Lecture Notes – Part 2

Linear Programming: Basics

1 Introduction to Linear Programming

In the end of Chapter 1, we saw how to solve two-variable LP problems graphically. Unfortunately, most real-life LPs have many variables, so a technique is needed to solve LPs with more than two variables.
Consider an optimisation problem

$$\begin{array}{rrll} \text{min (or max)} & z = f(\mathbf{x}) & & \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) & \geq \text{(or mixed with } \leq, =) & \mathbf{0} \\ & \mathbf{x} & \geq \text{(or mixed with } \leq, \text{ urs)} & \mathbf{0} \end{array}$$

If the objective function $f(\mathbf{x})$ and all of the constraints $\mathbf{g}(\mathbf{x})$ are linear, then the considered optimisation problem is an LP, which can be widely found and used in industry. The formulation shown above is called the general form of an LP. It can have both inequality and equality constraints, and also can have variables that are required to be nonnegative/non-positive as well as those allowed to be *unrestricted in sign* (urs).

It is possible that there will be a set of constraints that never can be simultaneously satisfied, in which case *no feasible solution* exists to the LP. Hence, there is no optimal solution to the LP.[1]

It is also possible that the feasible set/region is open or unrestricted in some direction. If this happens, an optimal solution may not be yielded to the LP, which is called *unbounded* (possible but not guaranteed, depending on the objective function).[2] For example, consider the following LP

$$\begin{array}{rrcl} \min z = & -2x_1 - 3x_2 & & \\ \text{s.t.} & x_1 + x_2 & \geq & 3 \\ & x_1,\, x_2 & \geq & 0 \end{array}$$

[1]cf. Case 1 at p.10 in Lecture Note Part 1
[2]cf. Case 4 at p.10 in Lecture Note Part 1

This LP is unbounded. Both decision variables x_1 and x_2 can be made as large as we wish, giving an objective function value below any bound (loosely, the objective function value heads to $-\infty$).

However, consider another LP with the same constraints

$$\begin{array}{rrcl} \min z = & 2x_1 + 3x_2 & & \\ \text{s.t.} & x_1 + x_2 & \geq & 3 \\ & x_1,\, x_2 & \geq & 0 \end{array}$$

This LP is not unbounded, and does have an optimal solution. This is because the feasible region is unrestricted in some direction, which is not the moving direction of the iso-cost line.

Before introducing another LP format called "standard" form, we first describe two conversion processes.

1.1 Nonnegativity of Decision Variables

Although in most practical situations the decision variables, e.g. the number of cars produced, quantity of raw material used, etc, are usually required to be nonnegative, some urs variables could exist in an LP. In solving LPs with the Simplex method, we will need to perform a value examination called "ratio test", which depends on the condition that any feasible solution requires all variables to be nonnegative. Thus, if some variables are allowed to be urs, the ratio test and therefore the Simplex algorithm are no longer valid.

Here the conversion of urs variables to nonnegative variables is introduced. Any variable not already constrained to be nonnegative (by the nature of the problem) can be converted to the difference of two new nonnegative variables. For example, decision variables x_1 and x_2 in the following LP are urs:

$$\begin{array}{rrcl} \min z = & 25x_1 + 30x_2 & & \\ \text{s.t.} & 4x_1 + 7x_2 & \geq & 1 \\ & 8x_1 + 5x_2 & \geq & 3 \\ & 6x_1 + 9x_2 & \geq & -2 \\ & x_1,\, x_2 & \text{urs} & \end{array}$$

Since a real number can be written as the difference between two nonnegative numbers, we can convert this problem to an equivalent one with nonnegative decision variables by introducing four new nonnegative decision variables p_1, q_1, p_2 and q_2, such that

$$\begin{aligned} x_1 &= p_1 - q_1 \\ x_2 &= p_2 - q_2 \end{aligned}$$

Then we have an equivalent LP as follows:

$$\begin{array}{rrcl} \min z = & 25p_1 - 25q_1 + 30p_2 - 30q_2 & & \\ \text{s.t.} & 4p_1 - 4q_1 + 7p_2 - 7q_2 & \geq & 1 \\ & 8p_1 - 8q_1 + 5p_2 - 5q_2 & \geq & 3 \\ & 6p_1 - 6q_1 + 9p_2 - 9q_2 & \geq & -2 \\ & p_1,\, q_1,\, p_2,\, q_2 & \geq & 0 \end{array}$$

1.2 Slack and Surplus Variables and the Matrix Form for Linear Constraints

In order to use the techniques for the solution of the system of equations, it is necessary to convert the inequality constraints of an LP into equality constraints.

We can convert any inequality constraint into an equality constraint by adding slack or surplus variables as appropriate. These variables are defined to be nonnegative. For example, the constraints

$$\begin{aligned} x_1 - 2x_2 &\leq 3 \\ x_1,\, x_2 &\geq 0 \end{aligned}$$

is equivalent to

$$\begin{aligned} x_1 - 2x_2 + s_1 &= 3 \\ x_1,\, x_2,\, s_1 &\geq 0 \end{aligned}$$

The variable s_1 is referred to as a *slack* variable. It results from the fact that, in order for the smaller left-hand-side (lhs) to equal the right-hand-side (rhs), some nonnegative value must be added to the lhs. Similarly, the constraints

$$\begin{aligned} 2x_1 + x_2 &\geq 3 \\ x_1,\, x_2 &\geq 0 \end{aligned}$$

is equivalent to

$$\begin{aligned} 2x_1 + x_2 - e_1 &= 3 \\ x_1,\, x_2,\, e_1 &\geq 0 \end{aligned}$$

The variable e_1 is referred to as a *surplus* (or *excess*) variable. As the lhs is bigger, some nonnegative value must be subtracted from it to achieve equality.

Hence, the system of constraints in any LP, regardless of the direction of the inequality, can be rearranged in the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

or, equivalently

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

or even more briefly as $\mathbf{Ax} = \mathbf{b}$.

It is usual that we have more variables than constraints, i.e. $n > m$. Besides, the built constraints normally shall be *consistent* (i.e. a row of the form $[0\ 0 \cdots 0 \mid c]$ with $c \neq 0$ does not exist after applying Gaussian-Jordan elimination); otherwise, there exists no feasible solution. In fact, if we have $n > m$ and consistent constraints, then there will be an infinite number of feasible solutions. In other words, the system of linear equations has $n - m$ degrees of freedom. This means that it is possible to find the "best" solution from those feasible ones with regard to the objective function, i.e. an optimal solution. And it makes sense to consider "optimisation" for the problems of this type.

1.3 The Standard Form of an LP

Before the Simplex algorithm can be used to solve an LP, the LP must be converted into an equivalent problem in which all constraints are equations and all variables are nonnegative. An LP in this form with nonnegative rhs is said to be in *standard form*. Any LP in general form can be transformed into an equivalent LP in standard form using the two fore-introduced conversion techniques, as shown below:

$$\begin{array}{rlcl}
\max \text{ (or min) } z = & c_1x_1 + c_2x_2 + \cdots + c_nx_n & & \\
\text{s.t.} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\
& & \vdots & \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \\
& x_1,\ x_2,\ \cdots,\ x_n\ \geq 0 & &
\end{array}$$

where $b_i \geq 0$ for $i = 1, 2, \cdots, m$.

Converting to the standard form can be done as follows:

- You can choose whether you want to maximise or minimise the objective function, or leave the objective as it is. If the problem is in minimisation (or maximisation) sense, multiply it by -1 to convert the objective to a maximisation (or minimisation) one.
- Convert any inequality to an equality constraint by the addition of slack or surplus variables (as appropriate).
- If any rhs b_i is negative, multiply the whole constraint by -1.
- Any urs x_j can be replaced by two nonnegative variables x'_j and x''_j: $x_j = x'_j - x''_j$.

In matrix format, an LP in standard form can be written as

$$\begin{array}{rrcl}
\max \text{ (or min) } z = & \mathbf{c}^T\mathbf{x} & & \\
\text{s.t.} & \mathbf{A}\mathbf{x} & = & \mathbf{b} \\
& \mathbf{x} & \geq & \mathbf{0}
\end{array}$$

where $\mathbf{x}$ and $\mathbf{c}$ are n-dimensional vectors, $\mathbf{A}$ is an $m \times n$ matrix, and $\mathbf{b}$ is an m-dimensional vector. Note that $\mathbf{b} \geq \mathbf{0}$.

1.4 Fundamental Law of LP

Convex set: A set $\mathbf{S}$ in n-dimensional space is said *convex* if whenever any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to $\mathbf{S}$ so does every point of the line segment connecting $\mathbf{x}_1$ and $\mathbf{x}_2$. In other words, a set $\mathbf{S}$ is a convex set if the line segment joining any pair of points in $\mathbf{S}$ is wholly contained in $\mathbf{S}$.

Closed half-space: Given an n-dimensional row vector $\mathbf{a}$ and a constant b, the set of all vectors (i.e. points) $\mathbf{x}$ in n-dimensional space satisfying $\mathbf{ax} \leq b$ is called a *closed half-space*. The set of vectors for which $\mathbf{ax} = b$ is called the boundary of the closed half-space.

Extreme point: Given a convex set $\mathbf{S}$ of n-dimensional vectors, a point $\mathbf{x}^*$ is called an *extreme point* (or corner point) of $\mathbf{S}$ if there do not exist any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbf{S}$ and any value $\alpha \in (0,1)$, such that

$$\mathbf{x}^* = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2.$$

In other words, for any convex set $\mathbf{S}$, a point $\mathbf{x}^*$ in $\mathbf{S}$ is an extreme point if each line segment that lies completely in $\mathbf{S}$ and contains the point $\mathbf{x}^*$ has $\mathbf{x}^*$ as an endpoint of the line segment.

Lemma 1 *Every closed half-space is a convex set.*

Suppose that $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in a closed half-space consisting of points which satisfy $\mathbf{ax} \leq b$. Let $\mathbf{x}_3$ be any point on the line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$. Then we have $\mathbf{ax}_1 \leq b$, $\mathbf{ax}_2 \leq b$, and $\mathbf{x}_3 = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ for some $\alpha \in [0,1]$.
So it can be shown that

$$\mathbf{ax}_3 = \mathbf{a}\big(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\big) = \alpha\mathbf{ax}_1 + (1-\alpha)\mathbf{ax}_2 \leq \alpha b + (1-\alpha)b = b,$$

which means $\mathbf{x}_3$ also lies in the considered closed half-space.

Lemma 2 *The intersection of any collection of convex sets is convex.*

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be any two points in the intersection. Then $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to each convex set of the collection. Each convex set contains the line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$. Thus, the line segment belongs to each convex set in the collection so that it belongs to the intersection.

Each feasible set of an LP, if it exists, consists of all vectors that simultaneously satisfy a finite number of linear constraints. Each constraint defines a closed half-space. Thus the feasible set is the intersection of a finite number of closed half-spaces, each of which is convex as per Lemma 1. Lemma 2 then gives us the following theorem.

Theorem 1 *The feasible set of an LP, if not infeasible, is a convex set.*

Since the LP feasible region is a convex set constructed by a finite number of closed half-space boundaries, it is a polyhedron with a finite number of sides (and therefore vertices). Then we have the following theorem.

Theorem 2 *The feasible region for any LP has a finite number of extreme points.*

Theorem 3 *If the feasible set is non-empty and one optimal solution exists to the LP, then there is an optimal solution at one of the extreme points.*

To see this, consider an optimal solution $\mathbf{x}^*$ to a minimisation LP existing on the boundary of the feasible region.[3] Due to the linearity of the constraints, the feasible region is actually a convex polyhedron. Suppose that $\mathbf{x}^*$ is not an extreme point. Then, from the definition of extreme point, there must be two other points $\mathbf{x}_1$ and $\mathbf{x}_2$ in the feasible region, which is convex, and a value α, $0 < \alpha < 1$, such that $\mathbf{x}^* = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$. Since $\mathbf{x}^*$ on the boundary, i.e. some face of the convex polyhedron, it is obvious that $\mathbf{x}_1$ and $\mathbf{x}_2$ are on the same face as well.

Since $\mathbf{x}^*$ is an optimal solution, so $\mathbf{c}^T\mathbf{x}^* \leq \mathbf{c}^T\mathbf{x}_1$ and $\mathbf{c}^T\mathbf{x}^* \leq \mathbf{c}^T\mathbf{x}_2$. Hence

$$\begin{aligned}\mathbf{c}^T\mathbf{x}^* &= \mathbf{c}^T(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) = \alpha\mathbf{c}^T\mathbf{x}_1 + (1-\alpha)\mathbf{c}^T\mathbf{x}_2 \\ &\geq \alpha\mathbf{c}^T\mathbf{x}^* + (1-\alpha)\mathbf{c}^T\mathbf{x}^* = \mathbf{c}^T\mathbf{x}^*\end{aligned}$$

The only situation that the above derivation can be satisfied is $\mathbf{c}^T\mathbf{x}_1 = \mathbf{c}^T\mathbf{x}_2 = \mathbf{c}^T\mathbf{x}^*$, which means that $\mathbf{x}_1$ and $\mathbf{x}_2$ are both optimal solutions as well. Repeating the above argument for optimal solution $\mathbf{x}_1$ or $\mathbf{x}_2$, we can get an optimal solution existing on the edge of the convex polyhedron. Then we can repeat the process until we find an corner point, i.e. extreme point, giving the same objective value and thus being optimal.

[3] It is obvious that we can find an optimal solution $\mathbf{x}^*$ on the boundary due to the linearity of the objective function.

1.5 Basic Feasible Solutions

1.5.1 Definition

Considering an LP in standard form, we have the following system of constraints

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} \qquad (1) \\ \mathbf{x} &\geq \mathbf{0} \qquad (2) \end{aligned}$$

with n variables and m constraints, where $n > m$. Suppose that the feasible region exists and there are an infinite number of solutions. How can we find a solution?

One intuitive approach is to set $n - m$ components of $\mathbf{x}$ equal to zeros. If the columns in $\mathbf{A}$ corresponding to the remaining m variables are linearly independent, solving for the values of these m variables will yield a unique solution. A solution produced by the unique values for the m variables coupled with the zeros for the other $n - m$ variables is called a *basic solution*.

Assume without loss of generality that we take the first m components[4] and call them *basic variables*. The vector constructed with them is named *basis* and denoted by $\mathbf{x_B}$ ($\mathbf{B}$ is for basic). Then denote the vector of the remaining $n - m$ components, i.e. *nonbasic variables*, by $\mathbf{x_N}$, which is called *nonbasis* ($\mathbf{N}$ is for nonbasic). By setting $\mathbf{x_N} = \mathbf{0}$, we ensure that we have the same number of unknowns as that of equations, and open up the possibility of yielding a unique solution

$$(x_1, x_2, \cdots, x_n) = (\mathbf{x_B}^T | \mathbf{x_N}^T).$$

The first m columns, associated with the basic variables, of $\mathbf{A}$ can be labeled $\mathbf{B}$ and the last $n - m$ columns $\mathbf{N}$. Since $\mathbf{A} = [\mathbf{B}|\mathbf{N}]$, Eq. (1) can be written as

$$\mathbf{Bx_B} + \mathbf{Nx_N} = \mathbf{b}.$$

By setting $\mathbf{x_N} = \mathbf{0}$, we get

$$\mathbf{Bx_B} = \mathbf{b}.$$

[4]The numbering in the subscripts of the variables is entirely arbitrary so that basic variables may not always be the "first" m components.

If it can be assumed that the columns of $\mathbf{B}$ are linearly independent, then we have a unique solution

$$\mathbf{x_B} = \mathbf{B}^{-1}\mathbf{b}.$$

Then we say that the variables in $\mathbf{x_B}$ are in the basis, and that $\mathbf{x}^T = (\mathbf{x_B}^T|\mathbf{0})$ is a basic solution. If Eq. (2) is satisfied, i.e. $\mathbf{x_B} = \mathbf{B}^{-1}\mathbf{b} \geq \mathbf{0}$, then we call it a *basic feasible solution* (bfs). A bfs exactly represents an extreme point of the considered LP feasible region.

And the objective value of the bfs is

$$\begin{aligned} z &= \mathbf{c}^T\mathbf{x} = \mathbf{c_B}^T\mathbf{x_B} + \mathbf{c_N}^T\mathbf{x_N} \\ &= \mathbf{c_B}^T\mathbf{B}^{-1}\mathbf{b}. \end{aligned}$$

Obviously, we could swap the order of the variables (and correspondingly, the columns of matrix $\mathbf{A}$), and choose any m variables to be in the basis. Thus, there can be a huge number of possible choices, so we have many possible bases.

We can rewrite Theorem 3 as:

Theorem 4 *If the feasible set is non-empty and one optimal solution exists to the LP, then there is a basic feasible solution giving the optimal value.*

Consider the following LP for example,

$$\begin{aligned} \min z = \quad & x_1 + x_2 \\ \text{s.t.} \quad & 2x_1 + 5x_2 \leq 7 \\ & x_1 + 8x_2 \leq 4 \\ & x_1,\, x_2 \geq 0 \end{aligned}$$

Its standard form is

$$\begin{aligned} \min z = \quad & x_1 + x_2 \\ \text{s.t.} \quad & 2x_1 + 5x_2 + s_1 = 7 \\ & x_1 + 8x_2 + s_2 = 4 \\ & x_1,\, x_2\ s_1,\, s_2 \geq 0 \end{aligned}$$

A bfs is $(x_1, x_2, s_1, s_2) = (0, 0, 7, 4)$ with $z = 0$.

1.5.2 Degeneracy

It is possible that more than one bfs represents the same extreme point of the feasible set. This is known as *degeneracy*, which would cause the inefficiency of the Simplex method for solving an LP.

Another feature of a degenerate bfs is that some of the elements of $\mathbf{x_B}$ are equal to 0s (along with $\mathbf{x_N} = \mathbf{0}$).

We will look at degeneracy again later on.