to mean the ith row of A and similarly, A_j to mean the jth column of A.

We now formally prove the assumption about A. Suppose there is a row A^i that is linearly dependent on the rest of the rows.

$A^i = \sum_{j\neq i} \lambda_j A^j$, then for any solution x, we have

$$\begin{aligned}b_i = \langle A^i, x\rangle &= \langle \sum_{j\neq i} \lambda_j A^j, x\rangle\\&= \sum \lambda_j \langle A^j, x\rangle\\&= \sum \lambda_j b_j.\end{aligned}$$

So if $b_i = \sum \lambda_j b_j$ then the i'th equation is redundant. Otherwise, no x satisfies the system. These two cases are easily detected by Gaussian elimination. In the former case, this row can be removed. In the latter case, we will just state that the system is infeasible and stop.

Equipped with this assumption let's start with a fairly trivial situation in which $m = n$. Notice that in this event A is a nonsingular square matrix, and so $Ax = b$ has a unique solution, $x = A^{-1}b$. If $x \geq 0$ then that is the only solution to the system, and otherwise, the system is infeasible.

We use this simple observation to characterize a certain (finite) set of feasible solutions that will play a vital role in the following discussion. Before we go on we require the following.

Definition 2.4. Let $P = \{x | x \geq 0, Ax = b\}$ be called the set of feasible solutions.

A solution to the system satisfies $\sum_j x_j A_j = b$. Let's pick a set $I \subset \{1, \ldots, n\}$ such that $A_j, j \in I$ is a set of linearly independent columns. There is at least one such set, since rank$(A) = m$. We associate with I a vector x, for which for every $j \notin I$, $x_j = 0$. By the linear independence of the columns $A_j, j \in I$, we know that there is exactly one solution to the rest of the veraibales that will satisfy $Ax = b$. Formally, if we let A_I be the matrix restircted to the columnes in I, and x_I to be the vector x restriced to the indices in I, then we define the vector x so that $x_I = A_I^{-1}b$ and $x_{\overline{I}} = 0$ (notice that by definition, A_I is a square nonsingular matrix).

Definition 2.5. A vector x defined as above is called a *basic solution* associated with I. If $x \geq 0$, then we get that $x \in P$ and we call it *Basic Feasible Solution*.

We now consider an example of a procedure that generates basic feasible solutions.

Example 2.6. Consider the LP

$$\begin{aligned}
&\min\langle x, (-1, 1, 0)^T\rangle \quad \text{subject to} \\
&Ax = b \quad \text{where} \\
&A = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 7 \end{pmatrix}.
\end{aligned}$$

First, we choose $I_1 = \{1, 3\}$ so

$$A_I = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

We have

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \Rightarrow x = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}.$$

Next, we choose $I_2 = \{2, 3\}$ so

$$A_I = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}.$$

We have

$$\begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \end{pmatrix} \Rightarrow \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \Rightarrow x = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}.$$

In the above example, we were "fortunate" to get $y \geq 0$. Otherwise the solutions would not have been feasible.

To motivate the concept of bfs, we state Claim 2.7. It will be proven later in our discussion.

Claim 2.7. *If an LP in standard form has an optimal solution, it has one which is a basic feasible solution.*

What does Claim 2.7 tells us? Instead of the initial infinite domain, we may restrict our attention to a special finite set of bfs. Consequently, there is a brute force algorithm that finds an optimal solution to an LP, provided an optimal solution exists.

Algorithm 2.8 (Brute force).

Input: A, c, b

Initialize: Cost, $Z = \infty$
For all subsets $I \subset \{1, \ldots, n\}$ of size m, do
    Check if A_I is nonsingular
    If it is, check if $A_I^{-1} b \geq 0$
        If so, let x be the corresponding bfs, do
            If $\langle x, c \rangle < Z$
              Set $Z = \langle x, c \rangle$
              $x_{\text{best}} = x$

Output: x_{best} corresponding to the current best solution.

When considering the running time, note that we have $\binom{n}{m}$ iterations, which for $m = n/2$, say, is exponential in the size of n. This is indeed a good estimation to the running time, since the involved calculation does not contain extremely large numbers as is suggested by the following claim, which was given in Assignment 1, question 3a.

Claim 2.9. *The size of representation of a bfs is polynomial in the size of the input of the LP.*

Definition 2.10. We call $x \in P$ extreme if it is not the average of two points, $y, z \in P; y, z \neq x$. Specifically, if x is the average of two points y, z, we mean any convex combination $x = \lambda y + (1 - \lambda) z$ where $x \neq y \neq z$.

For example, the extreme points in $[0, 1]$ are $\{0, 1\}$. Also, $\vec{0}$ is the extreme point of $\{x \geq 0\}$.

Example 2.11. Consider the set of real values, $\{\|x\|_2 \leq 1\}$. The extreme points are $\{||x|| = 1\}$.

Example 2.12. Consider a polygon in 2-D (Figure 2). Its extreme points are its vertices.

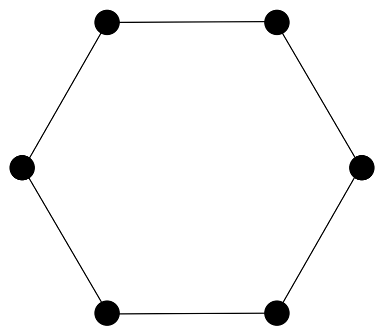

Figure 2: Extreme points example: 2-d polygon

Lemma 2.13. *x is a bfs iff x is extreme.*

Proof. Part 1. bfs $\Rightarrow$ extreme
Assume x is a bfs and is the average of two points, y and z

$$x = \lambda y + (1 - \lambda)z, \quad 0 < \lambda < 1$$
$$y, z \in P.$$

Equivalently, $\forall j, x_j = \lambda y_j + (1 - \lambda)z_j$. For $j \notin I$ we have

$$0 = x_j = \lambda y_j + (1 - \lambda)z_j$$

but $y_j, z_j \geq 0$. Therefore $y_j = z_j = 0$, for $j \notin I$. So

$$y_j = z_j = B^{-1}b = x_i, \text{for } j \in I$$

and we have $x = y = z$. It follows that x is extreme.
Part 2. extreme $\Rightarrow$ bfs

We first claim that x is a bfs iff $J = \{j | x_j > 0\}$ corresponds to a set of linearly independent columns. If this set is not independent it is immediate that x is not a bfs and if it is independent then simple linear algebra shows that there is a way to extetend the set of columns corresponding to J to a set of linearly independent columns, corresponding to a set I, $|I| = m, I \supset J$. Clearly, x is a bfs with a corresponding basis I.

Assume x is not a bfs. Let $J = \{j | x_j > 0\}$. We know that $x \in P$ such that the columns A_i for $x_i > 0$ are linearly independent iff x is a bfs. Therefore, the columns in A_J are linearly dependent. and so there is a nonzero vector v that is 0 outside J so that $Av = 0$. We have

$$A(x \pm \lambda v) = Ax \pm \lambda Av = Ax = b.$$

In other words, $x \pm \lambda v$ is a solution to the system $Ax = b$. For small enough $\lambda > 0$, both $x + \lambda v, x - \lambda v \geq 0$. Hence $x \pm \lambda v \in P$. Since $x = \frac{1}{2}\underbrace{(x + \lambda v)}_{=y} + \frac{1}{2}\underbrace{(x - \lambda v)}_{=z}$, x is not extreme.

□

Claim 2.14. *If x is a bfs, then there is a choice of vector c such that $\langle x, c\rangle < \langle y, c\rangle, y \in P$.*

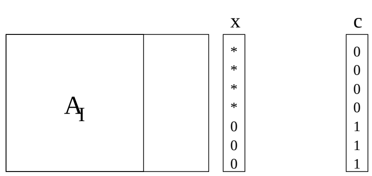

Figure 3: Choosing the objective function, c

Proof. Since x is a bfs, $x_j = 0$, for $j \notin I$. We choose c such that $c_j = 0$, for $j \in I$, and $c_j = 1$, for $j \notin I$. Clearly $\langle x, c\rangle = 0$. If $y \neq x$, then $y_{\overline{I}} \neq 0$ (otherwise $y_I = x_I$ and so $y = x$) but since $y_{\overline{I}} \geq 0$ we get $\langle y, c\rangle = \sum_{j \neq I} y_j > 0$ and so $\langle y, c\rangle > 0 = \langle x, c\rangle$.

Notice that the reverse direction is obvious, since if x is not a bfs, it is the average of y, z and so cannot achieve strictly smaller value than both. □

We now set out to prove Claim 2.7.

Proof. Assume that we have an optimal solution, x^*. If x^* is a bfs, then we are done. Otherwise, we may find a bfs through the following iterative procedure.

- Start with the initial solution, call it x^0
- Let $J = \left\{j | x_j^0 > 0\right\}$. If $A_j, j \in J$ are linearly independent, stop: x is a bfs. Otherwise, choose v (as in Lemma 2.13 such that $Av = 0$ and $v_j = 0$, if $x_j = 0$.
- Let k be the index for which
$$\left|\frac{x_k}{v_k}\right| = \min_{j: v_j \neq 0} \left|\frac{x_j}{v_j}\right|,$$
and set $\lambda = -\frac{x_k}{v_k}$

- Set $x^1 = x^0 + \lambda v$. We can easily verify that

$$\begin{aligned} &x^1 \geq 0 \quad \text{(hence } x \in P\text{)}, \\ &x^1_k = 0, \\ &x^1_j = 0, \quad \text{for } j \notin J. \end{aligned}$$

  Repeat.

This process ends when J is a set of indices for independent columns of A (notice that we may end up with $J = \phi$, in which case we know that $x = 0 \in P$, and this is a bfs).

We are left with showing that the bfs is as good as the optimal solution we started with. Indeed, at every iteration in the above procedure, the objective function changes by $\langle c, \lambda v\rangle$ which, by the following claim, is 0.

□

Claim 2.15. $\langle c, v\rangle = 0$

Proof. Assume x is the optimal solution. For $\lambda > 0$ small enough, $x \pm \lambda v \in P$,

$$\langle c, x \pm \lambda v\rangle = \langle c, x\rangle \pm \lambda\langle c, v\rangle$$

Since $\lambda \neq 0$, $\langle c, v\rangle \neq 0$ would change the value of the optimal solution at x. □

We now turn to an example to illustrate the process of finding a bfs.

Example 2.16. Given system

$Ax = b$ where

$$A = \begin{pmatrix} 1 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 3 & 5 \\ 0 & 1 & 1 & 1 & 8 \end{pmatrix}, b = \begin{pmatrix} 9 \\ 18 \\ 9 \end{pmatrix}, \text{and initial solution } x^0 = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 2 \\ 0 \end{pmatrix},$$

find a bfs.

Step 1. $J = \{1, 2, 3, 4\}$ (only the last element of x^0 is zero). A_J is not linearly independent, so we choose $v = (1, 1, -1, 0, 0)^T$ such that $Av = 0$ and $v_5 = 0$. Next we choose λ by comparing the ratios $|x_j/v_j|$ for $v_j \neq 0$

$$\left|\frac{x_1}{v_1}\right| = \left|\frac{1}{1}\right| \quad \left|\frac{x_2}{v_2}\right| = \left|\frac{3}{1}\right| \quad \left|\frac{x_3}{v_3}\right| = \left|\frac{4}{-1}\right|,$$

and a minimum is found for $j = 1$, so we set $\lambda = -1$.

$$x^1 = x^0 + \lambda v = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 2 \\ 0 \end{pmatrix} + (-1) \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 5 \\ 2 \\ 0 \end{pmatrix}.$$

Step 2. $J = \{2, 3, 4\}$ (the first and last elements of x^1 are zero). A_J is not linearly independent, so we choose $v = (0, -1, 2, -1, 0)^T$ such that $Av = 0$ and $v_1, v_5 = 0$. Next we choose λ by comparing the ratios $|x_j/v_j|$ for $v_j \neq 0$

$$\left|\frac{x_2}{v_2}\right| = \left|\frac{2}{-1}\right| \quad \left|\frac{x_3}{v_3}\right| = \left|\frac{5}{2}\right| \quad \left|\frac{x_4}{v_4}\right| = \left|\frac{2}{-1}\right|.$$

We see that there is a tie for the minimum at $j = 2$,$j = 4$ and so $\lambda = 2$.

$$x^2 = x^1 + \lambda v = \begin{pmatrix} 0 \\ 2 \\ 5 \\ 2 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ -1 \\ 2 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 9 \\ 0 \\ 0 \end{pmatrix}.$$

We are left with only one column of A, corresponding to the $x_j \neq 0$ at $j = 3$. This single vector is obviously linearly independent, so our solution, $x^2 = (0, 0, 9, 0, 0)^T$ is a bfs.