InFoMM – Optimisation
Lecture 1

Raphael Hauser

Oxford Mathematical Institute

Michaelmas Term 2019

What is integer programming?

Integer Programming concerns the mathematical analysis of and design of algorithms for optimisation problems of the following forms.

- (Linear) Integer Program:

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \; & c^{\mathsf{T}} x \\ \text{s.t.} \; & Ax \leq b, \quad \text{(componentwise)} \\ & x \geq 0, \quad \text{(componentwise)} \\ & x \in \mathbb{Z}^n, \end{aligned}$$

  where $A \in \mathbb{R}^{m \times n}$ is a matrix and $b \in \mathbb{R}^m, c \in \mathbb{R}^n$ are vectors with rational coefficients.

- Binary (Linear) Integer Program:

$$\begin{aligned}\max_x\ & c^{\mathsf{T}}x\\ \text{s.t.}\ & Ax \leq b\\ & x \in \mathbb{B}^n := \{0,1\}^n.\end{aligned}$$

This is a special case of a linear integer program, as it can be reformulated as

$$\begin{aligned}\max_x\ & c^{\mathsf{T}}x\\ \text{s.t.}\ & Ax \leq b\\ & x \leq \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}\\ & x \geq 0,\\ & x \in \mathbb{Z}^n.\end{aligned}$$

- (Linear) Mixed Integer Program:

$$\begin{aligned} \max_{x,y}\ & c^{\mathsf{T}} x + h^{\mathsf{T}} y \\ \text{s.t. } & Ax + Gy \leq b \\ & x, y \geq 0, \\ & y \in \mathbb{Z}^p, \end{aligned}$$

  where G is a matrix and h a vector with rational coefficients.

Linear Programming

An important special case of an integer programming problem is one without integrality constraints, e.g.,

$$\begin{aligned} \max_x \; & c^{\mathrm{T}} x \\ \text{s.t.} \quad & Ax \leq b, \\ & x \geq 0. \end{aligned}$$

Such problems are called *linear programming* problems (or LPs). We will see that LPs play an important role in algorithms designed to solve general IPs through the concept of of *LP relaxation*:

Consider the IP problem

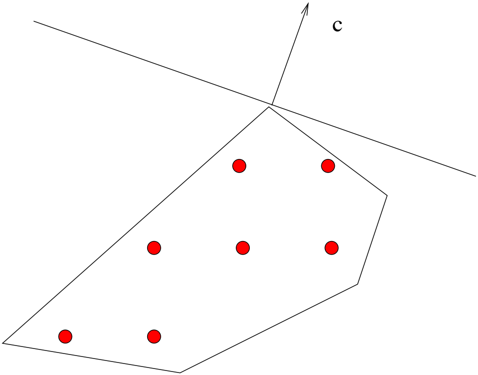
$$\text{(IP)} \quad z^* = \max_x c^{\mathsf{T}} x \quad \text{s.t.} \quad Ax \le b, \quad x \in \mathbb{Z}^n_+.$$

If we give up on the integrality constraints $x_i \in \mathbb{Z}$, we obtain an LP,

$$\text{(LP)} \quad \bar{z} = \max_x c^{\mathsf{T}} x \quad \text{s.t.} \quad Ax \le b, \quad x \ge 0.$$

Giving up on the integrality constraints has two effects on the feasible set $\mathscr{F}$ (the set of decision vectors x that satisfy the constraints of the problem)

- $\mathscr{F}$ becomes larger,
- $\mathscr{F}$ becomes convex.



Proposition (Relaxation implies dual bound)

The consequence of the first effect is that $\bar{z} \geq z^*$.

Proof.

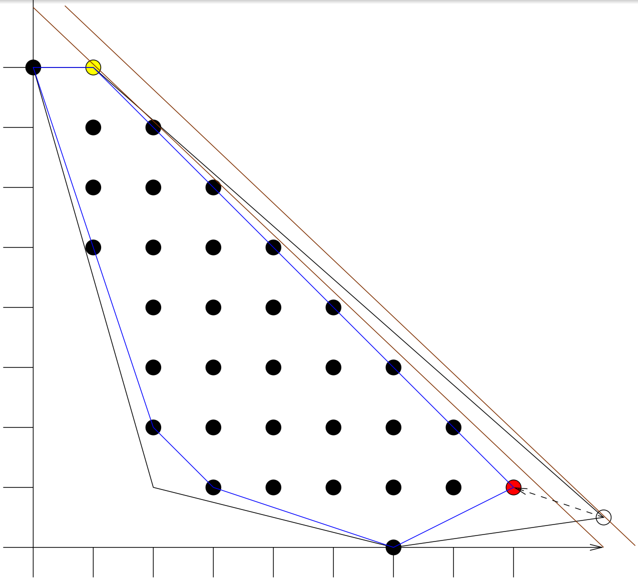
If the optimal objective value z^* of (IP) is achieved at the point x^*, then x^* is feasible for (IP), and hence it is also feasible for (LP). Therefore, $\bar{z} \geq c^{\mathrm{T}} x = z^*$. □

As we shall learn, the consequence of the second effect is that it is much easier to solve the problem (LP) than (IP).

A first idea for solving IPs is to solving the LP relaxation and round the optimal values of the decision variables to the nearest feasible integer valued feasible solution.

While this occasionally works, it is not always a good idea:

- Rounding may be non-trivial, e.g., when the LP relaxation of a binary program takes an optimal solution x^* with many values near 0.5.
- The rounded solution may be far from optimal.



The Simplex Algorithm in Dictionary Form

We will now discuss an algorithm for solving general linear programming problems.

Example (Simplex in dictionary form)

Consider the LP instance

$$
\begin{aligned}
z = \max_x \; & 5x_1 + 4x_2 + 3x_3 \\
\text{s.t.} \quad & 2x_1 + 3x_2 + x_3 \leq 5 \\
& 4x_1 + x_2 + 2x_3 \leq 11 \\
& 3x_1 + 4x_2 + 2x_3 \leq 8 \\
& x_1, x_2, x_3 \geq 0.
\end{aligned}
$$

Preliminary step I: introduce *slack variables* $x_4, x_5, x_6 \geq 0$ to reformulate inequality constraints as a system of linear equations,

$$
\begin{aligned}
z = \max\ & 5x_1 + 4x_2 + 3x_3 + 0x_4 + 0x_5 + 0x_6 \\
\text{s.t.}\quad & 2x_1 + 3x_2 + x_3 + x_4 = 5 \\
& 4x_1 + x_2 + 2x_3 + x_5 = 11 \\
& 3x_1 + 4x_2 + 2x_3 + x_6 = 8 \\
& x_1, x_2, x_3, x_4, x_5, x_6 \geq 0.
\end{aligned}
$$

Preliminary step II: express in *dictionary form*

$$\max z \quad \text{s.t.} \quad x_1, \ldots, x_6 \geq 0,$$

and where the variables are linked via the linear system

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

Step 0: $x_1, x_2, x_3 = 0$, $x_4 = 5$, $x_5 = 11$, $x_6 = 8$ is an initial feasible solution. x_1, x_2, x_3 are called the *nonbasic variables* and x_4, x_5, x_6 *basic variables*.

Note that basic variables are expressed in terms of nonbasic ones!

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

Step 1: We note that as long as x_1 is increased by at most

$$\frac{5}{2} = \min\left(\frac{5}{2}, \frac{11}{4}, \frac{8}{3}\right),$$

all x_i remain nonnegative, but z increases.

Setting $x_1 = 5/2$ and substituting into the dictionary, we find $x_2, x_3, x_4 = 0$, $x_5 = 1$, $x_6 = 1/2$, $z = 25/2$ as an improved feasible solution.

We call x_1 the *pivot* of the iteration.

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

We can now express the variables x_1, x_5, x_6, z in terms of the *new nonbasic variables* x_2, x_3, x_4 (those currently set to zero) to obtain a new dictionary.

To do this, use line 1 of the dictionary to express x_1 in terms of x_2, x_3, x_4,

$$x_1 = \frac{1}{2}\left(5 - 3x_2 - x_3 - x_4\right)$$

and substitute the right hand side for x_1 in the remaining equations.

The new dictionary then looks as follows,

$$x_1 = \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \tag{1}$$

$$x_5 = 1 + 5x_2 + 2x_4 \tag{2}$$

$$x_6 = \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \tag{3}$$

$$z = \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4. \tag{4}$$

Of course, we are still solving

$$\max z \quad \text{s.t.} \quad x_1, \dots, x_6 \geq 0,$$

subject to the relationships (1)–(4) holding between the variables, and the new LP instance is equivalent to the old one.

However, a better feasible solution can be read off the new dictionary by setting the nonbasic variables to zero!

$$
\begin{aligned}
x_1 &= \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\
x_5 &= 1 + 5x_2 + 2x_4 \\
x_6 &= \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\
z &= \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4.
\end{aligned}
$$

Step 2: We continue in the same vein: increasing the value of x_2 or x_4 is useless, as this would decrease the objective value z.

Thus, x_3 is our pivot, and we can increase its value up to

$$1 = \min\big(5, +\infty, 1\big),$$

leading to the improved solution $x_2, x_4, x_6 = 0$, $x_1 = 2$, $x_3 = 1$, $x_5 = 1$, $z = 13$ and the dictionary corresponding to x_2, x_4, x_6 as nonbasic variables:

$$\begin{aligned} x_3 &= 1 + x_2 + 3x_4 - 2x_6 \\ x_1 &= 2 - 2x_2 - 2x_4 + x_6 \\ x_5 &= 1 + 5x_2 + 2x_4 \\ z &= 13 - 3x_2 - x_4 - x_6. \end{aligned}$$

At this point we can stop the algorithm for the following reasons:

- from the last line of the dictionary we see that for any strictly positive value of x_2, x_4 or x_6 the objective value z is necessarily strictly smaller than 13,
- and from the other lines of the dictionary we see that as soon as the values of x_2, x_4 and x_6 are fixed, the values of x_3, x_1 and x_5 are fixed too.
- Thus, the last dictionary yields a certificate of optimality for the identified solution.

Direct Computation of Dictionaries

Let us now try to understand how the dictionary

$$
\begin{aligned}
x_3 &= 1 + x_2 + 3x_4 - 2x_6 \\
x_1 &= 2 - 2x_2 - 2x_4 + x_6 \\
x_5 &= 1 + 5x_2 + 2x_4 \\
z &= 13 - 3x_2 - x_4 - x_6,
\end{aligned}
\tag{5}
$$

(which was obtained after two pivoting steps) could have been obtained directly from the input data of the original LP instance

$$
\begin{aligned}
\text{(LPI)} \quad & \max 5x_1 + 4x_2 + 3x_3 + 0x_4 + 0x_5 + 0x_6 \\
& \text{s.t. } 2x_1 + 3x_2 + x_3 + x_4 = 5 \\
& \quad 4x_1 + x_2 + 2x_3 + x_5 = 11 \\
& \quad 3x_1 + 4x_2 + 2x_3 + x_6 = 8 \\
& \quad x_1, x_2, x_3, x_4, x_5, x_6 \geq 0
\end{aligned}
$$

if we had been given the relevant basic variables:

The constraints of (LPI) imply a functional dependence between the nonnegative decision variables x_i, expressed by the linear system

$$Ax = b, \tag{6}$$

where

$$A = \begin{bmatrix} 2 & 3 & 1 & 1 & 0 & 0 \\ 4 & 1 & 2 & 0 & 1 & 0 \\ 3 & 4 & 2 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 11 \\ 8 \end{bmatrix}.$$

The basic variables of dictionary (5) are x_3, x_1, x_5. Writing

$$x_B := \begin{bmatrix} x_3 & x_1 & x_5 \end{bmatrix}^{\mathrm{T}}, \qquad x_N := \begin{bmatrix} x_2 & x_4 & x_6 \end{bmatrix}^{\mathrm{T}}$$

$$A_B := \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 3 & 0 \end{bmatrix}, \qquad A_N := \begin{bmatrix} 3 & 1 & 0 \\ 1 & 0 & 0 \\ 4 & 0 & 1 \end{bmatrix}$$

(6) can be written as

$$A_B\, x_B + A_N\, x_N = b.$$

$$A_B\, x_B + A_N\, x_N = b.$$

Solving for the basic variables x_B, we obtain

$$x_B = A_B^{-1}\left(b - A_N x_N\right). \tag{7}$$

Likewise, the objective function can be written as

$$z = c_B^{\mathrm{T}}\, x_B + c_N^{\mathrm{T}}\, x_N,$$

where

$$c_B = \begin{bmatrix}3 & 5 & 0\end{bmatrix}^{\mathrm{T}}, \quad c_N = \begin{bmatrix}4 & 0 & 0\end{bmatrix}^{\mathrm{T}},$$

and substituting from (7), we find

$$z = c_B^{\mathrm{T}} A_B^{-1} b + \left(c_N^{\mathrm{T}} - c_B^{\mathrm{T}} A_B^{-1} A_N\right) x_N.$$

Dictionary (5) is now just the system of equations

$$\begin{aligned} x_B &= A_B^{-1} b - A_B^{-1} A_N x_N, \\ z &= c_B^{\mathrm{T}} A_B^{-1} b + \left(c_N^{\mathrm{T}} - c_B^{\mathrm{T}} A_B^{-1} A_N\right) x_N. \end{aligned}$$

Definition (Dictionary)

A *dictionary* of the LP problem (P) $\max_x\{c^{\mathrm{T}}x : Ax = b, x \geq 0\}$ is a system of equations

$$
\begin{aligned}
x_B &= A_B^{-1}b - A_B^{-1}A_N x_N, \\
z &= c_B^{\mathrm{T}}A_B^{-1}b + \left(c_N^{\mathrm{T}} - c_B^{\mathrm{T}}A_B^{-1}A_N\right)x_N,
\end{aligned}
$$

equivalent to

$$
\begin{aligned}
Ax &= b, \\
z &= c^{\mathrm{T}}x,
\end{aligned}
$$

where up to column perturbation $A = \left[\, A_B \; A_N \,\right]$ and $x = \left[\, x_B^{\mathrm{T}} \; x_N^{\mathrm{T}} \,\right]^{\mathrm{T}}$ is a block decomposition such that A_B is nonsingular.

A dictionary is called *feasible* if $A_B^{-1}b \geq 0$, so that $x = (x_B, x_N) = (A_B^{-1}b, 0)$ is a feasible (but generally suboptimal) solution. (x_B, x_N) is then called a *basic feasible solution*.

Algorithm (Simplex Method for the general LP instance (P))

1. choose a basic feasible solution (x_B, x_N)
2. until $c_N - A_N^{\mathbf{T}} A_B^{-T} c_B \leq 0$, repeat

   i) $i := \min\{\ell \in N : \ c_\ell > A_\ell^{\mathbf{T}} A_B^{-T} c_B\}$ % A_ℓ is ℓ-th column of A

   ii) if $A_B^{-\mathbf{1}} A_i \leq 0$, return "(P) unbounded" % (objective $\to \infty$)

   else $j := \min\left\{\ell \in B : \ \ell \in \arg\min\left\{(A_B^{-\mathbf{1}} b)_k / (A_B^{-\mathbf{1}} A_i)_k \ : \ k \in B, (A_B^{-\mathbf{1}} A_i)_k > 0\right\}\right\}$ % (where $A_B^{-\mathbf{1}} b$ is indexed by the ordered indices in B)

   iii) $N \leftarrow N \cup \{j\} \setminus \{i\}$, $B \leftarrow B \cup \{i\} \setminus \{j\}$
3. return $(x_B, x_N) = (A_B^{-\mathbf{1}} b, 0)$ "optimal basic solution"

Tableau Format of the Simplex Method

A tableau is a representation of the problem data obtained by moving all variables x_i of a dictionary to one side and constants to the other. For example, the dictionary

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

corresponds to the tableau

$$\begin{array}{cccccc||c} 2 & 3 & 1 & 1 & 0 & 0 & 5 \\ 4 & 1 & 2 & 0 & 1 & 0 & 11 \\ 3 & 4 & 2 & 0 & 0 & 1 & 8 \\ \hline\hline 5 & 4 & 3 & 0 & 0 & 0 & 0. \end{array}$$

Basic variables can be identified via the appearance of an identity submatrix, and an optimal tableau is characterised by the appearance of all non-positive entries on the l.h.s. of the last line.

Applying the same procedure to the second dictionary

$$
\begin{aligned}
x_1 &= \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\
x_5 &= 1 + 5x_2 + 2x_4 \\
x_6 &= \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\
z &= \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4,
\end{aligned}
$$

we obtain the following tableau,

$$
\begin{array}{cccccc||c}
1 & 1.5 & 0.5 & 0.5 & 0 & 0 & 2.5 \\
0 & -5 & 0 & -2 & 1 & 0 & 1 \\
0 & -0.5 & 0.5 & -2.5 & 0 & 1 & 0.5 \\
\hline\hline
0 & -3.5 & 0.5 & -2.5 & 0 & 0 & -12.5.
\end{array}
$$

Here is how the first tableau can be directly transformed into the second one:

1. Among the columns on the left, identify one whose last entry is positive.

$$
\begin{array}{|c|ccccc||c}
2 & 3 & 1 & 1 & 0 & 0 & 5 \\
4 & 1 & 2 & 0 & 1 & 0 & 11 \\
3 & 4 & 2 & 0 & 0 & 1 & 8 \\
\hline
\hline
5 & 4 & 3 & 0 & 0 & 0 & 0.
\end{array}
$$

We call this column the *pivot column*.

2	3	1	1	0	0	5
4	1	2	0	1	0	11
3	4	2	0	0	1	8
5	4	3	0	0	0	0.

2. For each row (apart from the last) of the pivot column with positive coefficient t, look up the corresponding coefficient u on the r.h.s.

- If no such row exists, the problem is unbounded.
- If such rows exist, pick the one for which t/u is smallest and call it the *pivot row*. In this example, $t = 2$ and $u = 5$.

2	3	1	1	0	0	5
4	1	2	0	1	0	11
3	4	2	0	0	1	8
5	4	3	0	0	0	0.

3. Divide the pivot row by t,

1	1.5	0.5	0.5	0	0	2.5
4	1	2	0	1	0	11
3	4	2	0	0	1	8
5	4	3	0	0	0	0.

4. For all other rows i of the tableau, subtract the "rescaled" pivot row t_i times, where t_i is the row-i entry of the pivot colum,

1	1.5	0.5	0.5	0	0	2.5
0	-5	0	-2	1	0	1
0	-0.5	0.5	-2.5	0	1	0.5
0	-3.5	0.5	-2.5	0	0	-12.5.

In the new tableau there once again appears a (possibly permuted) identity matrix that identifies the new basic variables.

Graphic Interpretation of the Simplex Method

Consider the polyhedron $\mathscr{P} = \{x \in \mathbb{R}^n : \ Ax = b, \ x \geq 0\}$ that constitutes the feasible domain of the LP problem

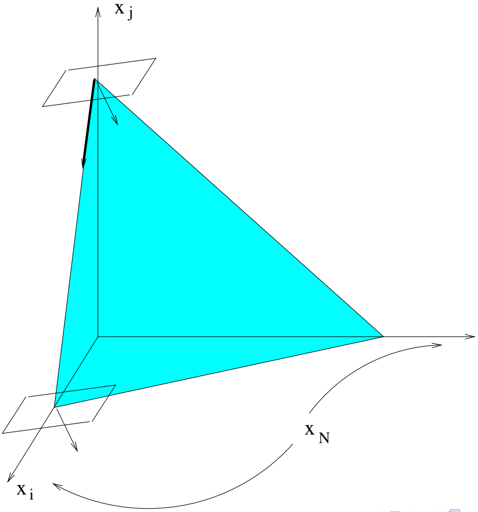
$$(\mathrm{P}) \quad \max \{c^{\mathsf{T}} x : \ Ax = b, \ x \geq 0\}.$$

Each set of nonbasic variables x_N of a basic feasible solution identifies a vertex (extreme point) of $\mathscr{P}$ via the system of equations

$$\begin{bmatrix} A_N & A_B \\ \mathrm{I} & 0 \end{bmatrix} \begin{bmatrix} x_N \\ x_B \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

Each pivoting operation changes N by only one entry i and moves to a neighbouring vertex where the objective value is greater. The move can be seen as occurring along the 1D facet

$$\{x \in \mathbb{R}^n : \ Ax = b, \ x_\ell = 0 \, \forall \ell \in N \setminus \{i\}\}.$$


x_j
x_N
x_i

Second Graphic Interpretation of the Simplex Method

A second graphic interpretation is of historic interest because it gave the method its name: Consider an LP instance in the following form,

$$
\begin{aligned}
\text{(CP)} \quad & \max\ c^{\mathrm{T}} x \\
& \text{s.t. } Ax = b, \\
& \qquad \mathbf{1}^{\mathrm{T}} x = 1, \\
& \qquad x \geq 0,
\end{aligned}
$$

where $\mathbf{1} = [\,1\ \ldots\ 1\,]^{\mathrm{T}}$.

Writing $z = c^{\mathrm{T}} x$, we define $(m+1)$-dimensional points

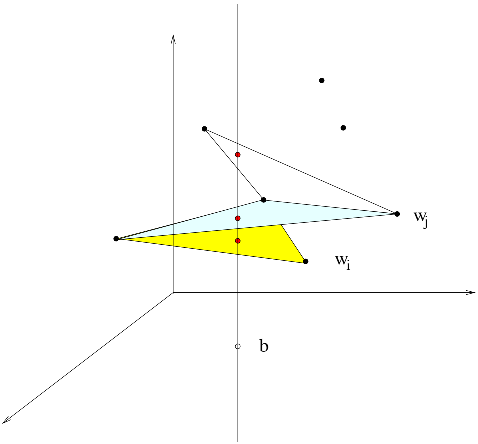
$$
v = \begin{bmatrix} b \\ z \end{bmatrix},\ w_1 = \begin{bmatrix} A_1 \\ c_1 \end{bmatrix}, \ldots, w_n = \begin{bmatrix} A_n \\ c_n \end{bmatrix},
$$

where A_i is the i-th column of A as before.
The objective of (CP) is thus to express v as a convex combination of $w_1, \ldots, w_n$ so as to maximise z.

The constraint $\mathbf{1}^{\mathsf{T}}x = 1$ implies that each basic feasible solution is associated with a collection of m points $\{w_j : j \in B\}$ for which $\{A_j : j \in B\}$ are linearly independent and contain b in their convex hull.

This is the same as requiring that the line $\{\left[\begin{smallmatrix} b \\ t \end{smallmatrix}\right] : t \in \mathbb{R}\}$ intersects the simplex $\mathrm{conv}(\{w_j : j \in B\})$.

Thus, the simplex algorithm proceeds via a sequence of simplices that are "tumbling upwards", and in each pivoting operation only one vertex of the simplex changes.


w
j
w
i
b