Theory of Linear Programming

Debasis Mishra*

April 6, 2011

1 Introduction

Optimization of a function f over a set S involves finding the maximum (minimum) value of f (objective function) in the set S (feasible set). Properties of f and S define various types of optimization. Primarily, optimization can be classified into three categories.

1. **Linear Programming**: If f is a linear function (e.g., $f(x_1, x_2) = x_1 + 2x_2$) and the set S is defined by a finite collection of linear inequalities and equalities, then it is called a **linear program**. As an example, consider the following linear program.

$$\max_{x_1,x_2} f(x_1, x_2) = \max_{x_1,x_2} [x_1 + 2x_2]$$
$$\text{s.t.}$$
$$x_1 + x_2 \leq 6$$
$$x_1 \geq 2$$
$$x_2 \geq 0$$

2. **Integer Programming**: An integer program is a linear program with further restriction that the solution be integers. In the previous example, if we impose that x_1 and x_2 can only admit integer values, then it becomes an integer program.

3. **Nonlinear Programming**: If f is a non-linear function and the feasible set S is defined by a finite collection of non-linear equations, then it is called a **non-linear program**. There are further classifications (and extensions) of non-linear programs

*Planning Unit, Indian Statistical Institute, 7 Shahid Jit Singh Marg, New Delhi 110016, India, E-mail: dmishra@isid.ac.in

depending on the specific nature of the problem. Typically, f is assumed to be continuous and differentiable. An example is the following non-linear program.

$$\max_{x_1,x_2} f(x_1, x_2) = \max_{x_1,x_2}[-2x_1^2 - 3x_2^2]$$
$$\text{s.t.}$$
$$x_1 + x_2 = 1.$$

In general, an optimization problem written in mathematical form is referred to as a **mathematical program**. In this course, we will learn about **linear and integer programming problems, their solution methods**.

To understand a little more about linear and integer programs, consider the above example. The feasible set/region can be drawn on a plane. Figure 1 shows the feasible regions for the linear program (dashed region), and the integer points inside that feasible region is the feasible region of the integer program. Notice that the optimal solution of this linear program has to lie on the boundary of the feasible region. Moreover, an *extreme* point is an optimal solution ($x_1 = 2,\ x_2 = 4$). This is no accident, as we will show. If we impose the integer constraints on x_1 and x_2, then the feasible region has a finite set of points. Again, the optimal solution is $x_1 = 2,\ x_2 = 4$ (this is obvious since this is an integral solution, and is an optimal solution of the linear program).

2 Steps in Solving an Optimization Problem

There are some logical steps to solve an optimization problem.

- **Modeling**: It involves reading the problem carefully to decipher the variables of the problem. Then, one needs to write down the objective and the constraints of the problem in terms of the variables. This defines the objective function and the feasible set, and hence the mathematical program. This process of writing down the mathematical program from a verbal description of the problem is called **modeling**. Modeling, though it does not give the solution, is an important aspect of optimization. A good modeling helps in getting solutions faster.

- **Solving**: Once the problem is modeled, the solution is sought. There are algorithms (techniques) to solve mathematical programs. Commercial software companies have come up with *solvers* that have built packages using these algorithms.

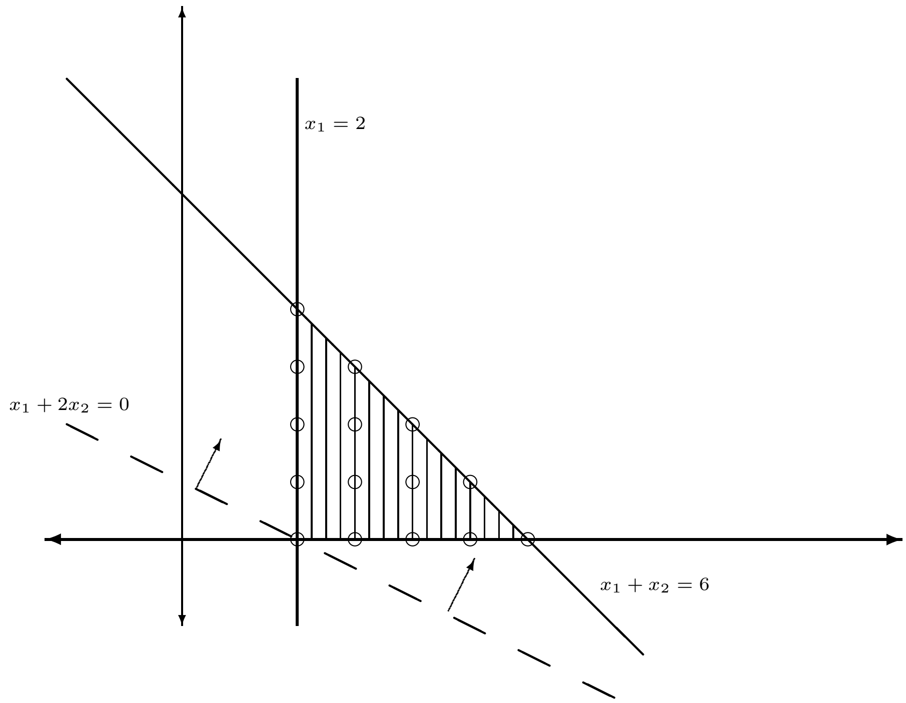

Figure 1: Feasible set and objective function of linear and integer programs

3 Linear Programming

3.1 An Example

Problem 1 *There is 100 units of water to be distributed among three villages. The water requirement of the villages are 30, 50, and 40 respectively. The water shortage costs of the three villages are 4, 3, and 5 respectively. Water supply to no two villages should exceed 70. Find a water distribution that minimizes the total water shortage cost.*

Modeling: Let x_i ($i \in \{1, 2, 3\}$) denote the amount of water supplied to village i. Since the total amount of water is 100, we immediately have

$$x_1 + x_2 + x_3 = 100.$$

Further, water supply of no two villages should exceed 70. This gives us,

$$\begin{aligned} x_1 + x_2 &\le 70 \\ x_2 + x_3 &\le 70 \\ x_1 + x_3 &\le 70. \end{aligned}$$

The water requirement of every village puts an upper bound on the supply. So, we can put $x_1 \leq 30$, $x_2 \leq 50$, $x_3 \leq 40$. Of course, the water supply should all be non-negative, i.e., $x_1, x_2, x_3 \geq 0$. Finally, the total water shortage costs of the three villages are $4(30 - x_1) + 3(50 - x_2) + 5(40 - x_3) = 470 - 4x_1 - 3x_2 - 5x_3$. If we want to minimize the total water shortage cost, then it is equivalent to just maximizing $4x_1 + 3x_2 + 5x_3$. So, the problem can be **formulated** as:

$$
\begin{aligned}
Z = \max_{x_1, x_2, x_3} \; & 4x_1 + 3x_2 + 5x_3 \\
\text{s.t.} \quad & \\
& \sum_{i=1}^{3} x_i = 100 \\
& x_i + x_j \leq 70 \qquad \forall\, i, j \in \{1, 2, 3\},\ i \neq j \\
& x_1 \leq 30 \\
& x_2 \leq 50 \\
& x_3 \leq 40 \\
& x_i \geq 0 \qquad i \in \{1, 2, 3\}
\end{aligned}
\tag{\textbf{P1}}
$$

Problems of this type are called linear programming formulations.

3.2 Standard Form

In general, if $c_1, \ldots, c_n$ are real numbers, then the function f of real variables $x_1, \ldots, x_n$ ($\mathbf{x} = (x_1, \ldots, x_n)$) defined by

$$f(\mathbf{x}) = c_1 x_1 + \ldots + c_n x_n = \sum_{j=1}^{n} c_j x_j$$

is called a **linear function**. If g is a linear function and b is a real number then

$$g(\mathbf{x}) = b$$

is called a **linear equation**, whereas

$$g(\mathbf{x}) \leq (\geq) b$$

is called a **linear inequality**. A **linear constraint** is one that is either a linear equation or a linear inequality. A **linear programming (LP)** problem is one which maximizes (minimizes) a linear function subject to (s.t.) a finite collection of linear constraints. Formally, any LP can be written in the following form:

$$
\begin{aligned}
& Z = \max_{\mathbf{x}} \sum_{j=1}^{n} c_j x_j \\
\text{s.t.} \quad & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathbf{LP}) \\
& \sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad \forall\, i \in \{1, \ldots, m\} \\
& x_j \ge 0 \qquad \forall\, j \in \{1, \ldots, n\}.
\end{aligned}
$$

In matrix notation, this can be written as $\max_x cx$ subject to $Ax \le b$ and $x \ge 0$. Problems in the form (**LP**) will be referred to as the problems in **standard form**. As we have seen any LP problem can be converted to a problem in standard form. The key difference between any LP problem not in standard form and a problem in standard form is that the constraints in standard form are all inequalities (written in a particular way). Also, the last collection of constraints say that variable have to be non-negative. This type of inequalities are special, and referred to as **non-negativity constraints**. The linear function that is to be maximized or minimized is called the **objective function**. In the standard form, the objective function will always be maximized (this is only our notation).

If $(x_1^*, \ldots, x_n^*)$ satisfy all the constraints of (**LP**), then it is called a **feasible solution** of (**LP**). For example, in the problem (**P1**), a feasible solution is $(x_1, x_2, x_3) = (30, 35, 35)$. A feasible solution that gives the maximum value to the objective function amongst all feasible solutions is called an **optimal solution**, and the corresponding value of the objective function is called the **optimal value**. The optimal solution of (**LP**) is $(x_1, x_2, x_3) = (30, 30, 40)$, and the optimal value is 410 (hence the minimum total water shortage cost is 60).

Not every LP problem has an optimal solution. As we will show later, every LP problem can be put in one of the following three categories.

1. **Optimal solution exists**: This is the class of LP problems whose optimal solution exists. An example is (**P1**).

2. **Infeasible**: This is the class of LP problems for which no feasible solution exists. An example is the following:

$$
\begin{aligned}
& Z = \max_{x_1, x_2} x_1 + 5x_2 \\
\text{s.t.} \quad & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathbf{INF\text{-}LP}) \\
& x_1 + x_2 \le 3 \\
& -3x_1 - 3x_2 \le -11 \\
& x_1, x_2 \ge 0
\end{aligned}
$$

3. **Unbounded**: This is the class of LP problems for which feasible solutions exist, but for every number M, there exists a feasible solution that gives the objective function a value more than M. So, none of the feasible solutions is optimal. An example is the following:

$$
\begin{aligned}
Z = \max_{x_1,x_2} x_1 - x_2 & \\
\text{s.t.} & \qquad\qquad (\textbf{UNBD-LP}) \\
-2x_1 + x_2 \leq -1 & \\
-x_1 - 2x_2 \leq -2 & \\
x_1, x_2 \geq 0 &
\end{aligned}
$$

To understand why (**UNBD-LP**) is unbounded, it is useful to look at its feasible region and objective function in a figure. Figure 2 shows how the objective function can increase indefinitely.

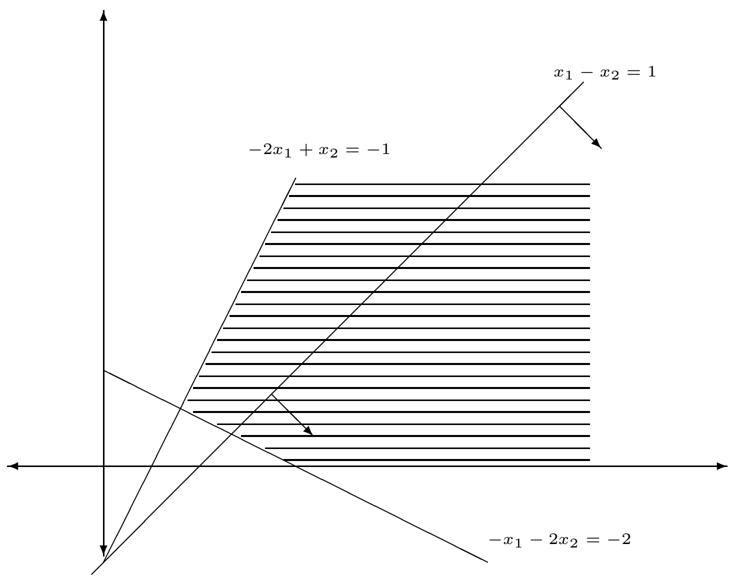

Figure 2: Unbounded LP

4 History of Linear Programming

The second world war brought about many new things to the world. This included the use and rapid growth of the field of linear programming. In 1947, **George B. Dantzig**, regarded by many as the founder of the discipline, designed the **simplex method** to solve linear programming problems for the U.S. Air Force. After that, the field showed rapid growth in terms of research. Production management and economics were the primary areas which applied linear programming in a variety of problems. Tremendous application potential of this field led to increase in theoretical research in the field, and thus, a new branch of applied mathematics was born.

In 1947, T.C. Koopmans pointed out several applications of linear programming in economic theory. Till today, a significant portion of economic theory is still governed by the fundamentals of linear programming (as we will discuss).

The fundamentals of linear programming has its roots as early as 1820s, when Fourier investigated techniques to solve systems of linear equations. L.V. Kantorovich pointed out the significance of linear programming and its applications in 1939. Unfortunately, his work was not noticed for a long time (since he carried out most of his research in U.S.S.R.). In 1975, Kantorovich and Koopmans were awarded the Nobel prize in economics "for their contributions to the theory of optimum allocation of resources". Another event in 1970s attracted a lot of media attention. Ever since the invention of simplex method, mathematicians were working for a *theoretically* satisfactory algorithm to solve linear programs. In 1979, L.G. Khachian published the description of such an algorithm - though its performance has been extremely unsatisfactory in practice. On the other hand, simplex method, whose theortical performance is not good, does a very good job in practice.

5 Simplex Preview

One of the first discovered, and immensely effective linear programming (LP) algorithms is the **simplex method**. The objective of this section is to give examples to illustrate the method.

5.1 First Example

Consider the following example.

$$Z = \max 5x_1 + 4x_2 + 3x_3$$

$$\text{s.t.} \qquad \textbf{(EX-1)}$$

$$2x_1 + 3x_2 + x_3 \leq 5 \qquad (1)$$

$$4x_1 + x_2 + 2x_3 \leq 11 \qquad (2)$$

$$3x_1 + 4x_2 + 2x_3 \leq 8 \qquad (3)$$

$$x_1, x_2, x_3 \geq 0.$$

The first step in the method consists of introducing **slack variables** for every constraint. For example, in Equation 1, the *slack* between 5 and $2x_1+3x_2+x_3$ is assigned a slack variable

x_4, i.e., $x_4 = 5 - 2x_1 - 3x_2 - x_3$. Notice that $x_4 \geq 0$. Thus, the equations in formulation (**EX-1**) can be rewritten using slack variables as

$$\begin{aligned} x_4 &= 5 - 2x_1 - 3x_2 - x_3 \\ x_5 &= 11 - 4x_1 - x_2 - 2x_3 \\ x_6 &= 8 - 3x_1 - 4x_2 - 2x_3 \\ x_1, x_2, x_3, x_4, x_5, x_6 &\geq 0. \end{aligned}$$

The new variables x_4, x_5, x_6 are called **slack variables**, and the old variables x_1, x_2, x_3 are called **decision variables**. Hence our new LP is to

$$\max z \qquad \text{s.t.} \qquad x_1, x_2, x_3, x_4, x_5, x_6 \geq 0, \tag{4}$$

where $z = 5x_1 + 4x_2 + 3x_3$ and x_4, x_5, and x_6 are determined by the equations above. This new LP is equivalent (same set of feasible solutions in terms of decision variables) to (**EX-1**), given the equations determining the slack variables. The simplex method is an iterative procedure in which having found a feasible solution $x_1, \ldots, x_6$ of (4), we look for another feasible solution $\bar{x}_1, \ldots, \bar{x}_6$ of (4) such that

$$5\bar{x}_1 + 4\bar{x}_2 + 3\bar{x}_3 > 5x_1 + 4x_2 + 3x_3.$$

If an optimal solution exists, we can repeat this finite number of iterations till there is no improvement in the objective function value, at which point we stop. The first step is to find a feasible solution, which is easy in our example: $x_1 = x_2 = x_3 = 0$, which gives $x_4 = 5, x_5 = 11, x_6 = 8$. This gives $z = 0$.

We now need to look for a solution that gives a higher value to z. For this, we look to increase values of any one of the variables x_1, x_2, x_3. We choose x_1. Keeping x_2 and x_3 at zero, we notice that we can increase x_1 to $\min(\frac{5}{2}, \frac{11}{4}, \frac{8}{3}) = \frac{5}{2}$ to maintain $x_4, x_5, x_6 \geq 0$. As a result of this, the new solution is

$$x_1 = \frac{5}{2}, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 1, x_6 = \frac{1}{2}, z = \frac{25}{2}.$$

Notice that by increasing the value of x_1, a variable whose value was positive (x_4) got a value of zero. Now, we have to create system of equation similar to previous iteration. For that we will write the value of z and variables having non-zero values (x_1, x_5, x_6) in terms of variables having zero values (x_4, x_2, x_3).

$$x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4.$$
$$x_5 = 1 + 5x_2 + 2x_4.$$
$$x_6 = \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4.$$

$$z = \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4.$$

Of x_2, x_3, x_4, the value of z decreases by increasing the values of x_2 and x_4. So, the only candidate for increasing value in this iteration is x_3. The amount we can increase the value of x_3 can again be obtained from the feasibility conditions of $x_1, x_5, x_6 \geq 0$, which is equivalent to (given $x_2 = x_4 = 0$) $\frac{5}{2} - \frac{1}{2}x_3 \geq 0$ and $\frac{1}{2} - \frac{1}{2}x_3 \geq 0$. This gives that the maximum possible value of x_3 in this iteration can be $\min(5, 1) = 1$. By setting $x_3 = 1$, we get a new solution as

$$x_1 = 2, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 1, x_6 = 0, z = 13. \quad (5)$$

Two things should be noticed here: (a) this solution is also a solution of the previous system of equations and (b) the earlier solution is also a solution of this system of equations. This is precisely because we are just rewriting the system of equations using a different set of decision and slack variables in every iteration, and that is the central theme of the simplex method.

So, the new variable that takes zero value is x_6. We now write the system of equations in terms of x_2, x_4, x_6.

$$x_3 = 1 + x_2 + 3x_4 - 2x_6$$
$$x_1 = 2 - 2x_2 - 2x_4 + x_6$$
$$x_5 = 1 + 5x_2 + 2x_4$$

$$z = 13 - 3x_2 - x_4 - x_6.$$

Now, the value of z will decrease by increasing the values of any of the variables x_2, x_4, x_6. So, we have reached a dead-end. In fact, we have reached an optimal solution. This is clear from the fact that any solution requires $x_2, x_4, x_6 \geq 0$, and by assigning any value not equal to zero to these variables, we will decrease the value of z. Hence, $z = 13$ is an optimal solution. The corresponding values of x_1, x_3, x_5 are $2, 1, 1$ respectively.

5.2 Dictionaries

Consider a general LP in standard form:

$$\begin{aligned} Z = \max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \\ \sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad \forall\, i \in \{1, \ldots, m\} \\ x_j \geq 0 \qquad \forall\, j \in \{1, \ldots, n\} \end{aligned} \tag{\textbf{LP}}$$

The first step in the simplex method is to introduce slack variables, $x_{n+1}, \ldots, x_{n+m} \geq 0$ corresponding to m constraints, and denote the objective function as z. So,

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j \qquad \forall\, i \in \{1, \ldots, m\} \tag{6}$$

$$z = \sum_{j=1}^{n} c_j x_j. \tag{7}$$

$$x_j \geq 0 \qquad \forall\, j \in \{1, \ldots, n, n+1, \ldots, n+m\} \tag{8}$$

In simplex method, we search for a feasible solution $\bar{x}_1, \ldots, \bar{x}_{m+n}$ given a feasible solution $x_1, \ldots, x_{m+n}$ so that the objective function is better, i.e.,

$$\sum_{j=1}^{n} c_j \bar{x}_j > \sum_{j=1}^{n} c_j x_j.$$

As we have seen in the example, a feasible solution is represented with a system of linear equations consisting of *dependent* variables. These system of equations corresponding to a feasible solution is called a **dictionary**. A dictionary will have the following features:

1. Every solution of the system of equations of the dictionary must be a solution of system of equations (6), (7), and (8), and vice versa.

2. The equations of every dictionary must express m of the variables $x_1, \ldots, x_{m+n}$ and the objective function z (dependent variables) in terms of the remaining n variables (independent variables).

Consider the following starting dictionary.

$$\begin{aligned} x_3 &= 5 - x_2 + x_1 \\ x_4 &= 3 - x_2 - 2x_1 \\ z &= x_1 + 3x_2 \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

In this dictionary, we can set $x_1 = x_2 = 0$ to get a feasible solution. Rewriting the first equation in terms of x_2, we get the following dictionary.

$$\begin{aligned} x_2 &= 5 + x_1 - x_3 \\ x_4 &= -2 - 3x_1 + x_3 \\ z &= 15 + 4x_1 - 3x_3 \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

Unlike the first dictionary, we cannot put the value of independent variables to zero to get a feasible solution: putting $x_1 = x_3 = 0$ gives us $x_2 = 5$ but $x_4 = -2 < 0$. This is an undesirable feature. To get over this feature, we need the following notion.

In the dictionary, the dependent variables are kept on the left hand side (LHS), and they are expressed in terms of the independent variables on the right hand side (RHS). An additional feature of a dictionary is

- setting the RHS variables at zero and evaluating the LHS variables, we arrive at a feasible solution.

A dictionary with this additional property is called a **feasible dictionary**. Hence, every feasible dictionary describes a feasible solution. The second dictionary above is not feasible but the first one is.

A feasible solution that can be described in terms of a feasible dictionary is called a **basic solution**. The characteristics feature of the simplex method is that it works with basic solutions only.

5.3 Second Example

We conclude the discussion by giving another example.

$$
\begin{aligned}
Z = \max 5x_1 + 5x_2 + 3x_3 \\
\text{s.t.} \\
x_1 + 3x_2 + x_3 &\le 3 \\
-x_1 + 3x_3 &\le 2 \\
2x_1 - x_2 + 2x_3 &\le 4 \\
2x_1 + 3x_2 - x_3 &\le 2 \\
x_1, x_2, x_3 &\ge 0.
\end{aligned}
$$

In this case, the initial feasible dictionary has all the slack variables as dependent variables, and it looks as follows:

$$
\begin{aligned}
x_4 &= 3 - x_1 - 3x_2 - x_3 \\
x_5 &= 2 + x_1 - 3x_3 \\
x_6 &= 4 - 2x_1 + x_2 - 2x_3 \\
x_7 &= 2 - 2x_1 - 3x_2 + x_3
\end{aligned}
$$

$$
z = 5x_1 + 5x_2 + 3x_3.
$$

The feasible dictionary describes the following solution:

$$
x_1 = x_2 = x_3 = 0, x_4 = 3, x_5 = 2, x_6 = 4, x_7 = 2.
$$

As before, we try to increase the value of z by increasing the value of one of the independent variables as much as we can. Right now, since x_1, x_2, x_3 have all positive coefficients in the z equation, we randomly choose x_1. From feasibility of $x_4, \ldots, x_7 \ge 0$, we get $x_1 \le 1$ to be the most stringent constraint. So, we make $x_1 = 1$, which in turn makes $x_7 = 0$. We now write the new dictionary with x_1 leaving the independent variables and x_7 entering the independent variables. First, substitute,

$$
x_1 = 1 - \frac{3}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_7.
$$

Substituting for x_1 in terms of new set of independent variables in the previous dictionary, we get

$$\begin{aligned}
x_1 &= 1 - \frac{3}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_7 \\
x_4 &= 2 - \frac{3}{2}x_2 - \frac{3}{2}x_3 + \frac{1}{2}x_7 \\
x_5 &= 3 - \frac{3}{2}x_2 - \frac{5}{2}x_3 - \frac{1}{2}x_7 \\
x_6 &= 2 + 4x_2 - 3x_3 + x_7
\end{aligned}$$

$$z = 5 - \frac{5}{2}x_2 + \frac{11}{2}x_3 - \frac{5}{2}x_7.$$

Some comments about terminology are in order:

1. Dependent variables, which appear on the LHS of any dictionary, are called **basic variables**. Independent variables are called **non-basic variables**. In the previous dictionary, x_1, x_4, x_5, x_6 are basic variables, and x_2, x_3, x_7 are non-basic variables.

2. Set of basic and non-basic variables change from iteration to iteration.

3. Choice of **entering** basic variable is motivated by the fact that we want to increase the value of z, and we choose one that does that, and increase its value the maximum possible.

4. Choice of **leaving** basic variable is motivated by the need to maintain feasibility. This is done by identifying the basic variable that poses the most stringent bound on the entering basic variable.

5. The formula for the entering basic variable appears in the **pivot row**, and the process of constructing a new dictionary is called **pivoting**. In the previous dictionary, x_3 is the next entering basic variable, and x_6 is the leaving basic variable. So, the formula for x_3 appears in

$$x_3 = \frac{2}{3} + \frac{4}{3}x_2 - \frac{1}{3}x_6 + \frac{1}{3}x_7,$$

   which is the pivot row.

Continuing with our example, the clear choice of entering basic variable is x_3. Calculations give that x_6 imposes the most stringent bound on x_3, and should be the leaving basic variable. So, we arrive at the new dictionary.

$$x_3 = \frac{2}{3} + \frac{4}{3}x_2 + \frac{1}{3}x_7 - \frac{1}{3}x_6$$
$$x_1 = \frac{4}{3} - \frac{5}{6}x_2 - \frac{1}{3}x_7 - \frac{1}{6}x_6$$
$$x_4 = 1 - \frac{7}{2}x_2 + \frac{1}{2}x_6$$
$$x_5 = \frac{4}{3} - \frac{29}{6}x_2 - \frac{4}{3}x_7 + \frac{5}{6}x_6$$

$$z = \frac{26}{3} + \frac{29}{6}x_2 - \frac{2}{3}x_7 - \frac{11}{6}x_6.$$

Now, the entering basic variable is x_2, and the leaving basic variable is x_5. Pivoting yields the following dictionary:

$$x_2 = \frac{8}{29} - \frac{8}{29}x_7 + \frac{5}{29}x_6 + \frac{6}{29}x_5$$

$$x_3 = \frac{30}{29} + \frac{1}{29}x_7 - \frac{3}{29}x_6 - \frac{8}{29}x_5$$
$$x_1 = \frac{32}{29} - \frac{3}{29}x_7 - \frac{9}{29}x_6 + \frac{5}{29}x_5$$
$$x_4 = \frac{1}{29} + \frac{28}{29}x_7 - \frac{3}{29}x_6 + \frac{21}{29}x_5$$

$$z = 10 - 2x_7 - x_6 - x_5.$$

At this point, no more pivoting is possible, and we arrive at the optimal solution described by the last dictionary as:

$$x_1 = \frac{32}{29}, x_2 = \frac{8}{29}, x_3 = \frac{30}{29},$$

and this yields an optimal value of $z = 10$.

6 Pitfalls and How to Avoid Them

Three kind of pitfalls can occur in simplex method:

1. **Initialization**: We may not be able to start. We may not have a feasible dictionary to start.

2. **Iteration**: We may get stuck in some iteration. Can we always choose a new entering and leaving variable?

3. **Termination**: We may not be able to finish. Can the simplex method construct an endless sequence of dictionaries without reaching an optimal solution?

We look at each of these three pitfalls. Before proceeding, let us review the general form of a dictionary. There is a set of basic variables B with $\#B = m$, and the linear program is written in the following form in this dictionary.

$$x_i = \bar{b}_i - \sum_{j \notin B} \bar{a}_{ij} x_j \qquad \forall\, i \in B$$
$$z = \bar{v} + \sum_{j \notin B} \bar{c}_j x_j$$

Here, for each $i \in B$, $\bar{b}_i \geq 0$ if the dictionary is a feasible dictionary.

6.1 Iteration

6.1.1 Choosing an Entering Variable

The entering variable is a *non-basic variable with a positive coefficient $\bar{c}_j$ in the last row of the current dictionary*. This rule is ambiguous in the sense that it may provide more than one candidate for entering or no candidate at all.

The latter alternative implies that the current dictionary has an optimal solution. This is because any solution which is not the current solution will involve some current non-basic variable taking on positive value. Since all the $\bar{c}_j$s are negative, this will imply objective function value decreasing from the current value.

If there are more than one candidate for entering the basis, then any of these candidates may serve.

6.1.2 Finding a Leaving Variable

The leaving variable is that *basic variable whose non-negativity imposes the most stringent upper bound on the increase of the entering variable*. Again, there may be more than one candidate or no candidate at all.

If there are more than one candidate, then we may choose any one of them.

If there are no candidate at all, then an interesting conclusion can be drawn. Recall that a linear program is **unbounded** if for every real number M there exists a feasible solution of the linear program such that the objective function value is larger than M.

Here is an example of a dictionary:

$$x_2 = 5 + 2x_3 - x_4 - 3x_1$$
$$x_5 = 7 - 3x_4 - 4x_1$$

$$z = 5 + x_3 - x_4 - x_1.$$

The entering variable is x_3. However, neither of the two basic variables x_2 and x_5 put an upper bound on x_3. Hence, we can increase x_3 as much as we want without violating feasibility. Set $x_3 = t$ for any positive number t, and we get the solution $x_1 = 0, x_2 = 5 + 2t, x_3 = t, x_4 = 0, x_5 = 7$, and $z = 5 + t$. Since t can be made arbitrarily large, so can be z, and we conclude that the problem is unbounded. The same conclusion can be reached in general: if there is no candidate for leaving the basis, then we can make the value of the entering variable, and hence the value of the objective function, as large as we wish. In that case, the problem is unbounded.

6.1.3 Degeneracy

The presence of more than one candidate for leaving the basis has interesting consequences. For example, consider the dictionary

$$x_4 = 1 - 2x_3$$
$$x_5 = 3 - 2x_1 + 4x_2 - 6x_3$$
$$x_6 = 2 + x_1 - 3x_2 - 4x_3$$

$$z = 2x_1 - x_2 + 8x_3.$$

Having chosen x_3 as the entering variable, we see that x_4, x_5, and x_6 are all candidates for leaving variable. Choosing x_4, and pivoting, we get the new dictionary as

$$x_3 = 0.5 - 0.5x_4$$
$$x_5 = -2x_1 + 4x_2 + 3x_4$$
$$x_6 = x_1 - 3x_2 + 2x_4$$

$$z = 4 + 2x_1 - x_2 - 4x_4.$$

This dictionary is different from others in one important aspect: along with the non-basic variables, two of the basic variables, x_5 and x_6 have value of zero. Basic solutions with one or more basic variables at zero are called **degenerate**.

Although harmless, degeneracy has annoying side effects. In the next iteration, we have x_1 as the entering variable, and x_5 as the leaving variable. But the value of x_1 can be increased by a maximum of zero. Hence, the objective function value does not change. Pivoting changes the dictionary to:

$$
\begin{aligned}
x_1 &= 2x_2 + 1.5x_4 - 0.5x_5 \\
x_3 &= 0.5 - 0.5x_4 \\
x_6 &= -x_2 + 3.5x_4 - 0.5x_5
\end{aligned}
$$

$$z = 4 + 3x_2 - x_4 - x_5.$$

but the solution remains the same. Simplex iterations that do not change the basic solution are called **degenerate**. One can verify that the next iteration is also degenerate, but the one after that is not - in fact, it is the optimal solution.

Degeneracy is an accident. Many practical problems face degeneracy, and when it happens the simplex goes through few (many a times quite a few) degenerate iterations before coming up with a non-degenerate solution. But there are occasions when this may not happen.

6.2 Cycling

Sometimes, a sequence of dictionary can appear again and again. This phenomenon is called **cycling**. To understand cycling let us look at a series of dictionaries.

$$
\begin{aligned}
x_5 &= -0.5x_1 + 5.5x_2 + 2.5x_3 - 9x_4 \\
x_6 &= -0.5x_1 + 1.5x_2 + 0.5x_3 - x_4 \\
x_7 &= 1 - x_1
\end{aligned}
$$

$$z = 10x_1 - 57x_2 - 9x_3 - 24x_4.$$

The following rule for selecting the entering and leaving variable is the following:

- The entering variable will always be the nonbasic variable that the largest coefficient in the z-row of the dictionary.
- If two or more basic variables compete for leaving the basis, then the candidate with the smallest subscript will be made to leave.

Now, the sequence of dictionaries constructed in the first six iterations goes as follows. After the first iteration:

$$\begin{aligned} x_1 &= 11x_2 + 5x_3 - 18x_4 - 2x_5 \\ x_6 &= -4x_2 - 2x_3 + 8x_4 + x_5 \\ x_7 &= 1 - 11x_2 - 5x_3 + 18x_4 + 2x_5 \end{aligned}$$

$$z = 53x_2 + 41x_3 - 20x_4 - 20x_5.$$

After the second iteration:

$$\begin{aligned} x_2 &= -0.5x_3 + 2x_4 + 0.25x_5 - 0.25x_6 \\ x_1 &= -0.5x_3 + 4x_4 + 0.75x_5 - 2.75x_6 \\ x_7 &= 1 + 0.5x_3 - 4x_4 - 0.75x_5 - 13.25x_6 \end{aligned}$$

$$z = 14.5x_3 - 98x_4 - 6.75x_5 - 13.25x_6.$$

After the third iteration:

$$\begin{aligned} x_3 &= 8x_4 + 1.5x_5 - 5.5x_6 - 2x_1 \\ x_2 &= -2x_4 - 0.5x_5 + 2.5x_6 + x_1 \\ x_7 &= 1 - x_1 \end{aligned}$$

$$z = 18x_4 + 15x_5 - 93x_6 - 29x_1.$$

After the fourth iteration:

$$\begin{aligned} x_4 &= -0.25x_5 + 1.25x_6 + 0.5x_1 - 0.5x_2 \\ x_3 &= -0.5x_5 + 4.5x_6 + 2x_1 - 4x_2 \\ x_7 &= 1 - x_1 \end{aligned}$$

$$z = 10.5x_5 - 70.5x_6 - 20x_1 - 9x_2.$$

After the fifth iteration:

$$\begin{aligned} x_5 &= 9x_6 + 4x_1 - 8x_2 - 2x_3 \\ x_4 &= -x_6 - 0.5x_1 + 1.5x_2 + 0.5x_3 \\ x_7 &= 1 - x_1 \end{aligned}$$

$$z = 24x_6 + 22x_1 - 93x_2 - 21x_3.$$

After the sixth iteration:

$$
\begin{aligned}
x_5 &= -0.5x_1 + 5.5x_2 + 2.5x_3 - 9x_4 \\
x_6 &= -0.5x_1 + 1.5x_2 + 0.5x_3 - x_4 \\
x_7 &= 1 - x_1
\end{aligned}
$$

$$z = 10x_1 - 57x_2 - 9x_3 - 24x_4.$$

Since the dictionary after the sixth iteration is identical with the initial dictionary, the simplex method will go through the same set of dictionaries again and again without ever finding the optimal solution (which is $z = 1$ in this example).

Notice that cycling means, we have a series of degenerate solutions, else we will have increase in objective function, and cannot have the same dictionaries repeating. It is important to note that **cycling implies that we get the same solution in every iteration, even though the set of basic variables change**. It is not possible that we are changing the value of some variable without changing the objective function value (because we always choose an entering variable that changes the objective function value when its value is changed).

THEOREM **1** *If the simplex method fails to terminate, then it must cycle.*

Proof: There are a total of $m + n$ variables. Since in every iteration of the simplex method we choose m basic variables, there are finite number of ways to choose them. Hence, if the simplex method does not terminate, then there will be two dictionaries with the same set of basic variables. Represent the two dictionaries as:

$$x_i = b_i - \sum_{j \notin B} a_{ij} x_j \qquad \forall\, i \in B$$

$$z = v + \sum_{j \notin B} c_j x_j.$$

and

$$x_i = b_i^* - \sum_{j \notin B} a_{ij}^* x_j \qquad \forall\, i \in B$$

$$z = v^* + \sum_{j \notin B} c_j^* x_j.$$

with the same set of basic variables $x_i (i \in B)$.

But there is a unique way of representing a (basic) variable in terms of a set of non-basic variable. Hence the two dictionaries must be exactly equal. ∎

Cycling is a rare phenomena, but sometimes they do occur in practice. In fact, constructing an LP problem on which the simplex method may cycle is difficult. It is known that if the simplex method cycles off-optimum on a problem that has an optimal solution, then the dictionaries must involve at least six variables and at least three equations. In practice, cycling occurs very rarely.

Two popular rules for avoiding cycling are: (a) perturbation method and lexicographic ordering (b) smallest subscript rule. We describe the smallest subscript rule here. The former requires extra computation to choose the entering and leaving variables while the latter leaves no choice in the hands of users to choose entering variables, which we can get in the former one.

To avoid cycling, we introduce a rule called (Bland's) **smallest subscript rule**. This refers to breaking ties in the choice of the entering and leaving variables by always choosing the candidate x_k that has the smallest subscript k.

THEOREM **2** *The simplex method terminates as long as the entering and leaving variables are selected by the smallest subscript rule (SSR) in each iteration.*

Proof: By virtue of previous theorem, we need to show that cycling is impossible when the SSR is used. Assume for contradiction that the simplex method with SSR generates a sequence of dictionaries $D_0, D_1, \ldots, D_k$ such that $D_k = D_0$.

Call a variable **fickle** if it is nonbasic in some iteration and basic in some other. Among all fickle variables, let x_t have the largest subscript. Due to cycling, there is a dictionary D in the sequence $D_0, \ldots, D_k$ with x_t leaving (basic in D but nonbasic in the next dictionary), and some other fickle variable x_s entering (nonbasic in D but basic in the next iteration). Further along in the sequence $D_0, D_1, \ldots, D_k, D_1, \ldots, D_k$, there is a dictionary D^* with x_t entering. Let us record D as

$$x_i = b_i - \sum_{j \notin B} a_{ij} x_j \qquad \forall\ i \in B$$

$$z = v + \sum_{j \notin B} c_j x_j.$$

Since all iterations leading from D to D^* are degenerate, the objective function z must have the same value v in both D and D^*. Thus, the last row of D^* may be recorded as

$$z = v + \sum_{j=1}^{m+n} c_j^* x_j,$$

where $c_j^* = 0$ wherever x_j is basic in D^*. Since this equation has been obtained from D by algebraic manipulations, it must satisfy every solution of D. In particular, it must be satisfied by

$$x_s = y, x_j = 0 \ (j \notin B, j \neq s), x_i = b_i - a_{is} y \ (i \in B), z = v + c_s y \qquad \forall \ y.$$

Thus, we have

$$v + c_s y = v + c_s^* y + \sum_{i \in B} c_i^* (b_i - a_{is} y).$$

and, after simplification,

$$\Big(c_s - c_s^* + \sum_{i \in B} c_i^* a_{is}\Big) y = \sum_{i \in B} c_i^* b_i$$

for every choice of y. Since the RHS of the previous equation is a constant independent of y, we have

$$c_s - c_s^* + \sum_{i \in B} c_i^* a_{is} = 0.$$

But x_s is entering in D, implying $c_s > 0$. Since x_s is not entering in D^* and yet $s < t$, we have $c_s^* \leq 0$. Hence for some $r \in B$, $c_r^* a_{rs} < 0$. Note two points now:

- Since $r \in B$, the variable x_r is basic in D; since $c_r^* \neq 0$, the same variable is nonbasic in D^*. Hence, x^r is fickle, and we have $r \leq t$.
- $r \neq t$: since x_t is leaving in D, we have $a_{ts} > 0$ and so $c_t^* a_{ts} > 0$ (since $c_t^* > 0$ with x_t entering in D^*).

This shows that $r < t$ and yet x_r is not entering in D^*. Thus, $c_r^* \leq 0$, and $a_{rs} > 0$. Since all iterations from D to D^* are degenerate, the two dictionaries describe the same solution. Since x_r is non-basic in D^*, its value is zero in both D and D^*, meaning $b_r = 0$. Hence, x_r was a candidate for leaving the basis of D - yet we picked x_t, even though $r < t$. This is a contradiction. ■

6.3 Initialization

The only remaining point that needs to be explained is getting hold of the initial feasible dictionary in a problem

$$\begin{aligned} &\max \sum_{j=1}^{n} c_j x_j \\ &\text{s.t.} \\ &\sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad \forall\, i \in \{1, \ldots, m\} \\ &x_j \geq 0 \qquad \forall\, j \in \{1, \ldots, n\}. \end{aligned}$$

with an infeasible origin. The problem with infeasible origin is that we may not know whether a feasible solution exists at all, and even we know what a feasible dictionary will be for that solution. One way of getting around these two solutions is the so called **auxiliary problem**:

$$\begin{aligned} &\min x_0 \\ &\text{s.t.} \\ &\sum_{j=1}^{n} a_{ij} x_j - x_0 \leq b_i \qquad \forall\, i \in \{1, \ldots, m\} \\ &x_j \geq 0 \qquad \forall\, j \in \{0, 1, \ldots, n\}. \end{aligned}$$

A feasible solution of the auxiliary problem is readily available: set $x_j = 0$ for $j \neq 0$ and make the value of x_0 sufficiently large.

Theorem 3 *The original LP problem has a feasible solution if and only if the optimal value of the associated auxiliary problem is zero.*

Proof: If the original problem has a feasible solution than the auxiliary problem has the same feasible solution with $x_0 = 0$. This is clearly the optimal value. Further if the auxiliary problem has a feasible (optimal) solution with $x_0 = 0$, then the original problem has the same feasible solution. ■

Hence, our objective is to solve the auxiliary problem. Consider the following example.

$$\begin{aligned}
\max x_1 - x_2 + x_3 \\
\text{s.t.} \\
2x_1 - x_2 + 2x_3 \leq 4 \\
2x_1 - 3x_2 + x_3 \leq -5 \\
-x_1 + x_2 - 2x_3 \leq -1 \\
x_1, x_2, x_3 \geq 0.
\end{aligned}$$

To avoid unnecessary confusion, we write the auxiliary problem in its maximization form, and construct the dictionary as

$$\begin{aligned}
x_4 &= 4 - 2x_1 + x_2 - 2x_3 + x_0 \\
x_5 &= -5 - 2x_1 + 3x_2 - x_3 + x_0 \\
x_6 &= -1 + x_1 - x_2 + 2x_3 + x_0 \\
\\
w &= -x_0,
\end{aligned}$$

which is an infeasible dictionary. But it can be made feasible by pivoting on the most negative b_i row, i.e., x_5 in this case, and choosing x_0 as the entering variable. The new (feasible) dictionary is:

$$\begin{aligned}
x_0 &= 5 + 2x_1 - 3x_2 + x_3 + x_5 \\
x_4 &= 9 - 2x_2 - x_3 + x_5 \\
x_6 &= 4 + 3x_1 - 4x_2 + 3x_3 + x_5 \\
\\
z &= -5 - 2x_1 + 3x_2 - x_3 - x_5.
\end{aligned}$$

In general, the dictionary corresponding to the auxiliary problem is:

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij}x_j + x_0 \qquad \forall\ i \in \{1, \ldots, m\}$$

$$w = -x_0.$$

which is infeasible. However, this can be transformed into a feasible dictionary. This is done by a single pivot in which x_0 enters and the "most infeasible" x_{n+i} leaves. More precisely, the leaving variable is that x_{n+k} whose b_k has the largest negative value among all negative b_is. After pivoting, x_0 assumes the positive value $-b_k$, and each basic x_{n+i} assumes the non-negative value $b_i - b_k$. Now, we can continue with our simplex method. In our example, two more iterations yield the following dictionary:

$$
\begin{aligned}
x_3 &= 1.6 - 0.2x_1 + 0.2x_5 + 0.6x_6 - 0.8x_0 \\
x_2 &= 2.2 + 0.6x_1 + 0.4x_5 + 0.2x_6 - 0.6x_0 \\
x_4 &= 3 - x_1 - x_6 + 2x_0 \\
\\
w &= -x_0.
\end{aligned}
$$

This dictionary is an optimal solution of the auxiliary problem with $x_0 = 0$. Further, this points to a feasible dictionary of the original problem.

$$
\begin{aligned}
x_3 &= 1.6 - 0.2x_1 + 0.2x_5 + 0.6x_6 \\
x_2 &= 2.2 + 0.6x_1 + 0.4x_5 + 0.2x_6 \\
x_4 &= 3 - x_1 - x_6 \\
\\
z &= -0.6 + 0.2x_1 - 0.2x_5 + 0.4x_6.
\end{aligned}
$$

So, we learned how to construct the auxiliary problem, and its first feasible dictionary. In the process of solving the auxiliary problem, **it may be possible that x_0 may be a candidate for the leaving variable in which case we *pick* x_0**. Immediately, after pivoting we get

- x_0 as a non-basic variable, in which case $w = 0$.

Clearly, this is an optimal solution. However, we may also reach an optimal dictionary of auxiliary problem with x_0 basic. If value of w is non-zero in that, then we simply conclude that the original problem is infeasible. Else, x_0 is basic and the optimal value of w is zero. We argue that this is not possible. Since the dictionary preceding the final dictionary was not optimal, the value of $w = -x_0$ must have changed from some negative value to zero in the final iteration. To put it differently, the value of x_0 must have changed from some positive level to zero in this iteration. This means, x_0 was also a candidate for leaving the basis, and we should have picked it according to our policy. This is a contradiction.

Hence, we either construct an optimal solution of the auxiliary problem where x_0 is a non-basic variable, and we proceed to the original problem by constructing a new feasible dictionary, or we conclude that the original problem is infeasible.

This strategy of solving an LP is known as the **two phase simplex method**. In the first phase, we set up and solve the auxiliary problem; if we find an optimal solution of the auxiliary problem, then we proceed to the **second phase**, solving the original problem.

THEOREM **4 (Fundamental Theorem of Linear Programming)** *Every LP problem in the standard form has the following three properties:*

1. *If it has no optimal solution, then it is either unbounded or infeasible.*

2. *If it has a feasible solution, then it has a basic feasible solution.*

3. *If it has an optimal solution, then it has a basic optimal solution.*

Proof: The first phase of the two phase simplex method either discovers that the problem is infeasible or else it delivers a basic feasible solution. The second phase either discovers that the problem is unbounded or gives a basic optimal solution. ■

Note that if the problem is not in standard form then the theorem does not hold, e.g., max x s.t. $x < 0$ has no optimal solution even though it is neither infeasible nor unbounded.

6.4 An Example Illustrating Geometry of the Simplex Method

We give an example to illustrate how the simplex method works. Consider the following linear program.

$$\max_{x_1,x_2} x_1 + 2x_2$$
$$\text{s.t.}$$
$$x_1 + x_2 \leq 4$$
$$x_2 \leq 2$$
$$x_1, x_2 \geq 0.$$

The feasible region for this LP is shown in Figure 3. Clearly, no first phase is required here since the origin is a feasible solution. Hence, the first dictionary looks as follows (x_3 and x_4 are the slack variables).

$$x_3 = 4 - x_1 - x_2$$
$$x_4 = 2 - x_2$$

$$z = x_1 + 2x_2.$$

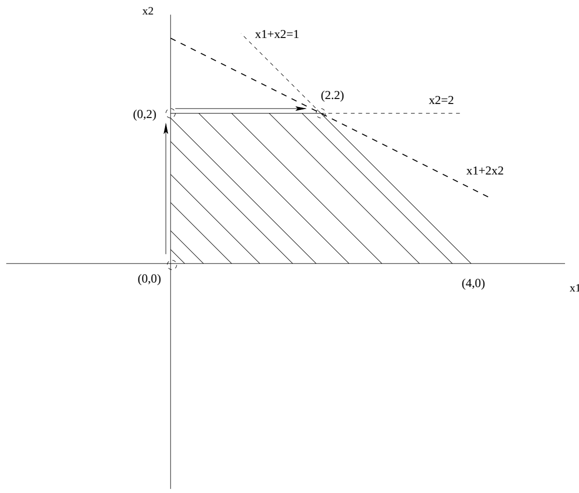

Figure 3: Illustrating the Simplex Method

Note that the feasible dictionary gives the solution $x_3 = 4$ and $x_4 = 2$. It shows the amount of slack in the two constraints. The two constraints $x_1 \geq 0$ and $x_2 \geq 0$ are tight. This describes the point $(0, 0)$.

Let us choose x_2 as the entering variable. In that case, the binding constraint is $x_4 = 2 - x_2$. So, x_4 is the leaving variable. Hence, $x_4 = 0$. This means the constraint $x_2 \leq 2$ will now be tight (along with $x_1 \geq 0$). This describes the point $(0, 2)$.

Finally, x_1 is the entering variable, in which case the constraint corresponding to x_3 became tight (along with the constraint corresponding to x_4). This describes the point $(2, 2)$, which is the optimal solution according to the simplex method.

Hence, we go from one *corner* point to the other in the simplex method as shown in Figure 3.

7 Polyhedra and Polytopes

Polyhedra are special classes of closed convex sets. We have already shown (in assignments) that a closed convex set is characterized by intersection of (possibly infinite) half-spaces. If it is the intersection of a finite number of half-spaces, then it is called a polyhedron.

Definition **1** *A set $P \subseteq \mathbb{R}^n$ is called a* **polyhedron** *if there exists a $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$ such that $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.*

A polytope is the convex hull of a finite set of points.

Definition **2** *A set $P \subset \mathbb{R}^n$ is called a* **polytope** *if there exists finite number of vectors $x^1, \ldots, x^t \in \mathbb{R}^n$ such that $P = H(x^1, \ldots, x^t)$.*

Definition 3 *Let P be a convex set. A point $z \in P$ is called an* **extreme point** *of P if P cannot be expressed as a convex combination of two other points in P, i.e., there do not exist $x, y \in P \setminus \{z\}$ and $0 < \lambda < 1$ such that $z = \lambda x + (1 - \lambda)y$.*

Figure 4 shows two polyhedra, out of which the one on the right is a polytope. It also shows some extreme points of these polyhedra.

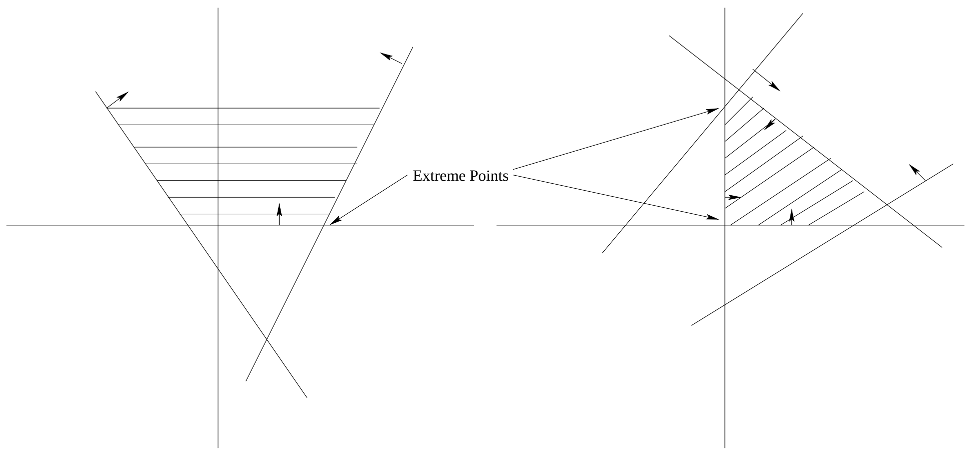

Figure 4: A polyhedron, a polytope, and extreme points

We prove a fundamental result characterizing the extreme points of a polyhedron. For this, we use the following notation. Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polyhedron and $z \in P$. Then, A_z denotes the submatrix of A for which $a_i z = b_i$ for every row a_i of A.

As an example consider $P = \{(x_1, x_2) : x_1 + 2x_2 \leq 2, -x_1 \leq 0, -x_2 \leq 0\}$. If we let $z = (0, 1)$, then A_z corresponds to a matrix with rows $(1, 2)$ and $(-1, 0)$.

Now, we remind ourselves of some basic definitions of linear algebra.

Definition 4 *The* **rank** *of a finite set S of vectors, denoted as $r(S)$, is the cardinality of the largest subset of linearly independent vectors in S.*

If $S = \{(1, 2), (-2, 4)\}$, then $r(S) = 1$. If $S = \{(0, 1, 0), (-2, 2, 0), (-2, 3, 0)\}$, then $r(S) = 2$.

Let A be a $m \times n$ matrix. Then the rank of row vectors of A and the rank of column vectors of A are same. So, for a matrix, we can talk about rank of that matrix. We denote rank of matrix A as $r(A)$.

Theorem 5 *Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polyhedron and $z \in P$. Then z is an extreme point of P if and only if $r(A_z) = n$.*

Proof: Suppose z is an extreme point of P. Assume for contradiction $r(A_z) < n$. This means there exists a vector $x \in \mathbb{R}^n$ and $x \neq 0$ such that $A_z x = 0$. By definition, for all rows a_i not in A_z we have $a_i z < b_i$. This means there exists a $\delta > 0$ such that for every a_i not in A_z we have

$$a_i(z + \delta x) \leq b_i \text{ and } a_i(z - \delta x) \leq b_i.$$

To see why this is true, consider the a row a_i. Suppose $a_i x \leq 0$. Then $\delta a_i x \leq 0$. This means, $a_i z + \delta a_i x < b_i$ since $a_i z < b_i$. Also, since δ can be chosen arbitrarily small, $a_i z - \delta a_i x \leq b_i$. Analogously, if $a_i x \geq 0$, we will have $a_i z + \delta a_i x \leq b_i$ and $a_i z - \delta a_i x < b_i$.

Since $A_z x = 0$ and $Az \leq b$, we get $A(z + \delta x) \leq b$ and $A(z - \delta x) \leq b$. Hence, $z + \delta x$ and $z - \delta x$ belong to P. Since z is a convex combination of these two points, z cannot be an extreme point. This is a contradiction.

Suppose $r(A_z) = n$. Assume for contradiction z is not an extreme point of P. Then there exists $x, y \in P$ with $z \neq x \neq y$ and $0 < \lambda < 1$ such that $z = \lambda x + (1 - \lambda)y$. Then for every row a_i in A_z we can write $a_i x \leq b_i = a_i z = a_i(\lambda x + (1 - \lambda)y)$. Rearranging, we get $a_i(x - y) \leq 0$. Similarly, $a_i y \leq b_i = a_i z = a_i(\lambda x + (1 - \lambda)y)$. This gives us $a_i(x - y) \geq 0$. Hence, $a_i(x - y) = 0$. This implies that $A_z(x - y) = 0$. But $x \neq y$ implies that $r(A_z) \neq n$, which is a contradiction. ■

Remark: This theorem implies that a polyhedron has only a finite number of extreme points. This follows from the fact that there can be only finite number of subrows of A.

Remark: Also, if the number of linearly independent rows (constraints) of A is less than n, then rank of every submatrix of A will be less than n. In that case, the polyhedron has no extreme points - in two dimension, if the constraints are all parallel lines then the rank of any submatrix is one, and clearly we cannot have any extreme point. Hence, if z is an extreme point of P, then we should have more constraints than variables.

Remark: Suppose z and z' are two distinct extreme points of a polyhedron. Then A_z and $A_{z'}$ are distinct. Else, we will have $A_z(z - z') = 0$, and $z - z' \neq 0$ will imply that $r(A_z) < n$.

The result can be used to prove the following theorem, which we state without proving.

Theorem **6** *Let P be a bounded polyhedron with extreme points $(x^1, \ldots, x^t)$. Then $P = H(x^1, \ldots, x^t)$, i.e., every bounded polyhedron is a polytope. Moreover, every polytope is a bounded polyhedron.*

8 Extreme Points and Simplex Method

We will discuss a fundamental property of simplex dictionary.

Theorem 7 *A feasible solution described by a feasible dictionary (i.e., a basic feasible solution) of a linear program* $\max cx$ *subject to* $x \in \{x \in \mathbb{R}_n : Ax \leq b\}$ *is an extreme point of* $\{x \in \mathbb{R}_n : Ax \leq b\}$.

Proof: Let z be a feasible solution of max cx s.t. $Ax \leq b$ ($x \geq 0$ is folded into the constraints) described by a feasible dictionary. An implication of this is exactly n non-basic variables have value zero in the dictionary.

Suppose $r(A_z) < n$. Then, we know that (from earlier proof) that there exists $\delta > 0$ and $z' \neq 0$ with $A_z z' = 0$ such that $z + \delta z'$ is a solution to the linear program. But $(z + \delta z')$ is also a solution to $A_z x = b_z$, where b_z is the part of b corresponding to rows in A_z. Hence, they must be describing the same solution (since it refers to the same set of non-basic variables set to zero). This implies that $z' = 0$, a contradiction. Hence, $r(A_z) = n$. From our characterization of extreme points of polyhedron, z is an extreme point of the polyhedron $Ax \leq b$. ■

9 A Matrix Description of Dictionary

The objective of this section is to write the dictionary in matrix form. Consider the following dictionary.

$$
\begin{aligned}
x_1 &= 54 - 0.5x_2 - 0.5x_4 - 0.5x_5 + 0.5x_6 \\
x_3 &= 63 - 0.5x_2 - 0.5x_4 + 0.5x_5 - 1.5x_6 \\
x_7 &= 15 + 0.5x_2 - 0.5x_4 + 0.5x_5 + 2.5x_6
\end{aligned}
$$

$$
z = 1782 - 2.5x_2 + 1.5x_4 - 3.5x_5 - 8.5x_6.
$$

The dictionary arises after two iterations of simplex method to the following linear program:

$$
\begin{aligned}
\max 19x_1 + 13x_2 + 12x_3 + 17x_4 & \\
\text{s.t.} & \\
3x_1 + 2x_2 + x_3 + 2x_4 &\leq 225 \\
x_1 + x_2 + x_3 + x_4 &\leq 117 \\
4x_1 + 3x_2 + 3x_3 + 4x_4 &\leq 420 \\
x_1, x_2, x_3, x_4 &\geq 0.
\end{aligned}
\qquad (\mathbf{EX})
$$

The slack variables are x_5, x_6, x_7, and they convert the inequalities to equations in the dictionary.

$$
\begin{aligned}
3x_1 + 2x_2 + x_3 + 2x_4 + x_5 &= 225 \\
x_1 + x_2 + x_3 + x_4 + x_6 &= 117 \\
4x_1 + 3x_2 + 3x_3 + 4x_4 + x_7 &= 420
\end{aligned}
\tag{9}
$$

The given dictionary is obtained by solving for x_1, x_3, and x_7 from these equations. In matrix terms the solution may be described very compactly. First, we record the system as $Ax = b$, where

$$
A = \begin{bmatrix} 3 & 2 & 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 4 & 3 & 3 & 4 & 0 & 0 & 1 \end{bmatrix}
$$

$$
b = \begin{bmatrix} 225 \\ 117 \\ 420 \end{bmatrix}
$$

$$
x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix}
$$

To write the system in terms of basic variables x_1, x_3, x_7, we rewrite $Ax = b$ as $A_B x_B + A_N x_N$, where

$$
A_B = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix}
$$

$$
A_N = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 3 & 4 & 0 & 0 \end{bmatrix}
$$

$$
x_B = \begin{bmatrix} x_1 \\ x_3 \\ x_7 \end{bmatrix}
$$

$$x_N = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

So, now we can write the system of equations as

$$A_B x_B = b - A_N x_N$$

If the square matrix A_B is non-singular, we can multiply both sides of the last equation by A_B^{-1} on left to get

$$x_B = A_B^{-1} b - A_B^{-1} A_N x_N$$

This is a compact record of the equations in the given dictionary. To write the objective function in matrix terms, we write it as cx with $c = [19, 13, 12, 17, 0, 0, 0]$, or more precisely as $c_B x_B + c_N x_N$, where $c_B = [19, 12, 0]$ and $c_N = [13, 17, 0, 0]$. Substituting for x_B we get

$$z = c_B(A_B^{-1} b - A_B^{-1} A_N x_N) + c_N x_N = c_B A_B^{-1} b + (c_N - c_B A_B^{-1} A_N) x_N.$$

The given dictionary can now be recorded quite compactly as

$$x_B = A_B^{-1} b - A_B^{-1} A_N x_N$$

$$z = c_B(A_B^{-1} b - A_B^{-1} A_N x_N) + c_N x_N = c_B A_B^{-1} b + (c_N - c_B A_B^{-1} A_N) x_N.$$

The only thing that we have not proved so far is that A_B is non-singular. This is equivalent to proving that the system of equations $A_B x_B = b$ has a unique solution. We already know that there is a solution: let x^* be the solution corresponding to the dictionary with B as the set of basic variables, then $Ax^* = b$ or $A_B x_B^* + A_N x_N^* = b$ or $A_B x_B^* = b$ (since x_N^* is zero). Suppose there is another solution $\bar{x}_B$. Create $\bar{x}$ by setting $\bar{x}_N$ to zero. Since $A_B \bar{x}_B = b$ and $\bar{x}_N = 0$, we get that $A_B \bar{x}_B + A_N \bar{x}_N = A\bar{x} = b$. So, $\bar{x}$ satisfies the original system of equations, and hence should satisfy any dictionary generated in the simplex method. But $\bar{x}_N = 0$ implies that $\bar{x} = x^*$.

10 Duality

Duality is probably the most used concept of linear programming in both theory and practice. The central motivation to look for a dual is the following: **How do we find bounds on the objective function of a linear program without solving it completely**? To understand further, let us look at the following example.

$$
\begin{aligned}
Z = \max 4x_1 + x_2 + 5x_3 + 3x_4 \\
\text{s.t.} \\
x_1 - x_2 - x_3 + 3x_4 &\le 1 \\
5x_1 + x_2 + 3x_3 + 8x_4 &\le 55 \\
-x_1 + 2x_2 + 3x_3 - 5x_4 &\le 3 \\
x_1, x_2, x_3, x_4 &\ge 0.
\end{aligned}
$$

Rather than solving this LP, let us try to find bounds on the optimal value z^* of this LP. For example, $(0, 0, 0, 0)$ is a feasible solution. Hence, $z^* \ge 0$. Another feasible solution is $(0, 0, 1, 0)$ which gives $z^* \ge 5$. Another feasible solution is $(3, 0, 2, 0)$ which gives $z^* \ge 22$. But there is no systematic way in which we were looking for the estimate - it was purely guess work. Duality provides one systematic way of getting this estimate.

Let us multiply the second constraint by $\frac{5}{3}$, which gives us

$$\frac{25}{3}x_1 + \frac{5}{3}x_2 + 5x_3 + 40x_4 \le \frac{275}{3}.$$

But notice that

$$4x_1 + x_2 + 5x_3 + 3x_4 \le \frac{25}{3}x_1 + \frac{5}{3}x_2 + 5x_3 + 40x_4 \le \frac{275}{3}.$$

Hence $z^* \le \frac{275}{3}$. With a little thinking, we can improve this bound further. In particular, add the second and third constraints to get

$$4x_1 + 3x_2 + 6x_3 + 3x_4 \le 58.$$

Using the same logic as before, we get $z^* \le 58$. Here, we are constructing a series of upper bounds for the objective function value, while earlier we were constructing a series of lower bounds.

Formally, we construct linear combinations of the inequalities. We multiply j^{th} constraint by y_j, and add them all up. The resulting inequality reads

$$(y_1 + 5y_2 - y_3)x_1 + (-y_1 + y_2 + 2y_3)x_2 + (-y_1 + 3y_2 + 3y_3)x_3 + (3y_1 + 8y_2 - 5y_3)x_4 \le y_1 + 55y_2 + 3y_3. \quad (10)$$

Of course, each of these multipliers must be non-negative. Next, we want to use the LHS of Equation (10) as an upper bound on $4x_1 + x_2 + 5x_3 + 3x_4$. This can be justified only if in (10), the coefficient of each x_i is at least as big as the corresponding coefficient in the objective function. More explicitly, we want

$$\begin{aligned}
y_1 + 5y_2 - y_3 &\geq 4 \\
-y_1 + y_2 + 2y_3 &\geq 1 \\
-y_1 + 3y_2 + 3y_3 &\geq 5 \\
3y_1 + 8y_2 - 5y_3 &\geq 3.
\end{aligned}$$

If the multipliers are non-negative (note here that if the constraints are equalities, then we do not need the multipliers to be non-negative - they can be free) and if they satisfy these inequalities, then we can get an upper bound on the objective function, i.e., for every feasible solution (x_1, x_2, x_3, x_4) of the original problem and every feasible solution (y_1, y_2, y_3) of the previous set of inequalities, we have

$$4x_1 + x_2 + 5x_3 + 3x_4 \leq y_1 + 55y_2 + 3y_3.$$

Further optimal solution z^* of the original LP satisfies

$$z^* \leq y_1 + 55y_2 + 3y_3.$$

Of course we want this bound to be as close to optimal as possible. This can be done by minimizing $y_1 + 55y_2 + 3y_3$. So, we are led to another LP problem that gives us an upper bound of the original problem.

$$\begin{aligned}
\min y_1 + 55y_2 + 3y_3 & \\
\text{s.t.} & \\
y_1 + 5y_2 - y_3 &\geq 4 \\
-y_1 + y_2 + 2y_3 &\geq 1 \\
-y_1 + 3y_2 + 3y_3 &\geq 5 \\
3y_1 + 8y_2 - 5y_3 &\geq 3 \\
y_1, y_2, y_3 &\geq 0.
\end{aligned}$$

10.1 Writing Down the Dual

From our discussion of the example, it is clear how to write the dual of an original problem. In general, the **dual problem** of

$$\begin{aligned} \max \sum_{j=1}^{n} c_j x_j & \\ \text{s.t.} & \\ \sum_{j=1}^{n} a_{ij} x_j \leq b_i & \qquad \forall\, i \in \{1, \ldots, m\} \\ x_j \geq 0 & \qquad \forall\, j \in \{1, \ldots, n\}. \end{aligned} \tag{\textbf{P}}$$

is defined as the problem

$$\begin{aligned} \min \sum_{i=1}^{m} b_i y_i & \\ \text{s.t.} & \\ \sum_{i=1}^{m} a_{ij} y_i \geq c_j & \qquad \forall\, j \in \{1, \ldots, n\} \\ y_i \geq 0 & \qquad \forall\, i \in \{1, \ldots, m\}. \end{aligned} \tag{\textbf{D}}$$

Notice the following things in the dual (**D**):

1. For every constraint of (**P**), we have a variable in (**D**).
2. Further, for every variable of (**P**), we have a constraint in (**D**).
3. The coefficient in the objective function of (**P**) appears on the RHS of constraints in (**D**) and the RHS of constraints in (**P**) appear as coefficients of objective function in (**D**).

As an exercise, verify that dual of (**D**) is (**P**).

Lemma 1 **(Weak Duality)** *Let $(x_1, \ldots, x_n)$ be a feasible solution of (**P**) and $(y_1, \ldots, y_m)$ be a feasible solution of (**D**). Then*

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i.$$

Proof:

$$\begin{aligned}\sum_{j=1}^{n} c_j x_j &\le \sum_{j=1}^{n} (\sum_{i=1}^{m} a_{ij} y_i) x_j \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} x_j) y_i \\ &\le \sum_{i=1}^{m} b_i y_i.\end{aligned}$$

■

Lemma 1 is useful since if we find feasible solutions of (**P**) and (**D**) at which their objective functions are equal, then we can conclude that they are optimal solutions. Indeed, Lemma 1 implies that if $(x_1^*, \ldots, x_n^*)$ is an optimal solution of (**P**) and $(y_1^*, \ldots, y_m^*)$ is an optimal solution of (**D**) such that $\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$, then for every feasible solution $(x_1, \ldots, x_n)$ of (**P**) and for every feasible solution $(y_1, \ldots, y_n)$, we can write

$$\sum_{j=1}^{n} c_j x_j \le \sum_{i=1}^{m} b_i y_i^* = \sum_{j=1}^{n} c_j x_j^* \le \sum_{i=1}^{m} b_i y_i.$$

11 The Duality Theorem

The explicit version of the theorem is due to Gale, but it is supposed to have originated from conversations between Dantzig and von Neumann in the fall of 1947.

Theorem 8 **(The Duality Theorem - Strong Duality)** *Let $(x_1^*, \ldots, x_n^*)$ be a feasible solution of (**P**) and $(y_1^*, \ldots, y_m^*)$ be a feasible solution of (**D**). $(x_1^*, \ldots, x_n^*)$ is an optimal solution of (**P**) and $(y_1^*, \ldots, y_m^*)$ is an optimal solution of (**D**) if and only if*

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*. \qquad \textbf{(SD)}$$

Before presenting the proof, let us illustrate the crucial point of the theorem: the optimal solution of (**D**) can be read off the z-row of the final dictionary for (**P**). For the example, the final dictionary of (**P**) is

$$\begin{aligned}x_2 &= 14 - 2x_1 - 4x_3 - 5x_5 - 3x_7 \\ x_4 &= 5 - x_1 - x_3 - 2x_5 - x_7 \\ x_6 &= 1 + 5x_1 + 9x_3 + 21x_5 + 11x_7 \\ \\ z &= 29 - x_1 - 2x_3 - 11x_5 - 6x_7.\end{aligned}$$

Note that the slack variables x_5, x_6, x_7 can be matched with the dual variables y_1, y_2, y_3 in a natural way. In the z-row of the dictionary, the coefficients of these slack variables are $(-11, 0, -6)$. As it turns out the optimal dual solution is obtained by reversing the signs of these coefficients, i.e., $(11, 0, 6)$. The proof of the duality theorem works on this logic.

Proof: Suppose Equation (**SD**) holds. Assume for contradiction that $(x'_1, \ldots, x'_n) \neq (x^*_1, \ldots, x^*_n)$ is an optimal solution of (**P**). Hence, $\sum_{j=1}^n c_j x'_j > \sum_{j=1}^n c_j x^*_j = \sum_{i=1}^m b_i y^*_i$. By Lemma 1, this is a contradiction. Hence, $(x^*_1, \ldots, x^*_n)$ is an optimal solution of (**P**). A similar argument shows that $(y^*_1, \ldots, y^*_m)$ is an optimal solution of (**D**).

For the other side of the proof, we assume that $(x^*_1, \ldots, x^*_n)$ is an optimal solution of (**P**), and find a feasible solution $(y^*_1, \ldots, y^*_m)$ that satisfies the claim in the theorem, and such a solution will be optimal by the first part of the proof. In order to find that feasible solution, we solve (**P**) using the simplex method using slack variables

$$x_{n+i} = b_i - \sum_{j=1}^n a_{ij} x_j \qquad \forall\, i \in \{1, \ldots, m\}.$$

Since an optimal solution exists, the simplex method finds it, and the final row of the final dictionary reads

$$z = z^* + \sum_{k=1}^{m+n} \bar{c}_k x_k,$$

where $\bar{c}_k = 0$ whenever x_k is a basic variable and $\bar{c}_k \leq 0$ otherwise. In addition, z^* is the optimal value of (**P**), hence

$$z^* = \sum_{j=1}^n c_j x^*_j.$$

We claim that

$$y^*_i = -\bar{c}_{n+i} \qquad \forall\, i \in \{1, \ldots, m\}$$

is a feasible solution of (**D**) satisfying the claim of our theorem. Substituting $z = \sum_{j=1}^n c_j x_j$ and substituting for slack variables, we get

$$\sum_{j=1}^n c_j x_j = z^* + \sum_{j=1}^n \bar{c}_j x_j - \sum_{i=1}^m y^*_i \Big(b_i - \sum_{j=1}^n a_{ij} x_j \Big),$$

which may be rewritten as

$$\sum_{j=1}^n c_j x_j = \Big(z^* - \sum_{i=1}^m b_i y^*_i \Big) + \sum_{j=1}^n \Big(\bar{c}_j + \sum_{i=1}^m a_{ij} y^*_i \Big) x_j.$$

This equation is obtained from algebraic manipulations of the definitions of slack variables and objective function, and must hold for all values of $x_1, \ldots, x_n$. Hence, we have

$$z^* = \sum_{i=1}^{m} b_i y_i^*; \quad c_j = \bar{c}_j + \sum_{i=1}^{m} a_{ij} y_i^* \qquad \forall\, j \in \{1, \ldots, n\}.$$

Since $\bar{c}_k \leq 0$ for every $k \in \{1, \ldots, m+n\}$ we get

$$\sum_{i=1}^{m} a_{ij} y_i^* \geq c_j \qquad \forall\, j \in \{1, \ldots, n\}$$
$$y_i^* \geq 0 \qquad \forall\, i \in \{1, \ldots, m\}.$$

This shows that $(y_1^*, \ldots, y_m^*)$ is a feasible solution of (**D**). Finally, $z^* = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$. ■

11.1 Relating the Primal and Dual Problems

First, notice that dual of a dual problem is the original primal problem, i.e., dual of (**D**) is (**P**). A nice corollary to this observation is that **a linear program has an optimal solution if and only if its dual has an optimal solution**.

By Lemma 1, if the primal problem is unbounded, then the dual problem is infeasible. To see this, assume for contradiction that the dual is feasible when the primal is unbounded. This means, the feasible dual solution provides an upper bound on the optimal value of the primal problem. This is a contradiction since the primal problem is unbounded. By the same argument, if the dual is unbounded, then the primal is infeasible.

This also shows that if the primal and dual are both feasible, then they both have optimal solutions, i.e., none of them is unbounded. However, both the primal and the dual can be infeasible. For example,

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

and its dual are infeasible. We summarize these observations in the Table 11.1.

Duality has important practical applications. In certain cases, it may be better to solve the dual problem than the primal problem, and then read the primal solution from the last row of the final dictionary. For example, a primal problem with 100 constraints and two

		Dual		
		Optimal	Infeasible	Unbounded
Primal	Optimal	$\surd$	$\times$	$\times$
	Infeasible	$\times$	$\surd$	$\surd$
	Unbounded	$\times$	$\surd$	$\times$

Table 1: Primal-dual combinations possibilities

variables will have two constraints in the dual. Typically, the number of simplex method iterations are insensitive to the number of variables and proportional to the number of rows/constraints. Hence, we may be better off solving the dual in this case.

11.2 Farkas Lemma and Duality Theory

Here, we prove the Farkas Lemma using duality theory.

Theorem **9** *Let A be a $m \times n$ matrix and b be a $m \times 1$ matrix. Suppose $F = \{x \in \mathbb{R}^n_+ : Ax = b\}$ and $G = \{y \in \mathbb{R}^m : yb < 0, yA \geq 0\}$. The set F is non-empty if and only if the set G is empty.*

Proof: Consider the linear program $\max_{x \in \mathbb{R}^n} 0 \cdot x$ subject to $x \in F$. Denote this linear program as (**P**). The dual of this linear program is $\min_{y \in \mathbb{R}^m} yb$ subject to $yA \geq 0$. Denote this linear program as (**D**).

Now, suppose F is non-empty. Then, (**P**) has an optimal value, equal to zero. By strong duality, the optimal value of (**D**) is zero. Hence, for any feasible solution y of (**D**), we have $yb \geq 0$. This implies that G is empty.

Suppose G is empty. Hence, for every feasible solution y of (**D**), $yb \geq 0$. This implies that (**D**) is not unbounded. Since, $y = 0$ is a feasible solution of (**D**), it is an optimal solution. This implies that (**P**) has an optimal solution. Hence, F is non-empty. ■

11.3 Complementary Slackness

The question we ask in this section is given a feasible solution of the primal problem (**P**) and a feasible solution of the dual problem (**D**), are there conditions under which these solutions are optimal. The following theorem answers this question.

Theorem **10 (Complementary Slackness)** *Let $(x^*_1, \ldots, x^*_n)$ and $(y^*_1, \ldots, y^*_m)$ be feasible solutions of (**P**) and (**D**) respectively. $(x^*_1, \ldots, x^*_n)$ is an optimal solution of (**P**) and*

$(y_1^*, \ldots, y_m^*)$ *is an optimal solution of (**D**) if and only if*

$$\Big[\sum_{i=1}^{m} a_{ij} y_i^* - c_j\Big] x_j^* = 0 \qquad \forall\, j \in \{1, \ldots, n\} \tag{\textbf{CS-1}}$$

$$\Big[b_i - \sum_{j=1}^{n} a_{ij} x_j^*\Big] y_i^* = 0 \qquad \forall\, i \in \{1, \ldots, m\}. \tag{\textbf{CS-2}}$$

Proof: Denote $y^* := (y_1^*, \ldots, y_m^*)$ and $x^* := (x_1^*, \ldots, x_n^*)$. Since x^* and y^* are feasible, we immediately get

$$\Big[\sum_{i=1}^{m} a_{ij} y_i^* - c_j\Big] x_j^* \geq 0 \qquad \forall\, j \in \{1, \ldots, n\}$$

$$\Big[b_i - \sum_{j=1}^{n} a_{ij} x_j^*\Big] y_i^* \geq 0 \qquad \forall\, i \in \{1, \ldots, m\}.$$

Now suppose that x^* and y^* are optimal. Assume for contradiction that one of the inequalities in the first set of constraints is not tight. In that case, adding up all the constraints in the first set will give us

$$\begin{aligned} 0 &< \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} y_i^* x_j^* - \sum_{j=1}^{n} c_j x_j^* \\ &\leq \sum_{i=1}^{m} b_i y_i^* - \sum_{j=1}^{n} c_j x_j^* \qquad \text{Because } x^* \text{ is feasible to } (\mathbf{P}) \\ &= 0 \qquad \text{Because } x^* \text{ and } y^* \text{ are optimal solutions and Theorem 8} \end{aligned}$$

This gives us a contradiction. A similar proof shows (**CS-2**) holds.

Now suppose (**CS-1**) and (**CS-2**) holds. Then, add all the equations in (**CS-1**) to get

$$\sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_j^* y_i^* = \sum_{j=1}^{n} c_j x_j^*.$$

Similarly, add all the equations in (**CS-2**) to get

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j^* y_i^* = \sum_{i=1}^{m} b_i y_i^*.$$

This gives us $\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$. By Theorem 8, x^* is an optimal solution of (**P**) and y^* is an optimal solution of (**D**). ∎

Theorem 10 gives us a certificate of proving optimality. The idea is clear from our earlier interpretation of optimal dual variable values as negative of coefficients of slack variables in the objective function row of the final simplex dictionary. If a dual variable has positive optimal value, this implies that coefficient of slack variable is negative. This further implies that the slack variable is non-basic in the final simplex dictionary. Hence, its value is zero in the primal optimal solution. This implies that the corresponding constraint is binding in the optimal solution. Similarly, if some constraint is non-binding, then the corresponding slack variable has positive value in the optimal solution. This implies that the slack variable is basic in the final simplex dictionary, which further implies that its coefficient is zero in the objective function row. Hence, the corresponding dual solution has zero value.

Consider the following example.

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

Consider an optimal solution (x_1^*, x_2^*) of this LP, and assume that $x_1^* < 1$. Now, let (y_1^*, y_2^*) be a dual optimal solution. This should provide a bound to $x_1^* + x_2^* \leq (y_1^* + 2y_2^*)x_1^* + (3y_2^*)x_2^* < y_1^* + 6y_2^*$ (since $x_1^* < 1$). By strong duality, this is not possible unless we set $y_1^* = 0$. This is exactly the idea behind complementary slackness.

11.4 Interpreting the Dual

In optimization, the dual variables are often called *Lagrange multipliers*. In economics, the dual variables are interpreted as prices of *resources*, where resources are constraints. Consider the following example.

Suppose there are n products that a firm can manufacture. Each product requires the use of m resources. To manufacture product j, the firm needs a_{ij} amount of resource i (naturally, it makes sense to assume $a_{ij} \geq 0$ here, but one need not). The amount of resource i available is b_i. The market price of product j is c_j (again, both b_i and c_j can be assumed to be non-negative in this story). The firm needs to decide how much to manufacture of each product to maximize his revenue subject to resource constraints. The problem can be formulated as a linear program - formulation (**PE**).

$$\begin{aligned} \max \sum_{j=1}^{n} c_j x_j & \\ \text{s.t.} & \\ \sum_{j=1}^{n} a_{ij} x_j \leq b_i & \qquad \forall\, i \in \{1, \ldots, m\} \\ x_j \geq 0 & \qquad \forall\, j \in \{1, \ldots, n\}. \end{aligned} \tag{PE}$$

Now, suppose an investor wants to buy this firm's resources. He proposes a price for every resource. In particular, he proposes a price of y_i for resource i. Moreover, the investor promises that he will set his prices high enough such that the firm gets at least as much selling the resources as he would turning the resources into products and then selling them at price vector c. Hence, the following constraints must hold

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j \qquad \forall\, j \in \{1, \ldots, n\}.$$

Another way to think of these constraints is that if the constraint for product j does not hold, then the firm will not sell resources required to produce product j since by selling them in the market he gets a per unit price of c_j which will be higher than $\sum_{i=1}^{m} a_{ij} y_i$ - the per unit profit from selling.

Of course, all prices offered by the investor must be non-negative. The investor must now try to minimize the price he needs to pay to buy the resources. Hence, he should

$$\min \sum_{i=1}^{m} b_i y_i.$$

In particular, the investor will solve the following linear programming problem (**DE**).

$$\begin{aligned} \min \sum_{i=1}^{m} b_i y_i & \\ \text{s.t.} & \\ \sum_{i=1}^{m} a_{ij} y_i \geq c_j & \qquad \forall\, j \in \{1, \ldots, n\} \\ y_i \geq 0 & \qquad \forall\, i \in \{1, \ldots, m\}. \end{aligned} \tag{DE}$$

Strong duality theorem says that the optimal value of investor's cost equals the optimal value of firm's profit. The dual variables are thus prices for the resources in the primal problem.

Now, let us interpret the complementarity slackness conditions here. Suppose the firm has an optimal production plan in place. In this plan, he does not use resource, say, i completely, i.e., the constraint corresponding to i is not binding. In that case, can sell the resources at unit price of zero. But if the price offered is strictly positive for resource i, then the production plan must be using all the resources. Intuitively, the demand for the input i is high, leading to positive prices.