Solving Linear Programs #2

1.0 Manipulations of the simplex method

In setting up the simplex method, we introduced *slack variables* which allowed us to write the LP in *equality form*. We also introduced the idea of a *basic solution*, which is just a corner point solution including the values of the slack variables; a *basic feasible solution* is just a feasible corner point solution including the values of the slack variables.

The problem under consideration, in equality form, is

$$\max F = 3x_1 + 5x_2$$

s.t.

$$x_1 \qquad\quad + x_3 \qquad\qquad\qquad = \quad 4$$

$$\qquad 2x_2 \qquad\quad + x_4 \qquad\quad = \quad 12$$

$$3x_1 \; + 2x_2 \qquad\qquad\quad + x_5 \; = \quad 18 \tag{1}$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0$$

Its feasible region is drawn in Fig. 1.

What we want to do now is illustrate the manipulations taken by the simplex method in solving this problem.
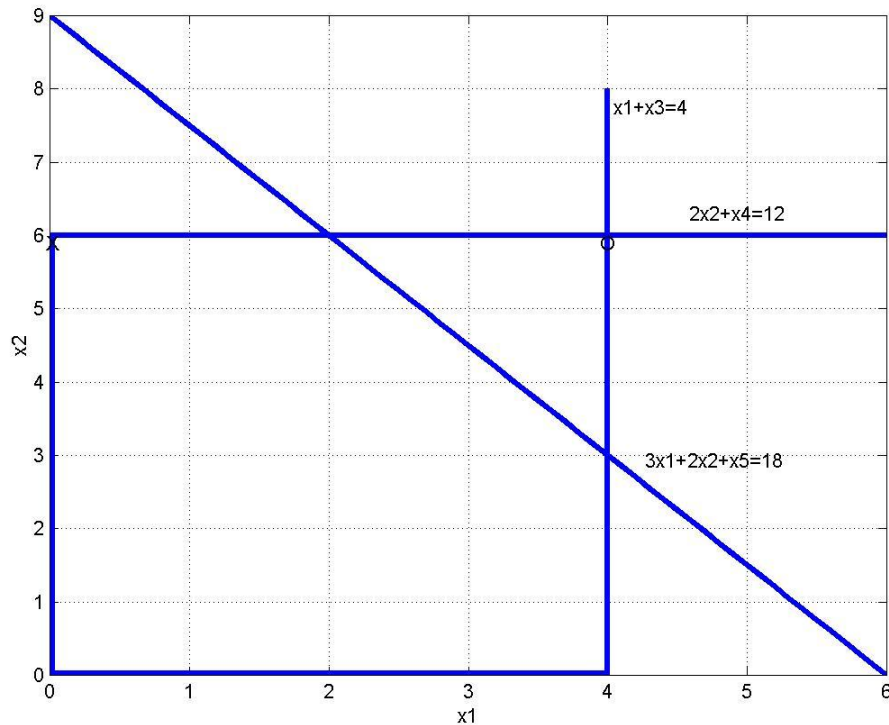
The material of this section is adapted from [1].

Fig. 1: Illustration of feasible region

2.0 Simplex initialization

We begin the simplex method by initializing the problem, i.e., we obtain a basic feasible solution.

But this is actually easy, for this problem, and for all LP problems in equality form. Consider any one of the constraints, which is in the form

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + x_{n+k} = b_k \qquad (2)$$

where the slack variable is x_{n+k}.

Notice that

$$x_1 = 0, \ x_2 = 0,..., \ x_n = 0, \ x_{n+k} = b_k \qquad (3)$$

satisfies (2).

We know all variables must be nonnegative, meaning $x_i \geq 0$, for any i, but (3) satisfies that for the first n variables $x_1, \ x_2, \ ...x_n$.

We also know that the right-hand-sides are all non-negative, meaning $b_k \geq 0$, for any k, and because of that, we know that (3) ensures that $x_{n+k} \geq 0$.

But what about the other constraints? Will (3) violate any of them?

The answer is no, because of our clever provision of slack variables for every constraint. Making all of the first n variables zero is fine, as long as we have the degree of freedom in each equation provided by the slack variable.

Effectively, we are stating that, in terms of the original decision variables, the origin is always a solution once we have transformed the problem into our desired equality form.

And so now let's initialize our example problem.

$$\max F = 3x_1 + 5x_2$$

s.t.

$$
\begin{array}{rcrcl}
x_1 & & + x_3 & & = & 4 \\
& 2x_2 & & + x_4 & = & 12 \\
3x_1 & + 2x_2 & & + x_5 & = & 18
\end{array}
$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0$$

Notice an initial BSF can be directly read off as the RHS when x_1 and x_2 are zero. We will come back to this in Section 3.3 below.

The initial basic feasible solution is

$$x_1 = 0, x_2 = 0, x_3 = 4, x_4 = 12, x_5 = 18$$

At this point, we provide two more definitions that are useful in describing the simplex method.

• *Basic variables*: These are variables that *are not* 0.

• *Non-basic variables*: These are variables that *are* 0.

In our initial basic feasible solution above, the basic variables are x_3, x_4, x_5; non-basic variables are x_1, x_2.

In this problem, there are two decision variables: x_1, x_2. We use n to denote the number of decision variables, and m to denote the number of constraints. So in our example, $n=2$, $m=3$.

Throughout the simplex method, there are always $n+m$ variables comprising the solution: n are non-basic, and m are basic. Each equation has one basic variable, with coefficient of 1. The simplex method manipulates equations so they continue to have this form for reading subsequent basic feasible solutions.

3.0 Simplex iteration

Recall in our geometric interpretation, the simplex method moves from one feasible corner point to an adjacent feasible corner point that is better.

A basic feasible solution is a corner-point solution that is feasible. And so our initialization procedure has provided that we are at a feasible corner point solution. What we desire to do now is to move from it to another feasible corner point solution (another basic feasible solution) that is better (has higher objective, assuming we are maximizing).

We explain the mechanical steps of simplex iteration. We would also like to understand why the steps are as they are. To this end, we extend our understanding in regards to two terms in the context of simplex manipulations.

1. <u>What is a corner point</u>? A corner point (basic solution) is the simultaneous solution of a system of n constraint equations. A feasible corner point (basic feasible solution) is the simultaneous solution of a set of n constraint equations that does not violate any constraint equations.

2. <u>What does adjacency mean</u>? Two corner points (or basic solutions) are said to be adjacent if the constraint equations which define them are the same, with exception of just one. This means they share *n-1* constraint equations, with one constraint equation for one corner point differing from one constraint equation for the other corner point. Adjacent feasible corner points (or adjacent basic feasible solutions) also satisfy this definition and in addition satisfy all constraint equations.

To fix the above ideas, let's examine the corner point solutions of our example. We can identify them from Fig. 1, repeated below for convenience. The feasible corner point solutions are given in Table 1. Notice in this table that consecutive rows contain adjacent corner point solutions; in addition, the first and last rows contain adjacent corner point solutions. One sees in this example that adjacent corner point solutions share one defining constraint equation.
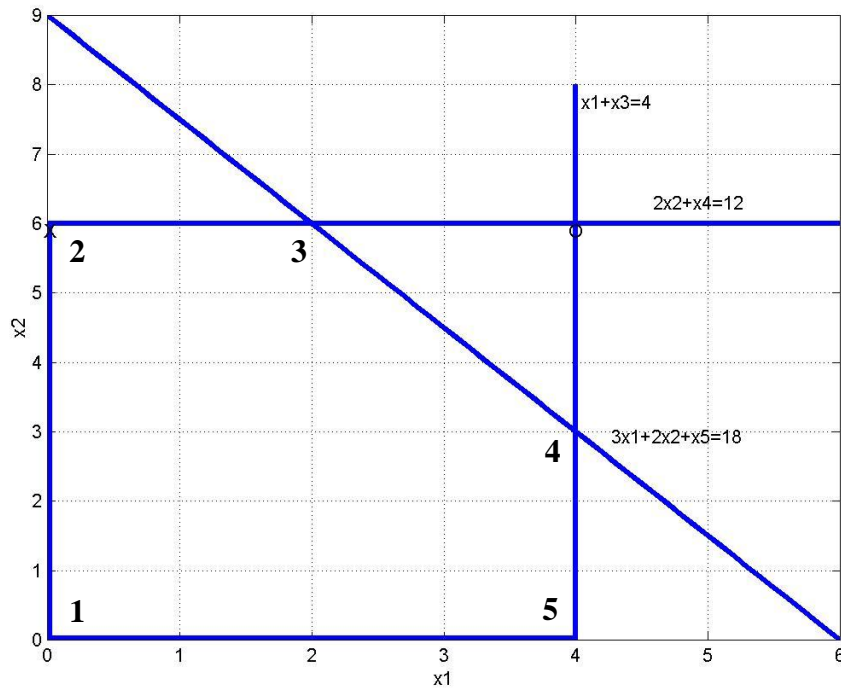
Fig. 1: Illustration of feasible region

Table 1: Feasible corner point solutions & defining constraint equations

Crnr pnt feasible soltns		Defining cnstrnt eqtns
1	(0,0)	$x_1=0$
		$x_2=0$
2	(0,6)	$x_1=0$
		$2x_2=12$
3	(2,6)	$2x_2=12$
		$3x_1+2x_2=18$
4	(4,3)	$3x_1+2x_2=18$
		$x_1=4$
5	(4,0)	$x_1=4$
		$x_2=0$

It is useful at this point to consider manipulations for circulating around the feasible boundary by moving from one corner point to another.

Start at (*0,0*).

1. The $x_2=0$ constraint becomes inactive.
2. Move along the $x_1=0$ constraint until you reach the next corner point, which is (0,6).
3. The $2x_2=12$ constraint becomes active.

1 to 2

1. The $x_1=0$ constraint becomes inactive.
2. Move along the $2x_2=12$ constraint until you reach the next corner point, which is (2,6).
3. The $3x_1+2x_2=18$ constraint becomes active.

2 to 3

1. The $2x_2=12$ constraint becomes inactive.
2. Move along the $3x_1+2x_2=18$ constraint until you reach the next corner point, which is (4,3).
3. The $x_1=4$ constraint becomes active.

3 to 4

1. The $3x_1+2x_2=18$ constraint becomes inactive.
2. Move along the $x_1=4$ constraint until you reach the next corner point, which is (4,0).
3. The $x_2=0$ constraint becomes active.

1. The $x_1=4$ constraint becomes inactive.
2. Move along the $x_2=0$ constraint until you reach the next corner point, which is (0,0).
3. The $x_1=0$ constraint becomes active.

5 to 1

8

Now let's examine the basic feasible solutions, as given in Table 2. Notice here that the defining constraint equations include the slack variables.

Table 2: Basic feasible solutions & defining constraint equations

Basic feasible soltns	Defining cnstrnt eqtns
1 $(0,0,4,12,18)$	$x_1=0$
	$x_2=0$
2 $(0,6,4,0,6)$	$x_1=0$
	$2x_2+x_4=12$
3 $(2,6,2,0,0)$	$2x_2+x_4=12$
	$3x_1+2x_2+x_5=18$
4 $(4,3,0,6,0)$	$3x_1+2x_2+x_5=18$
	$x_1+x_3=4$
5 $(4,0,0,12,6)$	$x_1+x_3=4$
	$x_2=0$

One can interpret the values of the slack variables, for a particular corner point, as the weighted distance between that corner point and the slack variable's constraint, where the weights are the coefficients from the constraint equation. It is easy to see this interpretation for the constraint corresponding to $x_1 \leq 4$ (or $x_1+x_3=4$ in equality form), because the coefficient for x_1 is 1. Therefore, the slack variable is exactly the "distance" between the given corner point and this constraint (from Table 2, this would be 4, 4, 2, 0, 0 for points 1, 2, 3, 4, 5, respectively).

One observation that we can make from Table 2 is:
All basic feasible solutions have
exactly $n=2$ variables equal to 0.

Why is this?
The reason it occurs in this example is because any basic feasible solution (any basic solution, for that matter) is the simultaneous solution of a system of $n=2$ constraint equations. There are two kinds of constraint equations that may comprise this pair.

- Nonnegativity constraints on decision variables: If this is one of the constraint equations defining the basic feasible solution, then the corresponding decision variable is 0. In Table 2, we observe there are two nonnegativity constraints defining basic feasible solution #1, and one each defining basic feasible solutions #2 and #5.

- Constraints with slack variables: There are three equations with slack variables. In a resource allocation problem, these are the constraints on the resources. If this is one of the constraint equations defining the basic feasible solution, then the corresponding slack variable is zero. In Table 2, we observe there are two such constraints defining basic feasible solution #3 and #4, and one each defining basic feasible solutions #2 and #5.

This will occur in every LP, except that the number of decision variables n may be different, and the number of equations with slack variables m may be different. But it will always be the case that any basic feasible solution will have n variables 0.

That is, as we have already said, any basic feasible solution will have n non-basic variables.

Another observation that we make from Table 2 is:

> **The way we move from one basic feasible solution to another is by exchanging exactly one zeroed variable with a non-zeroed variable.**

In other words,

> **The way we move from one basic feasible solution to another is by exchanging exactly one non-basic variable with a basic variable.**

When we do this, we say that one variable is a leaving basic variable, i.e., it leaves the set of basic variables or "leaves the basis," (it becomes zero) and the other variable is an entering basic variable, i.e., it enters the set of basic variables or "enters the basis" (it becomes non-zero).
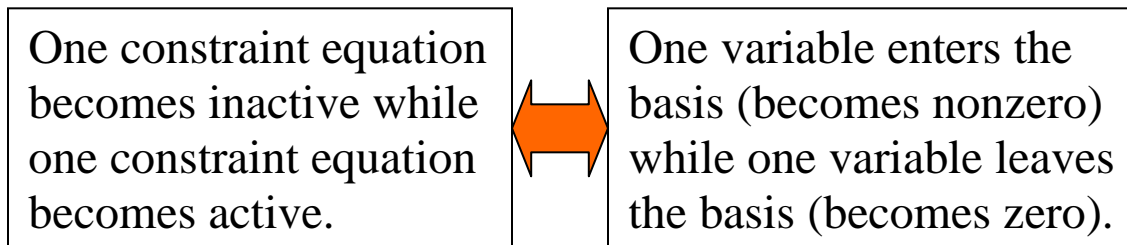
Let's repeat the manipulations for circulating around the feasible boundary by moving from one corner point to another, but this time, let's identify the leaving and entering basic variable. Table 2 is reproduced on this same page for ease of reference.

Start at $(0,0)$.

1. The $x_2=0$ constraint becomes inactive (x_2 enters).

2. Move along the $x_1=0$ constraint until you reach the next corner point, which is $(0,6)$.

3. The $2x_2=12$ constraint becomes active (x_4 leaves).

1. The $x_1=0$ constraint becomes inactive (x_1 enters).

2. Move along the $2x_2=12$ constraint until you reach the next corner point, which is $(2,6)$.

3. The $3x_1+2x_2=18$ constraint becomes active (var x_5 leaves).

1. The $2x_2=12$ constraint becomes inactive (x_3 enters)

2. Move along the $3x_1+2x_2=18$ constraint until you reach the next corner point, which is $(4,3)$.

3. The $x_1=4$ constraint becomes active (x_4 leaves).

1. The $3x_1+2x_2=18$ constraint becomes inactive (x_5 enters).

2. Move along the $x_1=4$ constraint until you reach the next corner point, which is $(4,0)$.

3. The $x_2=0$ constraint becomes active (x_2 leaves).

1. The $x_1=4$ constraint becomes inactive (x_1 enters).

2. Move along the $x_2=0$ constraint until you reach the next corner point, which is $(0,0)$.

3. The $x_1=0$ constraint becomes active (x_1 leaves).

Basic feasible soltns		Defining cnstrnt eqtns
1	$(0,0,4,12,18)$	$x_1=0$
		$x_2=0$
2	$(0,6,4,0,6)$	$x_1=0$
		$2x_2+x_4=12$
3	$(2,6,2,0,0)$	$2x_2+x_4=12$
		$3x_1+2x_2+x_5=18$
4	$(4,3,0,6,0)$	$3x_1+2x_2+x_5=18$
		$x_1+x_3=4$
5	$(4,0,0,12,6)$	$x_1+x_3=4$
		$x_2=0$

The above provides us with a sense of how the simplex method moves around the corner point solutions, where each move is defined by the following two equivalent statements.

One constraint equation becomes inactive while one constraint equation becomes active.	One variable enters the basis (becomes nonzero) while one variable leaves the basis (becomes zero).

Now we would like to gain insight into how this happens algebraically.

3.1 How is the entering basic variable identified?

The candidates for the entering basic variable are the n nonbasic variables. In the first step of our example (solution 1 to solution 2), the candidates are therefore x_1 and x_2. Our criterion is:

Select the one that improves the objective at the highest rate (i.e., the largest amount of objective per unit change in variable).

To decide this, we need to look at the objective function. In our example, our objective is

$$F = 3x_1 + 5x_2 \qquad (4)$$

Increasing either variable, x_1 or x_2, increases the objective, but x_2 increases it the most for a given unit change, since $5 > 3$; so x_2 is the entering variable.

3.2 How is the leaving basic variable identified?

What we intend to do, relative to the current basic feasible solution, is to increase the entering variable from its current value of 0 to some positive value. At the same time, we will hold the other non-basic variables (in this case only x_1) at zero, since we only want to have one variable enter the basis at a time.

In order to **increase** the entering variable while holding the other non-basic variables at zero, one or more of the basic variables must **decrease** towards zero in order to maintain equality in our (equality form) system of equations. Because we impose nonnegativity on all variables, the amount of decrease for any one basic variable is limited by 0. The balance between how much the entering variable may increase without pushing any current basic variable negative is controlled by our m constraints. So the criterion for selecting the leaving variable is:

> **Choose the leaving variable to be the one that hits 0 first as the entering variable is increased, as dictated by one of the m constraint equations.**

In other words, the leaving variable will be the current basic variable whose nonnegativity constraint imposes the smallest upper bound on how much the entering variable can be increased.

In the first step of our example (solution 1 to solution 2), the candidates for the leaving variable are therefore x_3, x_4, and x_5. Repeating below the equations we have,

$$
\begin{array}{rcrcrcrcrcr}
x_1 & & & + x_3 & & & & & & = & 4 \\
& & 2x_2 & & & + x_4 & & & & = & 12 \\
3x_1 & + 2x_2 & & & & & & + x_5 & & = & 18
\end{array}
\tag{5}
$$

we may summarize how each of our $m=3$ constraint equations limits the amount x_2 may be increased, per Table 3.

Table 3: Determination of leaving variable for first step of example, when x_2 is the entering variable

Basic variable	Constraint	Upper bound for x_2
x_3	$x_1+x_3=4$	No limit imposed
x_4	$2x_2+x_4=12$	$x_2=(12-0)/2=6$
x_5	$3x_1+2x_2+x_5=18$	$x_2=(18-3(0)-0)/2=9$

Table 3 tells us the following:
• We can increase x_2 up to 6, but if we go further, x_4 will become negative.
• We can increase x_2 up to 9, but if we go further, x_5 will become negative.
It is clear we cannot increase x_2 beyond 6, and x_4 becomes the leaving variable.

3.3 How is the new basic feasible solution found?

The basic feasible solution could be found from the following things that we know at this point

- x_1 must remain a non-basic variable, therefore $x_1=0$;
- From the conclusion of section 3.2, $x_2=6$, $x_4=0$;
- From the first-row equation of Table 3, together with the information from the above first bullet, $x_1+x_3=4 \rightarrow x_3=4$.
- From the third-row equation of Table 3, together with the information from the first two bullets above, $3x_1+2x_2+x_5=18 \rightarrow x_5=18-3x_1-2x_2=18-3(0)-2(6)=6$.

And so the new basic feasible solution is $(x_1,x_2,x_3,x_4,x_5)=(0,6,4,0,6)$, and this completes iteration 1 of the simplex method.

However, there is a more structured way to do this which requires that we rewrite the objective function according to

$$F = 3x_1 + 5x_2 \rightarrow F - 3x_1 - 5x_2 = 0 \qquad (6)$$

and then write it together with our constraint equations according to:

$$
\begin{array}{llllll}
F & -3x_1 & -5x_2 & & & = & 0 & (7) \\
& x_1 & & +x_3 & & = & 4 & (8) \\
& & 2x_2 & & +x_4 & = & 12 & (9) \\
& 3x_1 & +2x_2 & & +x_5 & = & 18 & (10)
\end{array}
$$

The above equations are set up for our first solution, where x_1 and x_2 are our nonbasic variables (equal to zero) and x_3, x_4, and x_5 are our basic variables (are nonzero). As noted in Section 2.0, we were able to directly read off the initial solution because all non-basic variables, being zero, did not contribute to the left-hand-sides. Therefore $(x_3, x_4, x_5)=(4, 12, 18)$.

In addition, we observe an interesting thing about (7). Again, because x_1 and x_2 are both zero (being non-basic variables), the value of F at the particular solution corresponding to the above form is on the right-hand-side of (7). That is, it is zero, which makes sense because our initial solution is the origin.

Our goal now is to have x_2 enter the basis and x_4 leave the basis so that, at this new solution, our system of equations is in the same convenient form as the above where
• the values of the basic variables at the solution can be directly read off as the right-hand-sides of those equations and
• the value of the objective at that solution can be directly read off as the right-hand-side of the objective equation.

We see what we must do to accomplish this…

• Because x_2 is going to be non-zero, it must show up in only one constraint equation, with a 1 as its coefficient.

• Because x_4 is going to be zero, it must show up in the objective equation.

The way to make the above happen is to use (9) to eliminate x_2 from all other equations. We choose (9) because it is the equation which forced us to choose x_4 as the leaving variable.

"Elimination" is a familiar notion to those of us that have taken linear algebra, or a power system analysis course like EE 456. It can be done effectively by converting our system of equations (7)-(10) into the following matrix relation:

$$
\begin{bmatrix} 1 & -3 & -5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} F \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 4 \\ 12 \\ 18 \end{bmatrix}
\tag{11}
$$

We may perform Gaussian elimination on (11), an algorithm that can be succinctly summarized as follows:

The algorithm is as follows (for an $n \times n$ matrix)

Let i=1. In each repetition below, row i is the *pivot row* and a_{ik} is the *pivot.*

1. Divide row i by a_{ik}.
2. If no rows remain, stop; else go to 3.
3. Eliminate all a_{jk}, j=1,…,n except j=i. This means to make all elements directly above and beneath the pivot equal to 0 by adding an appropriate multiple of the pivot row to each row above or beneath the pivot.
4. i=i+1, go to 2.

We illustrate on the matrix of (11) in what follows.

A first step is to join the matrix with the right-hand-side vector, like this:

$$\begin{bmatrix} 1 & -3 & -5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & \mathbf{2} & 0 & 1 & 0 & 12 \\ 0 & 3 & 2 & 0 & 0 & 1 & 18 \end{bmatrix} \leftarrow$$

Now, with row 3 as the pivot row (indicated with the arrow above) and element (3,3) as the pivot element (which is the bold-faced "2" above), we divide the pivot row by the pivot element to obtain:

$$\begin{bmatrix} 1 & -3 & -\mathbf{5} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0.5 & 0 & 6 \\ 0 & 3 & 2 & 0 & 0 & 1 & 18 \end{bmatrix} \leftarrow$$

Now eliminate the boldfaced "-5" in the first row by multiplying the third row by 5 and adding it to the first row.

$$\begin{bmatrix} 1 & -3 & 0 & 0 & 2.5 & 0 & 30 \\ 0 & 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0.5 & 0 & 6 \\ 0 & 3 & \mathbf{2} & 0 & 0 & 1 & 18 \end{bmatrix} \begin{matrix} \leftarrow \\ \\ \leftarrow 5 \times \\ \end{matrix}$$

Now eliminate the boldfaced "2" in the fourth row by multiplying the third row by -2 and adding it to the fourth row.

$$\begin{bmatrix} 1 & -3 & 0 & 0 & 2.5 & 0 & 30 \\ 0 & 1 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0.5 & 0 & 6 \\ 0 & 3 & 0 & 0 & -1 & 1 & 6 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow -2 \times \\ \leftarrow \end{matrix}$$

Recalling (x_2, x_3, x_5) are basic variables (with non-zero values) and (x_1, x_4) are non-basic variables (with zero values), one can immediately read off

$x_3=4$ (from second row)
$x_2=6$ (from third row)
$x_5=6$ (from fourth row)

In addition, from the first row, we can easily see the value of the objective function, which is $F=30$ (since all variables with nonzero coefficients are non-basic and therefore zero).

All of this may be a little easier to see if we write down the actual equations corresponding to the last matrix obtained in the Gaussian elimination procedure. These equations are:

$$F \quad -3x_1 \qquad\qquad\qquad +2.5x_4 \qquad\quad = \quad 30 \qquad (12)$$

$$x_1 \qquad\quad +\boxed{x_3} \qquad\qquad\qquad\quad = \quad 4 \qquad (13)$$

$$\boxed{x_2} \qquad +0.5x_4 \qquad\quad = \quad 6 \qquad (14)$$

$$3x_1 \qquad\qquad\qquad -x_4 \quad +\boxed{x_5} = \quad 6 \qquad (15)$$

The circled variables, (x_2, x_3, x_5), are the basic variables, i.e., the ones that are nonzero. The other ones (x_1, x_4) are the non-basic variables, i.e., the ones that are zero.

Let's explore the following question:

<u>Why, in each constraint equation, is there only one non-zero (circled) variable</u>?

➔ This is a result of (a) the fact that we began with a solution (via the initialization procedure) that also had this form and (b) we performed the elimination using a carefully selected pivot and pivot row, which had effects as described below.

- The pivot element was the one corresponding to the entering variable x_2 (which **was** zero and **became** non-zero), and thus, x_2 was eliminated from every equation except the one corresponding to the pivot row. This was important to do in order to ensure that we did not add a non-zero (basic) variable
  - to another constraint equation (only want 1 basic variable per constraint equation), and
  - to the objective equation (want only nonbasic variables in objective equation).

- The pivot row, corresponding to (14), was selected as the one that drove the previous basic variable x_4 to zero, which was what made x_4 the leaving variable. Because
  (a) this equation had previously contained only one non-zero (basic) variable, x_4, and
  (b) we had now converted x_4 to a zeroed (non-basic) variable,
  we could be sure that adding this equation to other equations would not introduce another non-zero (basic) variable into the other equations.

Let's summarize the iteration step with each sub-step and the key idea for that sub-step.

3.1 Identify the entering variable.
   Key idea:
   Select the variable that improves the objective at the highest rate (i.e., the largest amount of objective per unit change in variable).

3.2 Identify the leaving variable.
   Key idea:
   Choose the leaving variable to be the one that hits 0 first as the entering variable is increased, as dictated by one of the m constraint equations.

3.3 Find the new basic feasible solution.
   Key idea:
   Using the equation used to identify the leaving variable as the pivot row, eliminate the entering variable from all other equations.

4.0 Optimality test

The optimality test is easy… see if there are any remaining variables with positive coefficients in the current expression of the objective function. If so, the current solution may still be improved, and we repeat another iteration step. If not, the current solution may not be further improved, and it is therefore optimal.

In our example, the last iteration resulted in

$$F \quad -3x_1 \qquad\qquad +2.5x_4 \qquad\qquad = \quad 30 \qquad (12)$$
$$x_1 \qquad +x_3 \qquad\qquad\qquad = \quad 4 \qquad (13)$$
$$x_2 \qquad +0.5x_4 \qquad\qquad = \quad 6 \qquad (14)$$
$$3x_1 \qquad\qquad -x_4 \quad +x_5 = \quad 6 \qquad (15)$$

Here, the expression of the objective function is

$$F = 30 + 3x_1 - 2.5x_4 \qquad\qquad (16)$$

Since x_1 has a positive coefficient, F may still be improved. Therefore this solution is not optimal.