

# Contents

<b>1</b>	<b>Linear Programming</b>	<b>2</b>
1.1	Geometry of Linear Programming . . . . .	2
1.2	Standard Form of Linear Programs . . . . .	4
1.3	Convex Sets and Functions . . . . .	6
<b>2</b>	<b>The Simplex Method</b>	<b>9</b>
2.1	Basic Feasible Solutions . . . . .	9
2.2	The Simplex Method . . . . .	15

# Linear Programming

Recall that in general, an optimisation problem can be formulated as

$$\begin{aligned} \min_{\mathbf{x} \in X} & f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) = 0 \quad \text{for } i = 1, 2, \dots, p \\ & h_j(\mathbf{x}) \leq 0 \quad \text{for } j = 1, 2, \dots, m, \end{aligned}$$

where  $f$  is known as the *objective function*,  $g_i$ 's are known as *equality constraints* and  $h_j$ 's are known as *inequality constraints*. The set of all  $\mathbf{x}$  that satisfy all constraints is called the *feasible set*, where each of such  $\mathbf{x}$  is a *feasible solution*. The vector which minimises  $f$  is known as the *optimal solution* and is denoted by  $\mathbf{x}^*$ , with  $f(\mathbf{x}^*)$  being the *optimal value*.

The concept of a *linear program* is intuitive to understand: it is simply an optimisation problem whose objective function and constraint functions are all linear. In a 2-dimensional plane, we see that any region bounded by linear functions is a polygon. We will abstract this idea and generalise it for any finite-dimensional space.

## 1.1 Geometry of Linear Programming

In any Euclidean space  $\mathbb{R}^n$ , a linear function can be written as

$$f(x_1, x_2, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i,$$

where the  $a_i$ 's are real coefficients. In matrix notations, this becomes

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} + a_0$$

for some  $\mathbf{c} \in \mathbb{R}^n$ . Let  $f(\mathbf{x}) = a_0 + b$ , then we have  $\mathbf{c}^T \mathbf{x} = b$ . Note that this level set equation gives a linear function in  $\mathbb{R}^{n-1}$ , because obviously  $\mathbf{c}^T \mathbf{x} = c_n x_n + \bar{\mathbf{c}}^T \mathbf{x}'$ , so

$$x_n = \frac{b}{c_n} - \frac{1}{c_n} \bar{\mathbf{c}}^T \mathbf{x}'.$$

Apparently,  $-\frac{1}{c_n}\bar{\mathbf{c}} \in \mathbb{R}^{n-1}$ , so  $x_n$  is a linear function of  $\mathbf{x}'$ . This means every straight line in  $\mathbb{R}^n$  is defined by an equation

$$\mathbf{a}^T \mathbf{x} = b$$

for some  $\mathbf{a} \in \mathbb{R}^n$  and  $b \in \mathbb{R}$ . Take  $\mathbf{x}_0 \in \mathbb{R}^n$  which is on this line and let  $\mathbf{d} \in \mathbb{R}^n$  be the direction vector of the line, then for every  $\mathbf{x}$  with  $\mathbf{a}^T \mathbf{x} = b$ , we also have

$$\mathbf{x} = \mathbf{x}_0 + \lambda \mathbf{d}$$

for some  $\lambda \in \mathbb{R}$ . However, this implies that for any such  $\mathbf{x}$ , we have

$$\mathbf{a}^T(\mathbf{x}_0 + \lambda \mathbf{d}) = \mathbf{a}^T \mathbf{x} = b,$$

but  $\mathbf{a}^T \mathbf{x}_0 = b$ , so we have to have  $\mathbf{a}^T \mathbf{d} = 0$ . Therefore, the set

$$A_\perp := \{\mathbf{x} : \mathbf{a}^T \mathbf{x} = b\}$$

in fact is a set of vectors orthogonal to  $\mathbf{a}$  in  $\mathbb{R}^n$ .

#### Definition 1.1.1 ► Hyperplane

Let  $\mathbf{a} \in \mathbb{R}^n$  be a vector. For any  $b \in \mathbb{R}$ , the set

$$H_b := \{\mathbf{x} : \mathbf{a}^T \mathbf{x} = b\}$$

is said to be a **hyperplane** with normal vector  $\mathbf{a}$ .

It is easy to see that in  $\mathbb{R}^2$ , a hyperplane is a straight line perpendicular to  $\mathbf{a}$  and in  $\mathbb{R}^3$ , it is a plane whose normal vector is parallel to  $\mathbf{a}$ .

Intuitively, a hyperplane partitions the space  $\mathbb{R}^n$  into 2 halves. Therefore, we refer to the set

$$\{\mathbf{x} : \mathbf{a}^T \mathbf{x} \leq b\}$$

as a *half-space*. Intuitively, if we have  $m$  half-spaces, then their intersection is a polyhedron in the space, i.e., we define the set

$$P := \bigcap_{i=1}^m \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_i^T \mathbf{x} \leq b_i\}$$

as a *polyhedral set*.

*Remark.* Note that for  $P$  to be a polyhedron, the intersection must be finite. Otherwise, consider the counter example of the bounded set whose boundary is defined by all hyperplanes at a distance  $d$  away from a fixed point  $Q$ , which is a sphere.

We can let  $\mathbf{a}_i^T$  be the  $i$ -th row of the matrix  $\mathbf{A}$  and define a column vector  $\mathbf{b}$  whose  $i$ -th entry is  $b_i$ , then we can re-write the above intersection using matrix multiplication.

### Definition 1.1.2 ► Polyhedron

A **polyhedron** is defined as the set

$$P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$ .

## 1.2 Standard Form of Linear Programs

### Definition 1.2.1 ► Linear Programming Problem

A **linear programming** (LP) problem is an optimisation problem where the objective function  $f$  is linear and the feasible set  $P$  is a polyhedron.

Note that each linear constraint corresponds to a half-space, so we can formulate a linear programming problem as

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} & \mathbf{c}^T \mathbf{x} \\ \text{s.t. } & \mathbf{a}_i^T \mathbf{x} \leq b_i \quad \text{for } i = 1, 2, \dots, p \\ & \mathbf{a}_j^T \mathbf{x} = b_j \quad \text{for } i = 1, 2, \dots, m, \\ & \mathbf{a}_k^T \mathbf{x} \geq b_k \quad \text{for } i = 1, 2, \dots, q. \end{aligned}$$

where  $\mathbf{c} \in \mathbb{R}^n$  is called the *cost* or *profit* vector,  $\mathbf{a}_i^T \mathbf{x}$ ,  $\mathbf{a}_j^T \mathbf{x}$  and  $\mathbf{a}_k^T \mathbf{x}$  are called the *constraints* and  $\mathbf{x}$  is known as *decision variables*. Note that the number of constraints must be finite, or else we may have a feasible set which is not a polyhedron.

Given any linear program, it may present one of the following possibilities:

1. The program has a unique optimal solution.
2. The program has infinitely many optimal solution (but still a unique optimal value).
3. The program has no optimal solution.

4. The program has no feasible solution.

The first 2 cases are trivial. Suppose a linear program with objective function  $f$  has no optimal solution, then it means for every feasible  $\mathbf{x}$ , we can find a different feasible  $\mathbf{x}'$  such that  $f(\mathbf{x}') < f(\mathbf{x})$ .

Suppose a linear program with objective function  $f$  has no feasible solution, then this means that the feasible set is empty. In this case, we define the optimal value to be  $\infty$ . The reasoning is as follows.

Suppose  $f$  and  $g$  are objective functions of 2 optimisation problems with feasible set  $S_1$  and  $S_2$  respectively such that  $S_1 \subseteq S_2$ . Clearly,  $\min f(\mathbf{x}) \geq \min g(\mathbf{x})$ . Note that if  $S_1 = \emptyset$ , then for every  $S_2$ , we have the above inequality, which means that

$$\min_{\mathbf{x} \in S_1} f(\mathbf{x}) \geq y$$

for all  $y \in \mathbb{R}$ . Therefore,  $\min_{\mathbf{x} \in S_1} f(\mathbf{x}) = \infty$ .

For each inequality constraint in the form of  $\mathbf{a}_i^T \mathbf{x} \leq b_i$ , we can introduce a *slack variable*  $s_i \geq 0$  such that  $\mathbf{a}_i^T \mathbf{x} + s_i = b_i$ . For each constraint on  $x_i$  in the form of  $x_i \leq 0$ , we can replace every occurrence of  $x_i$  by  $-x_i^- = x_i$  such that  $x_i^- \geq 0$ . For each free variable  $x_j$ , we can express it as

$$x_j = x_j^+ - x_j^- \quad \text{for some } x_j^+, x_j^- \geq 0.$$

For instance, we can take  $x_i^+ = 0$  and  $x_i^- > 0$  whenever  $x_i < 0$  and vice versa for  $x_i > 0$ . Note that this correspondence is not unique.

After the above transformations, we see that every constraint is equivalent to either an equality constraint  $\mathbf{a}_i^T \mathbf{x} + s_i = b_i$  or an inequality constraint in the form of  $x_i \geq 0$ . Therefore, we define the following as the *standard form* of a linear program:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} & \mathbf{c}^T \mathbf{x} \\ \text{s.t. } & \mathbf{Ax} = \mathbf{b} \\ & x_i \geq 0, \quad \text{for } i = 1, 2, \dots, m. \end{aligned}$$

One should realise that a linear program in the standard form can be more easily solved by using linear algebra to find the optimal solution. Note that not every optimisation problem is given in the standard form. Fortunately, we can always convert a linear program into the standard form.

## 1.3 Convex Sets and Functions

Intuitively, we describe two types of shapes in natural languages: the shapes which, if you choose any of its edges, lies in the same side of that edge, and the shapes which span across both sides from some chosen edge of its.

Graphically, this means that some shapes are “convex” to all directions, where as some other shapes are “concave”. We shall define this rigorously as follows:

### Definition 1.3.1 ► Convex Set

A set  $D \subseteq \mathbb{R}^n$  is said to be **convex** if for all  $\mathbf{x}, \mathbf{y} \in D$  and for all  $\lambda \in [0, 1]$ ,

$$\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in D.$$

Analogously, we might want to say that a function is convex if, for any 2 points on its graph, the line segment joining the 2 points “lies within” the graph of the function. We can define convexity over functions as follows:

### Definition 1.3.2 ► Convex Function

A function  $f : D \rightarrow \mathbb{R}^n$  is said to be **convex** if for all  $\mathbf{x}, \mathbf{y} \in D$  and for all  $\lambda \in [0, 1]$ ,

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

### Definition 1.3.3 ► Concave Function

A function  $f : D \rightarrow \mathbb{R}^n$  is said to be **concave** if for all  $\mathbf{x}, \mathbf{y} \in D$  and for all  $\lambda \in [0, 1]$ ,

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \geq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

From another perspective, we can see that for any convex function  $f$ , the tangent plane to the graph of  $f$  at any point will lie below the graph.

*Remark.* A function which is not convex must be concave. However, a function which is convex may not be non-concave (consider  $f(x) = x$ ).

Therefore, it is easy to see that all functions in the form of  $f(\mathbf{x}) = d + \mathbf{c}^T \mathbf{x}$  are both convex and concave. Such functions are said to be *affine* functions. Equivalently, this means that all affine functions are neither strictly convex nor strictly concave.

In the above definitions, the expression  $\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}$  for  $\lambda \in [0, 1]$  is known as a *convex combination*. This notion can be generalised for any finite number of terms.

**Proposition 1.3.4 ► Generalised Convex Combination**

Let  $k \in \mathbb{N}^+$  and let  $f : S \rightarrow \mathbb{R}$  be a convex function on the convex set  $S \subseteq \mathbb{R}^n$  and let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in S$ , then

$$f\left(\sum_{i=1}^k \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^k \lambda_i f(\mathbf{x}_i),$$

where  $\sum_{i=1}^k \lambda_i = 1$  and  $\lambda_i \geq 0$  for  $i = 1, 2, \dots, k$ .

Using the idea of convex combinations, we can define the notion of a *convex hull*.

**Definition 1.3.5 ► Convex Hull**

The **convex hull** of  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  is defined as the set of all convex combinations of the vectors, denoted by

$$\text{conv}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) := \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \sum_{i=1}^n \lambda_i \mathbf{x}_i, \lambda_i \in [0, 1], \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Note that  $\text{conv}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  is the smallest convex set containing all of  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ .

Now we consider the following proposition:

**Proposition 1.3.6 ► Maximum of Convex Functions Is Convex**

Let  $f_1, f_2, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  be convex functions, then the function

$$f(\mathbf{x}) := \max_{i=1,2,\dots,m} f_i(\mathbf{x})$$

is convex.

*Proof.* Take any  $\mathbf{x} \neq \mathbf{y} \in \mathbb{R}^n$  and consider  $\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}$  for some  $\lambda \in [0, 1]$ . Note that for each of the  $f_i$ 's, we have

$$f_i(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f_i(\mathbf{x}) + (1 - \lambda) f_i(\mathbf{y}),$$

and so

$$\max_{i=1,2,\dots,m} f_i(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \max_{i=1,2,\dots,m} [\lambda f_i(\mathbf{x}) + (1 - \lambda) f_i(\mathbf{y})].$$

Therefore,

$$\begin{aligned} f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) &= \max_{i=1,2,\dots,m} f_i(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \\ &\leq \max_{i=1,2,\dots,m} [\lambda f_i(\mathbf{x}) + (1 - \lambda) f_i(\mathbf{y})] \\ &= \lambda \max_{i=1,2,\dots,m} f_i(\mathbf{x}) + (1 - \lambda) \max_{i=1,2,\dots,m} f_i(\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}). \end{aligned}$$

□

An immediate corollary of Proposition 1.3.6 allows us to define a piece-wise convex affine function.

### Corollary 1.3.7 ► Piece-wise Affine Functions Are Convex

*The piece-wise affine function*

$$f(\mathbf{x}) = \max_{i=1,2,\dots,n} (\mathbf{c}_i^T \mathbf{x} + d_i)$$

*is convex.*



# The Simplex Method

## 2.1 Basic Feasible Solutions

Recall that given any linear program, its feasible set is a polyhedron  $P$ . Note that in  $\mathbb{R}^n$ , the smallest possible number of hyperplanes intersecting at a point is  $n$ . Intuitively, any polyhedron can be completely defined by all of such “corner points” of itself. Here we provide three equivalent definitions.

### Definition 2.1.1 ▶ Extreme Point

Let  $P$  be a polyhedron, a point  $\mathbf{x}^* \in P$  is said to be an **extreme point** if whenever there are  $\mathbf{y}, \mathbf{z} \in P$  with  $\mathbf{x}^* = \lambda \mathbf{y} + (1 - \lambda) \mathbf{z} = \mathbf{x}^*$  for some  $\lambda \in (0, 1)$ , we have  $\mathbf{y} = \mathbf{z} = \mathbf{x}^*$ .

We can interpret the definition as follows: suppose  $\mathbf{x}^*$  is not a corner point, then there are 2 possibilities. If  $\mathbf{x}^*$  is an internal point, then there is some  $\delta > 0$  such that the neighbourhood  $V_\delta(\mathbf{x}^*) \subseteq P$ . Therefore, we can always find  $\mathbf{y}, \mathbf{z} \in V_\delta(\mathbf{x}^*)$  with  $\mathbf{y} \neq \mathbf{z} \neq \mathbf{x}^*$  such that  $\mathbf{x}^* \in (\mathbf{y}, \mathbf{z})$ . Otherwise,  $\mathbf{x}^*$  is on the boundary, i.e.,

$$\mathbf{x}^* \in H := \{\mathbf{x} : \mathbf{a}^T \mathbf{x} \leq b\}.$$

Clearly, we can also find  $\mathbf{y}, \mathbf{z} \in V_\delta(\mathbf{x}^*) \cap H$  with  $\mathbf{y} \neq \mathbf{z} \neq \mathbf{x}^*$  such that  $\mathbf{x}^* \in (\mathbf{y}, \mathbf{z})$ .

Alternatively, we consider a hyperplane defined by  $\mathbf{c}^T \mathbf{x} = b$  such that  $\mathbf{c}$  is not orthogonal to any of the boundaries of the polyhedron  $P$ . Clearly,  $\mathbf{c}$  is the gradient vector of the affine function

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}.$$

Therefore, by translating the hyperplane in the direction of  $\mathbf{c}$ , we are able to maximise  $b$ . Note that this hyperplane is not parallel to any boundary of the polyhedron, so when  $b$  reaches the maximum, the hyperplane will have a unique intersection with the polyhedron, which must be a corner point.

### Definition 2.1.2 ▶ Vertex

Let  $P$  be a polyhedron, a point  $\mathbf{x}^* \in P$  is said to be a **vertex** if there exists some  $\mathbf{c}$  such that  $\mathbf{c}^T \mathbf{x}^* > \mathbf{c}^T \mathbf{y}$  for all  $\mathbf{y} \in P - \{\mathbf{x}^*\}$ .

Note that here we require the inequality to be strict, because otherwise  $\mathbf{x}^*$  and  $\mathbf{y}$  may both be internal points of some boundary hyperplane of  $P$ .

Note that any boundary of a polyhedron  $P$  is uniquely determined by a constraint  $\mathbf{a}_i^T \mathbf{x} \leq b_i$ . Let  $\mathbf{x}^* \in P$ , it is clear that  $\mathbf{x}^*$  is “on the boundary” if and only if  $\mathbf{a}_i^T \mathbf{x} = b_i$  for some  $i$ . In such cases, we say that the corresponding constraint is *active/binding/tight* at  $\mathbf{x}^*$ .

Geometrically, a corner point of a polyhedron  $P$  in  $\mathbb{R}^n$  is the intersection of  $n$  boundary hyperplanes which are pair-wise non-parallel. Let  $\mathbf{a}_i^T \mathbf{x} = b_i$  and  $\mathbf{a}_j^T \mathbf{x} = b_j$  be any of the  $n$  hyperplanes, then it is clear that  $\mathbf{a}_i$  and  $\mathbf{a}_j$  must be linearly independent for the two hyperplanes to be non-parallel. This leads to the notion of *basic feasible solutions*.

### Definition 2.1.3 ► Basic Feasible Solution

Let  $P \subseteq \mathbb{R}^n$  be a polyhedron.  $\mathbf{x}^* \in P$  is said to be a **basic feasible solution** if there are  $n$  linearly independent constraints which are active at  $\mathbf{x}^*$ .

We can generalise Definition 2.1.3 to deal with even infeasible points. First we introduce some terminologies.

### Definition 2.1.4 ► Rank

Let  $P \subseteq \mathbb{R}^n$  be a polyhedron with constraints  $\mathbf{a}_i^T \mathbf{x} \leq b_i$  for  $i = 1, 2, \dots, m$ . For any  $\mathbf{x} \in \mathbb{R}^n$ , the **rank** of  $\mathbf{x}$  is defined as

$$\text{rank}(\mathbf{x}) := \dim(\text{span}\{\mathbf{a}_j : \mathbf{a}_j^T \mathbf{x} = b_j\}).$$

Clearly,  $\mathbf{x}^* \in \mathbb{R}^n$  is a basic feasible solution if and only if  $\text{rank}(\mathbf{x}^*) = n$  and  $\mathbf{x}^* \in P$ . Notice that not all  $\mathbf{x}$  with rank  $n$  is feasible, so we might consider the following definition:

### Definition 2.1.5 ► Basic Solution

Let  $P \subseteq \mathbb{R}^n$  be a polyhedron. A vector  $\mathbf{x} \in \mathbb{R}^n$  is a **basic solution** if  $\text{rank}(\mathbf{x}) = n$ .

One important thing to note here is that  $\text{rank}(\mathbf{x}) = n$  does not necessarily imply that there are exactly  $n$  constraints active at  $\mathbf{x}$ . Intuitively, there can be more than  $n$  hyperplanes in  $\mathbb{R}^n$  which intersect at a point  $\mathbf{x}$ . However, if  $\text{rank}(\mathbf{x}) = n$ , then some hyperplanes are “redundant”, i.e., their normal vectors can be expressed as linear combinations of the gradients of some  $n$  linearly independent constraints.

### Definition 2.1.6 ► Degeneracy

A basic solution  $\mathbf{x} \in \mathbb{R}^n$  is said to be **degenerate** if there are more than  $n$  constraints active at  $\mathbf{x}$ .

Let  $P \subseteq \mathbb{R}^n$  be a polyhedron defined by  $m$  constraints. Clearly, for each basic feasible solution, we require at least  $n$  different constraints to be active. Therefore, the number of distinct basic feasible solutions in  $P$  is at most  $\binom{m}{n}$ . This justifies the fact that any polyhedron in  $\mathbb{R}^n$  determined by less than  $n$  constraints has no basic feasible solution, and any polyhedron in a finite-dimensional space must have finitely many basic feasible solutions.

Suppose we are given a standard linear program, then we can write its feasible set as the polyhedron

$$P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\},$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . Obviously,  $P$  has no basic feasible solution if  $m < n$ . Suppose  $m \geq n$ , we will devise a way to compute the basic feasible solutions systematically.

### Theorem 2.1.7 ► Basic Solution Characterisation

Let

$$P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

be a polyhedron for some  $\mathbf{A} \in \mathbb{R}^{m \times n}$  with  $m \geq n$ . A vector  $\mathbf{x}^* \in \mathbb{R}^n$  is a basic solution if and only if

- $\mathbf{Ax}^* = \mathbf{b}$ , and
- There exists an index set  $B \subseteq \{1, 2, \dots, n\}$  such that the set

$$\{\mathbf{A}_i : i \in B\}$$

is linearly independent and  $x_j^* = 0$  for all  $j \notin B$ , where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \end{bmatrix}.$$

*Proof.* Write  $B = \{B(1), B(2), \dots, B(m)\}$  and define

$$N = \{N(1), N(2), \dots, N(n-m)\} := \{1, 2, \dots, n\} - B.$$

For each  $i \in N$ , since  $x_i^* = 0$ , we have  $\mathbf{e}_i^T \mathbf{x}^* = 0$ . Therefore, the matrix representation for the active constraints is

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{e}_{N(1)}^T \\ \mathbf{e}_{N(2)}^T \\ \vdots \\ \mathbf{e}_{N(n-m)}^T \end{bmatrix} \mathbf{x}^* = \mathbf{0}.$$

Re-arranging the columns, the above matrix can be re-written as

$$\begin{bmatrix} \mathbf{A}_B & \mathbf{A}_N \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix} \bar{\mathbf{x}}^* = \mathbf{0},$$

where  $\bar{\mathbf{x}}^*$  is obtained by re-arranging the rows of  $\mathbf{x}^*$  accordingly. Note that the columns of  $\mathbf{A}_B$  is linearly independent, so  $\det(\mathbf{A}_B) \neq 0$ . Therefore,

$$\begin{vmatrix} \mathbf{A}_B & \mathbf{A}_N \\ \mathbf{0} & \mathbf{I}_N \end{vmatrix} = \det(\mathbf{A}_B) \det(\mathbf{I}_N) \neq 0,$$

and so the matrix is invertible. Therefore, the rows of the matrix are linearly independent. This means that there are  $n$  linearly independent constraints active at  $\mathbf{x}^*$ . Therefore,  $\mathbf{x}^*$  is a basic feasible solution.

Suppose conversely that  $\mathbf{x}^*$  is a basic feasible solution, then clearly  $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ . Since there are  $m$  equality constraints, then we must have  $(n - m)$  active inequality constraints at  $\mathbf{x}^*$ , indexed by  $N = \{N(1), N(2), \dots, N(n - m)\}$ , such that the constraints are linearly independent. Therefore, the matrix

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{e}_{N(1)}^T \\ \mathbf{e}_{N(2)}^T \\ \vdots \\ \mathbf{e}_{N(n-m)}^T \end{bmatrix}$$

is invertible and that for all  $i \in N$ ,  $x_i^* = 0$ . Let

$$B = \{B(1), B(2), \dots, B(m)\} := \{1, 2, \dots, n\} - N$$

be an index set, then the above matrix can be re-arranged as

$$\begin{bmatrix} \mathbf{A}_B & \mathbf{A}_N \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix},$$

which is invertible. Therefore,  $\{\mathbf{A}_{B(1)}, \mathbf{A}_{B(2)}, \dots, \mathbf{A}_{B(m)}\}$  is linearly independent.  $\square$

Note that according to Theorem 2.1.7, for every  $i \in N$ ,  $x_i = 0$ . Note that  $\mathbf{A} = \begin{bmatrix} \mathbf{A}_B & \mathbf{A}_N \end{bmatrix}$ ,

so the linear system  $\mathbf{Ax} = \mathbf{b}$  is equivalent to  $\mathbf{A}_B \mathbf{x}_B = \mathbf{b}$ . Therefore, given any standard polyhedron  $P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ , we can construct a basic solution in  $P$  using the following procedures:

1. Choose  $m$  linearly independent columns from  $\mathbf{A}$  with  $B$  being the index set for the columns to form the set  $\{\mathbf{A}_{B(1)}, \mathbf{A}_{B(2)}, \dots, \mathbf{A}_{B(m)}\}$ ;
2. For each  $i \in N := \{1, 2, \dots, n\} - B$ , set  $x_i = 0$ . The vector consisting of all of these zero entries is denoted by  $\mathbf{x}_N$ ;
3. Solve the linear system  $\mathbf{A}_B \mathbf{x}_B = \mathbf{b}$  to obtain  $\mathbf{x}_B = \mathbf{A}_B^{-1} \mathbf{b}$ ;
4. Let  $x_i$  be the  $i$ -th entry of  $\mathbf{x}$ , then

$$x_i = \begin{cases} 0, & \text{if } i \in N \\ (\mathbf{x}_B)_i & \text{if } i \in B \end{cases}.$$

The  $\mathbf{x}$  obtained this way is alternatively denoted as  $\mathbf{x} := (\mathbf{x}_B, \mathbf{x}_N)$ .

Geometrically, 2 distinct basic feasible solutions of  $P$  are *adjacent* if there is an edge on the boundary joining the 2 points. Here we give an equivalent definition algebraically.

#### Definition 2.1.8 ► Adjacency

Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be distinct basic solutions with respect to polyhedron  $P$ .  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are said to be **adjacent** if there are exactly  $(n - 1)$  linearly independent constraints active at both points, or their corresponding bases only contain 1 different basic column.

Recall that not all polyhedrons have basic feasible solutions. Informally, we can see that if a polyhedron does not contain any basic feasible solution, it contains at least 2 “openings” which allows us to place a straight line into the polyhedron. This observation is summarised rigorously as follows:

#### Theorem 2.1.9 ► Conditions for the Existence of Basic Feasible Solution

Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and

$$P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\} \neq \emptyset,$$

then the following statements are equivalent:

1.  $P$  does not contain any straight line.
2.  $P$  has a basic feasible solution.
3.  $P$  has  $n$  linearly independent constraints.

*Proof.* Suppose  $P$  has a basic feasible solution  $\mathbf{x}^*$ , then there are  $n$  linearly independent constraints active at  $\mathbf{x}^*$ . Therefore, it is trivial that  $P$  must contain at least  $n$  linearly independent constraints.

We shall prove that (3) implies (1) by considering the contrapositive statement. Suppose that  $P$  contains a straight line  $\{\mathbf{x}_0 + \lambda \mathbf{d} : \lambda \in \mathbb{R}\}$  for some fixed point  $\mathbf{x}_0 \in P$  and direction vector  $\mathbf{d} \in \mathbb{R}^n$ , then for any  $\lambda \in \mathbb{R}$ , we have  $\mathbf{A}(\mathbf{x}_0 + \lambda \mathbf{d}) \leq \mathbf{b}$ . Notice that  $\mathbf{A}\mathbf{x}_0 \leq \mathbf{b}$ , so  $\mathbf{A}\mathbf{d} = \mathbf{0}$ . However,  $\mathbf{d} \neq \mathbf{0}$ , so  $\mathbf{A}$  does not contain  $n$  independent rows, which implies that  $P$  does not have  $n$  linearly independent constraints.

Suppose that  $P$  does not contain any straight line. Take some  $\mathbf{x} \in P$  and let

$$I(\mathbf{x}) := \{\mathbf{a}_i : \mathbf{a}_i^T \mathbf{x} = b_i\}$$

be the set of gradient vectors of all active constraints at  $\mathbf{x}$ . If  $I(\mathbf{x})$  contains  $n$  linearly independent vectors, then we are done. Otherwise,  $I(\mathbf{x})$  does not span  $\mathbb{R}^n$ , so there is some  $\mathbf{d} \in \mathbb{R}^n$  with  $\mathbf{d} \neq \mathbf{0}$  such that it is normal to  $\text{span}(I(\mathbf{x}))$ . Since  $P$  contains no straight lines, there exists some constraints with gradient vector  $\mathbf{a}_j \notin I(\mathbf{x})$  such that we can find some  $\mu \in \mathbb{R}$  such that  $\mathbf{a}_j^T(\mathbf{x} + \mu \mathbf{d}) = b_j$ . Set  $\mathbf{x}' = \mathbf{x} + \mu \mathbf{d}$ . Note that for all  $\mathbf{a}_i \in I(\mathbf{x})$ , we have  $\mathbf{a}_i^T \mathbf{d} = 0$ , and so  $\mathbf{a}_i^T(\mathbf{x} + \mu \mathbf{d}) = b_i$ . Therefore,  $\mathbf{a}_i \in I(\mathbf{x}')$ , and so we have  $I(\mathbf{x}) \cup \{\mathbf{a}_j\} \subseteq I(\mathbf{x}')$ . We claim that  $\mathbf{a}_j$  linearly independent with  $I(\mathbf{x})$ . Otherwise,  $\mathbf{a}_j^T \mathbf{d} = 0$ , which means that the boundary defined by  $\mathbf{a}_j^T \mathbf{x} = b_j$  is parallel to  $\mathbf{d}$  and so if the constraint is active at  $\mathbf{x}'$ , it must also be active at  $\mathbf{x}$ , which is a contradiction. Therefore,  $I(\mathbf{x}')$  has more linearly independent vectors than  $I(\mathbf{x})$ . Repeat this process and we will eventually obtain a set of  $n$  linearly independent gradient vectors,  $I(\mathbf{x}^*)$ , where  $\mathbf{x}^*$  is a basic feasible solution.  $\square$

Consider the level set defined by  $f(\mathbf{x}) = k$  for some affine function  $f$ . Suppose the level set has an intersection with some polyhedron  $P$ , then by translating the level set in the direction of  $-\nabla f$ , it will eventually intersect  $P$  at a basic feasible solution such that any further translation will make the intersection empty. Therefore, a reasonable guess is that an optimal solution for any linear program is closely linked to the basic feasible solutions of the feasible set.

### Theorem 2.1.10 ► Optimality of Basic Feasible Solutions

Consider the linear program

$$\min_{\mathbf{x} \in P} f(\mathbf{x})$$

where  $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$  and  $P$  is a polyhedron. If  $P$  has a basic feasible solution and the linear program has an optimal solution, then there exists a basic feasible solution of  $P$  which is an optimal solution to the linear program.

*Proof.* Let the optimal value of  $f$  be  $v^*$ , then the set of optimal solutions is

$$Q := P \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T \mathbf{x} = v^*\}.$$

Note that  $Q$  is a polyhedron. By Theorem 2.1.9, since  $P$  contains a basic feasible solution, it does not contain a straight line. Therefore,  $Q \subseteq P$  cannot contain a straight line and so it contains a basic feasible solution.

Let  $\mathbf{x}^*$  be any basic feasible solution of  $Q$ . We claim that  $\mathbf{x}^*$  is a basic feasible solution of  $P$ . Suppose on contrary that  $\mathbf{x}^*$  is not a basic feasible solution of  $P$ , then there exists some  $\mathbf{d} \in \mathbb{R}^n$  with  $\mathbf{c}^T \mathbf{d} = 0$  such that  $\mathbf{x}^* + \lambda \mathbf{d} \in Q$  is an internal point of  $P$  for some  $\lambda \in \mathbb{R}$ . However, this means that there exists some  $\mu > 0$  such that  $\mathbf{x}^* - \mu \mathbf{d} \in P$ . Note that  $\mathbf{c}^T (\mathbf{x}^* - \mu \mathbf{d}) < v^*$ , which is impossible. Therefore,  $\mathbf{x}^*$  must be a basic feasible solution of  $P$ .  $\square$

Note that this essentially implies that  $Q$  is the convex hull of basic feasible solutions of  $P$ . Therefore, any optimal solution of the original linear program can be expressed as a convex combination of basic feasible solutions of  $P$ , and so it suffices to first find all basic feasible solutions if our feasible set contains any.

## 2.2 The Simplex Method

Note that for any linear program, we can convert it to a standard linear program with feasible region

$$P := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}.$$

Let  $\{\mathbf{x}_0 + \lambda \mathbf{d} : \lambda \in \mathbb{R}\}$  be any straight line, then clearly we can find some  $\lambda \in \mathbb{R}$  such that there is some entry  $x_i$  of  $\mathbf{x}$  with  $x_i < 0$ . Therefore,  $P$  does not contain any straight line, and so  $P$  always contains at least one basic feasible solution if  $P \neq \emptyset$ . Therefore, as long as a linear program has a finite optimal value, it has optimal solutions which are basic feasible solutions.

Therefore, to find the optimal solution of a linear program, we only need to start at any basic feasible solution and try to reach an adjacent basic feasible solution which can improve our objective value. Continue this search and eventually we will be able to collect all optimal solutions.

Given any basic feasible solution  $\mathbf{x}$ , we wish to find a direction  $\mathbf{d}$  such that  $\mathbf{x} + \theta\mathbf{d}$  is some adjacent basic feasible solution for some  $\theta \in \mathbb{R}$ . First, we need to ensure that  $\mathbf{x} + \theta\mathbf{d}$  is still feasible.

### Definition 2.2.1 ► Feasible Direction

Let  $P$  be a polyhedron and  $\mathbf{x} \in P$  be a feasible point. A vector  $\mathbf{d}$  is a **feasible direction** if  $\mathbf{x} + \lambda\mathbf{d} \in P$  for some  $\lambda > 0$ .

Suppose  $\mathbf{d}$  is a feasible direction and  $\mathbf{x}$  is a basic feasible solution. Consider  $\mathbf{x}' = \mathbf{x} + \theta\mathbf{d} \in P$  for some  $\theta \in \mathbb{R}$ . If  $\mathbf{x}'$  is on an edge, then clearly there is exactly one less active constraint at  $\mathbf{x}'$  than at  $\mathbf{x}$ .

### Theorem 2.2.2 ► Characterisation of A Direction Connecting Basic Feasible Solutions

Let  $\mathbf{x} = (\mathbf{x}_B, \mathbf{x}_N)$  with  $\mathbf{x}_B \geq \mathbf{0}$  and  $\mathbf{x}_N = \mathbf{0}$  be a basic feasible solution, then a direction that connects  $\mathbf{x}$  to an adjacent basic feasible solution is in the form of  $\mathbf{d}^j = (\mathbf{d}_B^j, \mathbf{d}_N^j)$  for some  $j \in N$ , such that  $\mathbf{d}_N^j = \mathbf{e}_j$  and  $\mathbf{d}_B^j = -\mathbf{A}_B^{-1}\mathbf{A}_j$ .

*Proof.* Note that for any feasible  $\mathbf{x}' = \mathbf{x} + \theta\mathbf{d}^j$  which is not a basic feasible solution, we have  $\mathbf{A}\mathbf{x}'_B = \mathbf{b}$ . This means that all constraints corresponding to  $B$  are active along the edge. Therefore, the edge frees one constraint of the form  $x_j \geq 0$  for some  $j \in N$ . This means that the index set of active constraints along the edge is given by  $B \cup N - \{j\}$ . Note that

$$\mathbf{x} + \theta\mathbf{d}^j = (\mathbf{x}_B + \theta\mathbf{d}_B^j, \mathbf{x}_N + \theta\mathbf{d}_N^j).$$

Notice that  $\mathbf{x}_N = \mathbf{0}$  and  $\mathbf{x}_N + \theta\mathbf{d}_N^j = \theta\mathbf{e}_j$ , so  $\mathbf{d}_N^j = \mathbf{e}_j$ . Since  $\mathbf{x} + \theta\mathbf{d}^j$  is feasible, we have

$$\mathbf{A}(\mathbf{x} + \theta\mathbf{d}^j) = \mathbf{b} = \mathbf{A}\mathbf{x}.$$

Therefore,  $\mathbf{A}\mathbf{d}^j = \mathbf{0}$ . However, note that

$$\begin{aligned} \mathbf{A}\mathbf{d}^j &= \begin{bmatrix} \mathbf{A}_B & \mathbf{A}_N \end{bmatrix} \begin{bmatrix} \mathbf{d}_B^j \\ \mathbf{d}_N^j \end{bmatrix} \\ &= \mathbf{A}_B\mathbf{d}_B^j + \mathbf{A}_N\mathbf{e}_j \\ &= \mathbf{A}_B\mathbf{d}_B^j + \mathbf{A}_j. \end{aligned}$$

Since  $\mathbf{A}_B$  has linearly independent columns, it is invertible, so  $\mathbf{d}_B^j = -\mathbf{A}_B^{-1}\mathbf{A}_j$ . □

Simply traversing between basic feasible solutions is not very useful, because our ultimate goal is to minimise our objective function, i.e., we wish to find a direction  $\mathbf{d}^j$  such that for



some  $\theta > 0$ ,

$$\mathbf{c}^T (\mathbf{x} + \theta \mathbf{d}^j) < \mathbf{c}^T \mathbf{x}.$$

Clearly, we require  $\mathbf{c}^T \mathbf{d}^j < 0$ . Note that

$$\begin{aligned} \mathbf{c}^T \mathbf{d}^j &= (\mathbf{c}_B, \mathbf{c}_N)^T \begin{bmatrix} \mathbf{d}_B^j \\ \mathbf{d}_N^j \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{c}_B^T & \mathbf{c}_N^T \end{bmatrix} \begin{bmatrix} -\mathbf{A}_B^{-1} \mathbf{A}_j \\ \mathbf{e}_j \end{bmatrix} \\ &= -\mathbf{c}_B^T \mathbf{A}_B^{-1} \mathbf{A}_j + c_j, \end{aligned}$$

so our target is simply  $c_j - \mathbf{c}_B^T \mathbf{A}_B^{-1} \mathbf{A}_j < 0$ .

### Definition 2.2.3 ► Reduced Cost

Let  $\mathbf{x}$  be a basic feasible solution with respect to objective function  $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$ . Let  $\mathbf{c} = (\mathbf{c}_B, \mathbf{c}_N)$ . For each  $j = 1, 2, \dots, n$ , the **reduced cost** of variable  $x_j$  is defined as

$$\bar{c}_j = c_j - \mathbf{c}_B^T \mathbf{A}_B^{-1} \mathbf{A}_j.$$

*Remark.* For any  $j \in B$ , note that  $\mathbf{A}_j = \mathbf{A}_B \mathbf{e}_j$ , we have  $\mathbf{A}_B^{-1} \mathbf{A}_j = \mathbf{e}_j$ . Therefore,  $\bar{c}_j = 0$ .

We say that a direction  $\mathbf{d}^j$  is an *improving direction* if and only if  $\bar{c}_j < 0$ .

Next, we only need to determine a  $\bar{\theta}_j > 0$  such that  $\mathbf{x} + \bar{\theta}_j \mathbf{d}^j$  gives us another basic feasible solution. Since we already know that  $\mathbf{A}(\mathbf{x} + \bar{\theta}_j \mathbf{d}^j) = \mathbf{b}$  for all  $\bar{\theta}_j > 0$ , a natural idea is to take  $\bar{\theta}_j$  to be the greatest positive real number such that  $\mathbf{x} + \bar{\theta}_j \mathbf{d}^j \geq \mathbf{0}$ , i.e., we will move in the direction  $\mathbf{d}^j$  until we can barely stay in the feasible polyhedron. Consider

$$\mathbf{x} + \bar{\theta}_j \mathbf{d}^j = (\mathbf{x}_B + \bar{\theta}_j \mathbf{d}_B^j, \mathbf{x}_N + \bar{\theta}_j \mathbf{d}_N^j).$$

Note that  $\mathbf{x}_N \geq \mathbf{0}$  and  $\bar{\theta}_j \mathbf{d}_N^j = \bar{\theta}_j \mathbf{e}_j \geq \mathbf{0}$ , so it suffices to check that  $\mathbf{x}_B + \bar{\theta}_j \mathbf{d}_B^j \geq \mathbf{0}$ . This implies that for each  $i \in B$ , we have  $\bar{\theta}_j \geq -\frac{x_i}{d_i^j}$ . Therefore, we only need to take

$$\bar{\theta}_j = \min \left\{ -\frac{x_i}{d_i^j} : i \in B, d_i^j < 0 \right\}.$$

Note that here we do not consider those  $d_i^j$ 's with  $d_i^j > 0$  because in such cases  $x_i + \bar{\theta}_j d_i^j \geq 0$  for all  $\bar{\theta}_j > 0$ . Now, we would verify that the new point we reach is indeed another basic feasible solution.

**Proposition 2.2.4 ▶  $\mathbf{x} + \bar{\theta}_j \mathbf{d}_j$  Is A Basic Feasible Solution**

If  $\{i \in B : d_i^j < 0\} \neq \emptyset$ , then  $\mathbf{x} + \bar{\theta}_j \mathbf{d}_j$  is a basic feasible solution adjacent to  $\mathbf{x}$ .

*Proof.* Note that  $\mathbf{x} + \bar{\theta}_j \mathbf{d}_j$  is feasible, so  $\mathbf{A}(\mathbf{x} + \bar{\theta}_j \mathbf{d}_j) = \mathbf{b}$ . Note that by the definition of  $\bar{\theta}_j$ , we have

$$\bar{\theta}_j = -\frac{x_\ell}{d_\ell^j}$$

for some  $\ell \in B$ . Therefore, we have  $(\mathbf{x} + \bar{\theta}_j \mathbf{d}_j)_\ell = 0$ . Notice that at  $\mathbf{x} := (\mathbf{x}_B, \mathbf{x}_N)$ , we have  $\ell \in B$  and  $j \in N$ . Consider  $\bar{B} := (B - \{\ell\}) \cup \{j\}$  and  $\bar{N} := (N - \{j\}) \cup \{\ell\}$ . One may check that  $\bar{B}$  is linearly independent, so by Theorem 2.1.7,  $\mathbf{x} + \bar{\theta}_j \mathbf{d}_j$  is a basic feasible solution. □