
1. Convex sets and corner points.

In this note, we present the mathematical background for the SIMPLEX METHOD which is used in Section 3.3: Linear Programming, without referring to LINDO, or any other software, and which allows to do necessary computations by hand in the simplest cases. As the author points out on page 75, "typical large-scale problems involve thousands of decision variables and thousands of constraints". The computational difficulties are convincingly exposed on page 77 "for a moderate size linear programming problem". But why be moderate if we want to be prepared for real challenges? Consider a very simple example of a unit cube in $\mathbb{R}^{500}$ described by $1000 = 2 \times 500$ constraints.

**Example 1.** Consider the cube

$$Q := \{ x = (x_1, x_2, \ldots, x_{500})^T \in \mathbb{R}^{500} : 0 \leq x_j \leq 1 \text{ for all } j \}.$$  \hspace{1cm} (1)

Here the pair of inequalities $0 \leq x_j \leq 1$ can be written as the couple $x_j \leq 1$ and $-x_j \leq 0$, using only inequalities $\leq$ with constants in the right side. The corner points can be described as the points $x$ with coordinates $x_j \in \{0, 1\}$, i.e. each of 500 coordinates can take values 0 or 1. Therefore, the total number of the corner points $N = 2^{500}$. Note that $2^{10} = 1024 > 10^3$, so that

$$N = (2^{10})^{50} > (10^3)^{50} = 10^{150}. \hspace{1cm} (2)$$

For comparison, consider the range of distances (in meters)

$$10^{27} > \text{Edge of Universe} > 1.5 \times 10^{11} \approx \text{Distance from Earth to Sun}$$

$$> 10^{-15} \approx \text{Radius of a nucleus}. \hspace{1cm} (3)$$

*(credit for these data is due to: The Feynman Lectures on Physics, by Feynman-Leighton-Sands, Vol. I, Sec. 5-7: Short distances). This means that if one could prepare a box large enough to put the whole universe in it, and pack it with nuclei without "gaps", then the total number of nuclei would be something like

$$N_0 = \left(10^{27}/10^{-15}\right)^3 = 10^{42.3} = 10^{126}. \hspace{1cm} (4)$$

Then counting those nuclei "one by one" would be much faster job then doing same thing for the number of corner points, or vertices of $Q$ in (1).

In general, especially in high dimensions, we need the definition a a corner point which does not refer to geometric concepts. Note that we consider the corner points only for convex sets in $\mathbb{R}^n$, therefore we define these two concepts together.

**Definition 1 (a).** For fixed $x^{(0)}$ and $x^{(1)}$ in $\mathbb{R}^n$, the closed segment

$$[x^{(0)}, x^{(1)}] := \{ x^{(t)} := (1 - t) x^{(0)} + t x^{(1)} : \quad 0 \leq t \leq 1 \}, \hspace{1cm} (3)$$

and the open segment

$$(x^{(0)}, x^{(1)}) := \{ x^{(t)} := (1 - t) x^{(0)} + t x^{(1)} : \quad 0 < t < 1 \}. \hspace{1cm} (4)$$
(b). A set \( S \subset \mathbb{R}^n \) is convex if from \( x^{(0)}, x^{(1)} \in S \) it follows that the segment \( [x^{(0)}, x^{(1)}] \) is contained in \( S \).

(c). A point \( x^* \in S \) is a corner point of a (convex) set \( S \), if

\[
\text{from } x^* \in [x^{(0)}, x^{(1)}] \subset S \text{ it follows that either } x = x^{(0)} \text{ or } x = x^{(1)} \tag{5}
\]

In other words, it is impossible to have \( x^* \in (x^{(0)}, x^{(1)}) \subset [x^{(0)}, x^{(1)}] \subset S \), which exactly means that \( x^* \) is NOT a corner point of \( S \).

**Lemma 1.** Let vectors \( a^\alpha \in \mathbb{R}^n \) and scalars \( b^\alpha \in \mathbb{R}^1 \) be defined for every \( \alpha \) in an (arbitrary) set \( A \). Then the set

\[
S := \{ x \in \mathbb{R}^n : \; (a^\alpha, x) \leq b^\alpha \text{ for all } \alpha \in A \} \text{ is convex in } \mathbb{R}^n. \tag{6}
\]

Here \( (a, x) := \sum_j a_j x_j \) for \( a := (a_1, \ldots, a_n)^T \) and \( x := (x_1, \ldots, x_n)^T \) in \( \mathbb{R}^n \).

**Proof.** Let \( x^{(0)}, x^{(1)} \in S \). Then \( x^{(t)} := (1-t)x^{(0)} + tx^{(1)} \) satisfies

\[
(a^\alpha, x^{(t)}) = (1-t)(a^\alpha, x^{(0)}) + t(a^\alpha, x^{(1)}) \leq (1-t)b^\alpha + tb^\alpha = b^\alpha
\]

for all \( 0 \leq t \leq 1 \) and \( \alpha \in A \). According to Definition 1, (a,b), the set \( S \) is convex. \( \square \)

**Remark 1.** Lemma 1 is automatically extended to the case when some of inequalities 
"\( \leq \)" in (6) are replaced by the opposite inequalities "\( \geq \)" or equalities "\( = \)". Because \( (a^\alpha, x) \geq b^\alpha \) is equivalent to
\( (-a^\alpha, x) \leq -b^\alpha \), and \( (a^\alpha, x) = b^\alpha \) is equivalent to two inequalities \( (a^\alpha, x) \leq b^\alpha \) and \( (-a^\alpha, x) \leq -b^\alpha \).

**Remark 2.** Consider the unit ball \( B := \{ x \in \mathbb{R}^n : |x| \leq 1 \} \). By taking \( a^\alpha := \alpha \in A := B \), and using the Cauchy-Schwarz inequality \( |(x, y)| \leq |x| \cdot |y| \), it is easy to show that \( B \) can be represented in the form (6) as follows:

\[
B = \{ x \in \mathbb{R}^n : (a^\alpha, x) \leq 1 \text{ for all } \alpha \in A \}.
\]

Note that every point on the boundary of \( B \) is its corner point, according to Definition 1 (c), which may not match the intuitive understanding that \( B \) is smooth and does not have corners. However, in Linear Programming problems, we deal exclusively with cases when the set \( A \) is finite. In addition, if the set \( S \) in (6) is bounded, there is the following representation of \( S \), which is convenient for some applications.

**Theorem 1.** Suppose that in (6), the set \( A \) is finite, and the set \( S \) is bounded. Then \( S \) has finitely many corner points \( x^{(1)}, x^{(2)}, \ldots, x^{(N)} \), and

\[
S = \left\{ x = \sum_{j=1}^N t_j x^{(j)} : t_j \geq 0 \text{ for all } j, \text{ and } \sum_{j=1}^N t_j = 1 \right\}. \tag{7}
\]

This theorem can be considered as a generalization of the representation (3) of the set \( S := [x^{(0)}, x^{(1)}] \) with corner points \( x^{(0)} \) and \( x^{(1)} \). The proof is not difficult, but we do not give it here, we only point out that the claim on page 77 that it is easy to verify (check if you indeed can do so) comes immediately, under the assumption that \( S \) is bounded. Namely, we have
Corollary 1. For fixed non-zero vector \( c \in \mathbb{R}^n \), the linear function \( f(x) := (c, x) \) attains its maximum on \( S \) at least at one of its corner point \( x^{(j)} \).

Proof. Indeed, according to (7), for an arbitrary \( x \in S \), we have

\[
f(x) := (c, x) = \sum_{j=1}^{N} t_j (c, x^{(j)}) \leq \sum_{j=1}^{N} t_j M = M,
\]

where \( M := \max_j (c, x^{(j)}) = \max_j f(x^{(j)}) \), and any function \( f \) always attains its maximum on a finite number of points.

Remark 3. A more direct proof of this fact, without using Theorem 1, can be proceeded in the following way. For certainty, consider the unit cube \( Q \) in \( \mathbb{R}^3 \):

\[
Q := \left\{ x = (x_1, x_2, x_3)^T \in \mathbb{R}^3 : 0 \leq x_1, x_2, x_3 \leq 1 \right\}.
\]

One can represent it as the union \( Q = S_3 \cup S_2 \cup S_1 \cup S_0 \) of disjoint sets, where \( S_k \) denotes a subset with \( k \) "degrees of freedom":

- \( S_3 \) — the interior points of \( Q \) (without the boundary);
- \( S_2 \) — the union of 6 faces (without edges);
- \( S_1 \) — the union of 12 edges (without vertices);
- \( S_0 \) — the union of 8 vertices.

Note the sets \( S_2, S_1, S_0 \) are not convex, but each of point in these sets belongs to a component (face, edge, or vertex) which is convex. Consider the problem of maximization of a linear function \( f(x) := (c, x) \), starting from the initial point \( y^{(0)} \in S_k, k \geq 1 \). Then \( y^{(0)} \) is not a corner point of \( S_k \), because \( y^{(0)} \in (x^{(0)}, x^{(1)}) \) for some \( x^{(0)}, x^{(1)} \in S_k \). Now consider the values of \( f(x) \) along the line passing through the points \( x^{(0)} \) and \( x^{(1)} \), with the origin at \( y^{(0)} \):

\[
y(t) := y^{(0)} + t \cdot h, \quad \text{where} \quad h := x^{(1)} - x^{(0)} \neq 0.
\]

Note that \( y(t) \in S_k \) for small \( |t| > 0 \). Then the function \( f(y(t)) = (c, y^{(0)}) + (c, h)t \) is non-decreasing either in the direction \( t > 0 \), or \( t < 0 \). Choose this direction, and proceed to the point \( y^{(1)} = y^{(0)} + t_1 \cdot h \) at which the line \( y(t) \) hits the boundary of \( S_k \). By this procedure, we have

\[
f(y^{(0)}) \leq f(y^{(1)}), \quad \text{and} \quad y^{(1)} \in S_j \quad \text{for some} \quad j \leq k - 1.
\]

After finite number of such steps, we arrive at some corner point \( y^* \in S_0 \), at which \( f(y^{(0)}) \leq f(y^{(1)}) \leq \cdots \leq f(y^*) \). This argument can be extended to more general case of a finite number of constraints in (6), having in mind that the number of "degrees of freedom" is reduced at least by 1 when the non-strict inequalities \( (a^\alpha, x) \leq b^\alpha \) are replaced by equalities \( (a^\alpha, x) = b^\alpha \) along the path \( y^{(0)}, y^{(1)}, \ldots \). In a more rigorous and constructive way, we will use the equivalence between geometric and algebraic definitions of a corner point in Theorem 2 below.
2. Redundant and binding restrictions.

Consider a situation when the feasible region $S$ is described by a finite number of restrictions $(a^\alpha, x) \leq b^\alpha$ or $(a^\alpha, x) = b^\alpha$ similarly to (6).

**Definition 2 (a).** A constraint $(a^\alpha, x) \leq b^\alpha$ or $(a^\alpha, x) = b^\alpha$ is called redundant is the feasible region $S$ does not change after this particular constraint is removed. Note that if three lines

$$a_1^\alpha x_1 + a_2^\alpha x_2 = b^\alpha, \quad \alpha = 1, 2, 3,$$

in $\mathbb{R}^2$ intersect at one point, then separately each of these three constraints is redundant, but if we remove more than one of them, then $S$ (which originally is just their common point of intersection) will change.

**(b).** A constraint $(a^\alpha, x) \leq b^\alpha$ is binding if in fact in combination with the remaining constraints we have $(a^\alpha, x) = b^\alpha$. All the constraints $(a^\alpha, x) = b^\alpha$ are automatically binding.

Note that there is no universal agreement on binding constraints. Some sources take into consideration the optimal solution, which means that this is a property not only of the set $S$ by itself, but also of the function $f(x)$ to be optimized. For demonstration, consider the following example.

**Example 2.** A set $S$ in $\mathbb{R}^3$ is defined by the following constraints:

\begin{align*}
4x_1 - 2x_2 - x_3 &\leq -2, \quad (1) \\
-4x_1 + 4x_2 + x_3 &\leq 5, \quad (2) \\
x_2 - 5x_3 &\leq 10, \quad (3) \\
-2x_2 &\leq -3, \quad (4) \\
-2x_1 + x_2 + x_3 &= 1. \quad (5)
\end{align*}

Determine which of these constraints are binding and which are redundant.

**Solution.** By adding together (1), (2), (4), we get $0 \leq 0$. Therefore, in these constraints we in fact have "=", and they are binding. (5) is also binding. Solution to this system is a single point $x \in \mathbb{R}^3$ with components

$$x_1 = \frac{1}{4}, \quad x_2 = \frac{3}{2}, \quad x_3 = 0.$$

In (3), we have strict inequality, so that it is not binding and redundant. Since (1), (2), (4) are dependent (the sum is 0), each of them is redundant. Any two of them together with (5) compose a system of 3 equations with 3 unknowns, none of them can be removed without changing the set of solutions $S = \{x\}$.


We start with convex sets described by a finite number of linear constraints

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i, \quad x_j \geq 0 \quad \text{for all} \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n, \quad (8)$$
where $a_{ij}$ and $b_i$ are given real numbers. Introducing the additional slack variables

$$x_{n+i} := b_i - \sum_{j=1}^{n} a_{ij}x_j \quad \text{for} \quad i = 1, 2, \ldots, m,$$

we reduce (8) to the equivalent system with $N = n + m$ variables $x_j$:

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \quad \text{for all} \quad i = 1, 2, \ldots, m; \quad x_j \geq 0 \quad \text{for all} \quad j = 1, 2, \ldots, N. \quad (9)$$

More generally, consider the system

$$\sum_{j=1}^{N} a_{ij}x_j = b_i \quad \text{for all} \quad i = 1, 2, \ldots, m; \quad x_j \geq 0 \quad \text{for all} \quad j = 1, 2, \ldots, N \geq m. \quad (10)$$

One can rewrite this system in the matrix and vector form

$$Ax = \sum_{j=1}^{N} x_j A_j = b, \quad \text{and} \quad x_j \geq 0 \quad \text{for all} \quad j = 1, 2, \ldots, N \geq m, \quad (11)$$

where $A$ is an $m \times N$ matrix with entries $a_{ij}$, which are grouped into columns $A_1, A_2, \ldots, A_N$ in $\mathbb{R}^m$, and $x \in \mathbb{R}^N$, $b \in \mathbb{R}^m$ are with coordinates $x_j, b_i$, correspondingly. We assume that the rank of $A$ is $m$, i.e.

There exists a basis in $\mathbb{R}^m$ of its columns $A_{j_1}, A_{j_2}, \ldots, A_{j_m}$. \quad (12)

Introduce the set of $m$ indices

$$J := \{1 \leq j_1 < j_2 < \cdots < j_m \leq N\} \subseteq \{1, 2, \ldots, N\}. \quad (13)$$

The set of solutions of (11) has dimension $N - m \geq 0$ (note that a single point has dimension 0). It order to get a unique solution of (11), it suffices to set $x_j = 0$ for all remaining $j \notin J$. The following theorem states that this is exactly the characterization of corner points in algebraic terms.

**Theorem 2.** Let $S$ be a convex set in $\mathbb{R}^N$ described by (10), or equivalently, by (11), where the matrix $A$ is of rank $m$. Then the following statements (I) and (II) are equivalent.

(I). A point $x^* := (x_1^*, x_2^*, \ldots, x_N^*)^T \in \mathbb{R}^N$ is a corner point of $S$, according to Definition 1(c).

(II). There exists a set $\{A_j, j \in J\}$, of $m$ linearly independent columns of $A$, such that $x_j^* = 0$ for all $j \notin J$.

**Proof.** (I) $\Rightarrow$ (II). Denote $J_0 := \{j : x_j^* > 0\}$. We claim that the vectors $\{A_j : j \in J_0\}$ are linearly independent. Indeed, otherwise we would have

$$\sum_{j \in J_0} c_j A_j = 0 \quad \text{for some constants} \quad c_j, \text{not all zeros}. \quad (14)$$
Choose $\epsilon > 0$ so small that $x_j^0 \pm \epsilon c_j > 0$ for all $j \in J_0$. Consider the point $x^{(0)}, x^{(1)}$ with coordinates

$$x_j^{(0)} := x_j^0 - \epsilon c_j, \quad x_j^{(1)} := x_j^* + \epsilon c_j \quad \text{if} \quad j \in J_0; \quad \text{and} \quad x_j^{(0)} = x_j^{(1)} = 0 \quad \text{if} \quad j \notin J_0.$$

We have $x^{(0)} \neq x^{(1)}$, and $x_j^{(0)} \geq 0$, $x_j^{(1)} \geq 0$ for all $j$. Moreover, by virtue of (14),

$$Ax^{(0)} = Ax^* - \epsilon \cdot \sum_{j \in J_0} c_j A_j = Ax^* = b, \quad \text{and similarly,} \quad Ax^{(1)} = b.$$

Therefore, we have $x^{(0)}, x^{(1)} \in S$. However, $x^* = \frac{1}{2} \cdot (x^{(0)} + x^{(1)})$, which means that $x^*$ is NOT a corner point of $S$.

This contradiction shows that $\{A_j: j \in J_0\}$ are linearly independent vectors in $\mathbb{R}^m$, hence the number of indices in $J_0$ cannot exceed $m$. If it is equal to $m$, then we set $J := J_0$. Otherwise, if it is strictly less than $m$, then one can add more indices to compose a set of $m$ vectors $\{A_j: j \in J\}$ with $J \supset J_0$ in such a way that these vectors are still linearly independent. This is possible because the rank of $A$ is $m$. This set satisfies (II).

(II) $\Rightarrow$ (I). Suppose that $x^*$ is NOT a corner point of $S$, i.e.

$$x^* = (1 - t) x^{(0)} + tx^{(1)} \quad \text{for some} \quad x^{(0)}, x^{(1)} \in S, \quad x^{(0)} \neq x^{(1)}, \quad \text{and} \quad 0 < t < 1. \quad (15)$$

Note that

$$0 = x_j^* = (1 - t) x_j^{(0)} + tx_j^{(1)} \quad \text{for all} \quad j \notin J.$$

Together with $x_j^{(0)} \geq 0$, $x_j^{(1)} \geq 0$, and $0 < t < 1$, this implies $x_j^{(0)} = x_j^{(1)} = 0$ for all $j \notin J$. Then the vector

$$h := (h_1, h_2, \ldots, h_N)^T := x^{(1)} - x^{(0)} \neq 0, \quad \text{and} \quad h_j = 0 \quad \text{for all} \quad j \notin J.$$

Further, note that all the point $x^*, x^{(0)}, x^{(1)}$ satisfy $Ax = b$. Now from (15) it follows $t \cdot h = x^* - x^{(0)}$, and

$$t \cdot \sum_{j \in J} h_j A_j = A (t \cdot h) = A (x^* - x^{(0)}) = Ax^* - Ax^{(0)} = b - b = 0.$$

Since $t > 0$, the sum in the left side is zero for some non-zero $h_j$. This means that the vectors $\{A_j, j \in J\}$ are linearly dependent, contrary to our assumption. This contradiction shows that $x^*$ is a corner point in terms of (I).

4. Elementary row operations (ERO).

Here we review the key elements of the Gauss-Jordan procedure for solving systems (10), which can be written in the matrix form as $Ax = b$ in (11), where $A$ is a $m \times N$ matrix, $x \in \mathbb{R}^N, b \in \mathbb{R}^m$.

All the numerical information about this system is contained in its augmented matrix $[A \mid b]$ of size $m \times (N + 1)$. Using the following elementary row operations (ERO), this matrix is reduced to more simple (diagonal, triangle, or echelon) form.

1. $R_i \rightarrow \lambda R_i$. Multiply the $i^{th}$ row by a constant $\lambda \neq 0$.
2. $R_i \rightarrow R_i + \lambda R_j, \quad i \neq j$. Add the $j^{th}$ row multiplied by $\lambda$ to the $i^{th}$ row.
3. $R_i \leftrightarrow R_j, \quad i \neq j$. Interchange the $i^{th}$ and $j^{th}$ rows.

These operations are not independent: $R_i \leftrightarrow R_j$ can be obtained as the result of a sequence:

(1) $R_i \rightarrow R_i + R_j$; (2) $R_j \rightarrow R_j - R_i$; (3) $R_i \rightarrow R_i + R_j$; (4) $R_j \rightarrow -R_j$. 

6
You can keep track of results of each step in \(i^{th}\) and \(j^{th}\) rows as follows:

\[
\begin{bmatrix}
R_i \\
R_j
\end{bmatrix} \xrightarrow{(1)} \begin{bmatrix}
R_i + R_j \\
R_j
\end{bmatrix} \xrightarrow{(2)} \begin{bmatrix}
R_i + R_j \\
-R_i
\end{bmatrix} \xrightarrow{(3)} \begin{bmatrix}
R_i \\
-R_i
\end{bmatrix} \xrightarrow{(4)} \begin{bmatrix}
R_j \\
R_i
\end{bmatrix}.
\]

Each if these operations corresponds to multiplication by a certain elementary matrix. For illustration, consider the following examples.

**Example 3.** (i). \(R_2 \rightarrow \lambda R_2\).

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & \lambda & 0 \\
0 & 0 & 1
\end{bmatrix} \cdot \begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix} = \begin{bmatrix}
1 & 2 & 3 \\
4\lambda & 5\lambda & 6\lambda \\
7 & 8 & 9
\end{bmatrix}.
\]

(ii). \(R_1 \rightarrow R_1 + \lambda R_3\).

\[
\begin{bmatrix}
1 & \lambda & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \cdot \begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix} = \begin{bmatrix}
1 + 7\lambda & 2 + 8\lambda & 3 + 9\lambda \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix}.
\]

(iii). \(R_2 \leftrightarrow R_3\).

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix} \cdot \begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix} = \begin{bmatrix}
1 & 2 & 3 \\
7 & 8 & 9 \\
4 & 5 & 6
\end{bmatrix}.
\]

Note that the elementary matrices \(E\) are square \(m \times m\) matrices which can be applied to more general rectangle \(m \times N\) matrices \(A\) in such a way that \(j^{th}\) column \(A_j\) is replaced by \(EA_j\). In particular, if \(A\) is a non-degenerate \(m \times m\) matrix, a sequence of ERO which transforms the \(m \times (2m)\) matrix \([A | I_m]\) to \([I_m | B]\) can be described as \(E_k \cdots E_2 E_1 \cdot [A | I_m] = [I_m | B]\), which means that

\[
I_m = E_k \cdots E_2 E_1 A = BA, \quad \text{so that} \quad B = A^{-1}.
\]

The following example combines together a few techniques.

**Example 4.** Find the exact expression for \(A^{10}\), where \(A = \begin{bmatrix}
4 & 3 \\
1 & 2
\end{bmatrix}\).

**Step 1.** Find the eigenvalues \(\lambda\) of \(A\). They are roots of the characteristic equation

\[
p(\lambda) = \det (\lambda \cdot I - A) = \det \begin{bmatrix}
\lambda - 4 & -3 \\
-1 & \lambda - 2
\end{bmatrix} = (\lambda - 4) (\lambda - 2) - 3 = \lambda^2 - 6\lambda + 5 = (\lambda - 1) (\lambda - 5),
\]

so that the eigenvalues are \(\lambda_1 = 1\) and \(\lambda_2 = 5\).

**Step 2.** Find the corresponding eigenvectors \(v = (\alpha, \beta)^T\). They satisfy \(Av = \lambda v = \lambda I v\), i.e. \((A - \lambda I) v = 0\). For \(\lambda = \lambda_1, \lambda = \lambda_2\) we find the corresponding eigenvectors \(v_1, v_2\) as follows:

\[
\begin{bmatrix}
3 & 3 \\
1 & 1
\end{bmatrix} v_1 = 0, \quad v_1 = \begin{bmatrix}
\alpha_1 \\
\beta_1
\end{bmatrix} = \begin{bmatrix}
1 \\
-1
\end{bmatrix}; \quad \begin{bmatrix}
-1 & 3 \\
1 & -3
\end{bmatrix} v_2 = 0, \quad v_2 = \begin{bmatrix}
\alpha_2 \\
\beta_2
\end{bmatrix} = \begin{bmatrix}
3 \\
1
\end{bmatrix}.
\]
Step 3. Introduce the matrix $V$ with columns $v_1$ and $v_2$. Then $AV$ has columns $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$, which can be written as

$$AV = VA, \quad \text{where} \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \quad \text{hence} \quad A = V \Lambda V^{-1}.$$  

In turn, this yields a convenient representation

$$A^{10} = V \Lambda V^{-1} \cdot V \Lambda V^{-1} \cdot \ldots \cdot V \Lambda V^{-1} = V \Lambda^{10} V^{-1}$$

(10 times)

$$= V \cdot \begin{bmatrix} \lambda_1^{10} & 0 \\ 0 & \lambda_2^{10} \end{bmatrix} \cdot V^{-1} = V \cdot \begin{bmatrix} 1 & 0 \\ 0 & 5^{10} \end{bmatrix} \cdot V^{-1}$$

(17)

Step 4. For completion, we need to find $V^{-1}$. For $2 \times 2$ matrices, this is easy:

$$V = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \quad \text{has inverse matrix} \quad V^{-1} = \frac{1}{\det V} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & -3 \\ 1 & 1 \end{bmatrix}.$$  

However, for higher dimensions, the method in (16) is more efficient. In our case, it produces

$$[V \mid I] = \begin{bmatrix} 1 & 3 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{\ (R_2 \rightarrow R_2 + R_1) \ } \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & 4 & 1 & 1 \end{bmatrix} \xrightarrow{\ (R_1 \rightarrow 4R_1) \ } \begin{bmatrix} 4 & 12 & 4 & 0 \\ 0 & 4 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow{\ (R_1 \rightarrow R_1 - 3R_2) \ } \begin{bmatrix} 4 & 0 & 1 & -3 \\ 0 & 4 & 1 & 1 \end{bmatrix} \quad \implies \quad V^{-1} = \frac{1}{4} \begin{bmatrix} 1 & -3 \\ 1 & 1 \end{bmatrix}.$$  

Step 5. Finally, we can now rewrite (17) as

$$A^{10} = V \cdot \begin{bmatrix} 1 & 0 \\ 0 & 5^{10} \end{bmatrix} \cdot V^{-1} = \frac{1}{4} \cdot \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 5^{10} \end{bmatrix} \cdot \begin{bmatrix} 1 & -3 \\ 1 & 1 \end{bmatrix}$$

$$= \frac{1}{4} \cdot \begin{bmatrix} 1 & -3 \\ -1 & 3 \end{bmatrix} + \frac{5^{10}}{4} \cdot \begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix}.$$  

Note that in Step 4, non-zero diagonal elements were used as pivot elements to annihilate the remaining non-zero elements in the same column. A similar feature is present in the Simplex Method below.

5. Application of ERO to the Simplex Method.

Let a non-zero vector $c = (c_1, c_2, \ldots, c_N)^T \in \mathbb{R}^N$ be fixed. Consider the problem of maximization of the objective function

$$y = f(x) := (c, x) = \sum_{j=1}^{N} c_j x_j$$

(18)

on a set $S$ (the feasible region), which is described, as in (10–12), by a system of constraints

$$Ax = \sum_{j=1}^{N} x_j A_j = b, \quad \text{and} \quad x_j \geq 0 \quad \text{for all} \quad j = 1, 2, \ldots, N \geq m,$$  

(19)
where $A$ is an $m \times N$ matrix of rank $m$. The relations (18) and (19) can be combined together as the matrix representation of a system with a matrix of size $(m + 1) \times (N + 1)$

$$
\begin{bmatrix}
1 & -c^T \\
0_m & A
\end{bmatrix}
\begin{bmatrix}
y \\
x
\end{bmatrix} =
\begin{bmatrix}
0 \\
b
\end{bmatrix},
$$

(20)

where $0_m$ is a zero vector in $\mathbb{R}^m$. It is convenient to count numeration of rows and columns starting from 0. Then $a_{ij}$ in an element in $i^{th}$ row and $j^{th}$ column not only for $A$, but also for this extended matrix. In this matrix form, the $0^{th}$ equation corresponds to (18):

$$
1 \cdot y - c^T x = y - (c, x) = 0.
$$

Throughout the rest of this section, we assume that the system (19) is non-degenerate.

**Definition 3.** The system $S$ in (19) is non-degenerate is the matrix $A$ of size $m \times N$ has rank $m \leq N$, and for every corner point $x^* = (x^*_1, x^*_2, \ldots, x^*_N)^T$ of $S$, there exists a set of $m$ indices $J := \{1 \leq j_1 < j_2 < \cdots < j_m \leq N\}$, such that the columns $\{A_{j}, j \in J\}$ of the matrix $A$ are linearly independent, and

$$
x^*_j > 0 \text{ for all } j \in J, \quad x^*_j = 0 \text{ for all } j \notin J.
$$

(21)

Note that the part (II) if Theorem 2 claims only the second part of (21), whereas $x^* \in S$ implies that $x^*_j \geq 0$ for all $j$. This means that the system $S$ is degenerate if some of strict inequalities ”>” in (21) are replaced by equalities ”=” , so that such points belong to subsets of lower dimension rather than specified by (21).

**Remark 4.** If the objective function is originally defined as $f(x) := (c, x)$ with $c, x \in \mathbb{R}^n$, on a set described by $m$ inequalities in (8), then after reducing to to the system of $N = m + n$ equalities (9), we automatically get a corner point $x^{(0)}$ with $J = \{n + 1, n + 2, \ldots, n + m = N\}$. Then the assumption $x^{(0)}_j = 0$ for all $j \notin J$, i.e. for all $j = 1, 2, \ldots, n$, implies that $y^{(0)} := f(x^{(0)}) = (c, x^{(0)}) = 0$. This is not necessarily the case for $f$ and $S$ in (18), (19).

We can apply the ERO to the augmented matrix for the system (20) presented as the table

$$
T :=
\begin{bmatrix}
1 & -c^T \\
0_m & A
\end{bmatrix} \begin{bmatrix}
0 \\
b
\end{bmatrix}
$$

(22)

Let $x^{(0)}$ be a corner point of $S$ in (19), with the corresponding set $J$ of $m$ indices, i.e. $x^{(0)}_j > 0$ for all $j \in S$ (by our assumption of non-degeneracy in (21)), and $x^{(0)}_j = 0$ for all $j \notin S$. The variables $x_j$ for $j \in J$ are called basic variables, because the columns $A_j$, $j \in J$, of the matrix $A$ compose a basis in $\mathbb{R}^m$, so that these variables can be uniquely defined from the system $Ax = \sum x_j A_j = b$ in terms of remaining non-basic variables and constants. Then the $(m + 1) \times (m + 1)$ matrix

$$
\begin{bmatrix}
1 & -c_j^T \\
0_m & A_j
\end{bmatrix}
$$

with $m + 1$ columns $\begin{bmatrix}1 \\
0_m
\end{bmatrix}$ and $\begin{bmatrix}-c_j \\
A_j
\end{bmatrix}$ for $j \in J$, (23)
has determinant \( \det A_j \neq 0 \), and therefore, after a finite number of ERO can be reduced to \( I_{m+1} \) – the unit \( (m+1) \times (m+1) \) matrix. As it was mentioned before fromula (16), these ERO applied to the table (22) produce another table

\[
T^{(0)} := \begin{bmatrix}
1 & -c_j^T \\
0_m & A_j
\end{bmatrix}^{-1} \cdot \begin{bmatrix}
1 & -c^T \\
0_m & A
\end{bmatrix} = \begin{bmatrix}
1 & -c_j^T \\
1 & c^T
\end{bmatrix} A^{(0)} \begin{bmatrix}
y^{(0)} \\
x_j^{(0)}
\end{bmatrix}.
\]

Note that the last column contains the vector \( x_j^{(0)} \in \mathbb{R}^m \) which is composed of components \( x_j^{(0)} > 0 \) for \( j \in J \), and \( y^{(0)} = f(x^{(0)}) = (c, x^{(0)}) \). Indeed, we know that the systems corresponding to the tables \( T \) and \( T^{(0)} \) are equivalent, i.e. they have same set of solutions. We also know that \( x = x^{(0)} \), \( y = y^{(0)} \) satisfy the system for \( T \). The system for \( T^{(0)} \) is

\[
y - (c^{(0)}, x) = y^{(0)}, \quad A^{(0)} x = x_j^{(0)}.
\]

Note that as the result of matrix multiplication in (24) the columns in (23) have been replaced by those of \( I_{m+1} \), which implies that the "new" values \( c_j^{(0)} = 0 \) for all \( j \in J \). We also have \( x_j^{(0)} = 0 \) for all \( j \notin J \). Therefore \( (c^{(0)}, x^{(0)}) = \sum c_j^{(0)} x_j^{(0)} = 0 \), and the pair \( x = x^{(0)}, y = y^{(0)} \) satisfies the first equality in (25). Similarly, \( A^{(0)} \) restricted to columns \( A_j, j \in J \), is just the unit matrix \( I_m \), hence \( x = x^{(0)} \) satisfies the second equality in (25).

Since \( c_j^{(0)} = 0 \) for all \( j \in J \), from (25) it follows

\[
y = f(x) = y^{(0)} + (c^{(0)}, x) = y^{(0)} + \sum_{j \notin J} c_j^{(0)} x_j.
\]

We also have \( x_j^{(0)} = 0 \) for all \( j \notin J \), whereas \( x_j \geq 0 \) for \( x \in S \). Thus we have the following easy case.

(I). If \( c_k^{(0)} \leq 0 \) for all \( k \notin J \), then \( y = f(x) \) attains its maximum \( y^{(0)} = f(x^{(0)}) \) on \( S \) at the point \( x^{(0)} \in S \).

Now suppose that (I) fails, and fix some \( k \notin J \) such that \( c_k^{(0)} > 0 \). Then \( x^{(0)} \) certainly cannot be the point of maximum of \( f \), because the value of \( y \) increases when we increase the value of \( x_k \) from \( x_k^{(0)} = 0 \) to \( x_k > 0 \). This will affect the values of basic variables \( x_j, j \in S \), because they are tied up by the relation \( Ax = b \), but because of non-degeneracy assumption (21) for \( x^* = x^{(0)} \), for small \( x_k > 0 \) the change of basic coordinates will also be small to guarantee \( x_j \geq 0 \) for all \( j \). We cannot do so if \( x_j^{(0)} = 0 \) for some \( j \in S \), and this is exactly the point at which we need the non-degeneracy condition on the corner points of \( S \).

We split the alternative case \( c_k^{(0)} > 0 \) for some \( k \notin J \) between the following cases (II) and (III). Note that the case (II) can appear only if \( S \) is unbounded.

(II). If \( c_k^{(0)} > 0 \) for some \( k \notin J \), and the \( j \)th column \( A_k^{(0)} \) of the matrix \( A^{(0)} \) contains only non-positive elements:

\[
a_{ik}^{(0)} \leq 0 \quad \text{for all} \quad i,
\]

then \( y = f(x) \) is unbounded on \( S \), i.e. \( y \) can attain arbitrarily large values on \( S \).

Indeed, consider the points

\[
x = (x_1, x_2, \ldots, x_N)^T \in S \quad \text{with} \quad x_k > 0, \quad \text{and} \quad x_j = 0 \quad \text{for all} \quad j \notin J.
\]
Then according to (26),
\[ y = y^{(0)} + c_k^{(0)} x_k \]
can be made arbitrarily large for large \( x_k > 0 \). However, we need to make sure that \( x_j \geq 0 \) for all \( j \in J \) to guarantee that \( x \in S \).

Since (25) holds true for all \( x \in S \), including \( x^{(0)} \), we have
\[
A^{(0)} (x - x^{(0)}) = A^{(0)} x - A^{(0)} x^{(0)} = x^{(0)} - x^{(0)} = 0,
\]
so that for every \( i \), we have
\[
0 = \sum_j a_{ij}^{(0)} (x_j - x_j^{(0)}) = \Sigma_1 + \Sigma_2, \tag{29}
\]
where \( \Sigma_1 \) and \( \Sigma_2 \) are sums extended over \( j \in J \) and \( j \notin J \), correspondingly. However, as the result of matrix multiplication in (24), the columns \( A_j^{(0)} \) of \( A^{(0)} \) for \( j \in J \) compose the unit matrix \( I_m \). Therefore, \( \Sigma_1 = x_i - x_i^{(0)} \). In addition, by the choice of \( x \) in (28), we also have \( \Sigma_2 = a_{ik}^{(0)} x_k \leq 0 \). These relations provide the desired property
\[
x_i = x_i^{(0)} + \Sigma_1 = x_i^{(0)} - \Sigma_2 \geq 0, \quad \text{i.e.} \quad x \in S.
\]

Now it remains to consider the case when both (I) and (II) fail.

(III). \( c_k^{(0)} > 0 \) for some \( k \notin J \), and the set \( P := \{ i : a_{ik}^{(0)} > 0 \} \) is non-empty.

Note that the first column of \( T^{(0)} \) in (24) together with \( j^{th} \) columns for \( j \in J \), compose the set of standard unit basis vectors in \( \mathbb{R}^{m+1} \);

\[
e_0 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_1 := \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad e_{m-1} := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \quad e_m := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{30}
\]

The idea is to apply the ERO to the table \( T^{(0)} \) in (24) with the pivot entry \( a_{rk}^{(0)} \) for some \( r \in P \) in such a way that the column \( e_r \) will be moved from \( j^{th} \) column for some \( j \in J \) to \( k^{th} \) column. In the new set of basic variables, \( x_k \) will take place of \( x_j \). Geometrically, we move from the corner point \( x^{(0)} \) with \( x_k^{(0)} = 0 \) to another corner point \( x^{(1)} \) with \( x_j^{(1)} = 0 \), along the one-dimensional line passing through these two corner points. In a new set \( J \), we will have the same set of vectors \( e_1, e_2, \ldots, e_m \) in (30) in a different order. In general, only \( e_0 \) stays permanently at \( 0^{th} \) column, but the order of remaining \( e_i \) is not important.

Using notations in Section 4, these ERO can be specified as follows, with \( a_{0k}^{(0)} := -c_k^{(0)} \):

\[
R_r \rightarrow \frac{1}{a_{rk}^{(0)}} \cdot R_r; \quad \text{and} \quad R_i \rightarrow R_i - \lambda_i R_r \quad \text{for} \quad 0 \leq i \neq r, \quad \text{where} \quad \lambda_i := \frac{a_{ik}^{(0)}}{a_{rk}^{(0)}}. \tag{31}
\]
We know that at the end we will have a new table \( T^{(1)} \) with \( e_r \) in place of the \( k \)th column of \( T^{(0)} \), call it \( f_k := [-c_k^{(0)}, A_k^{(0)}]^T \), and \( e_i \) for all \( i \neq r \) remain unchanged. Let \( M \) be the matrix obtained from the unit matrix \( I_{m+1} \) by replacing its \( r \)th column \( e_r \) by \( f_k \). Then

\[
Me_r = f_k, \quad \text{and} \quad Me_i = e_i \quad \text{for all} \quad i \neq r,
\]

\[
M^{-1}f_k = e_r, \quad \text{and} \quad M^{-1}e_i = e_i \quad \text{for all} \quad i \neq r.
\]

These properties uniquely define the resulting table as

\[
T^{(1)} := M^{-1} \cdot T^{(0)} = \left[ \begin{array}{c|c|c|c|c|c|c|c|c} 
1 & -400 & -200 & -250 & 0 & 0 & 0 & 0 \\
0 & 3 & 1 & 1.5 & 1 & 0 & 0 & 1000 \\
0 & 0.8 & 0.2 & 0.3 & 0 & 1 & 0 & 300 \\
0 & 1 & 1 & 1 & 0 & 0 & 1 & 625 \\
\end{array} \right].
\]

In order to guarantee that \( x^{(1)} \in S \), we must have

\[
0 \leq x_i^{(1)} = x_i^{(0)} - \lambda_i x_r^{(0)} = x_i^{(0)} - \frac{a_{ik}^{(0)} x_r^{(0)}}{a_{rk}^{(0)}} \quad \text{for all} \quad i \geq 1, i \neq r.
\]

This is automatically true if \( a_{ik}^{(0)} \leq 0 \). In order to have these restrictions for all \( i \geq 1 \), we must choose \( r \in P \) from the relation

\[
\frac{x_r^{(0)}}{a_{rk}^{(0)}} = \min_{i \in P} \frac{x_i^{(0)}}{a_{ik}^{(0)}}, \quad \text{where} \quad P := \{ i : a_{ik}^{(0)} > 0 \}.
\]

Under this choice of \( r \), we will come from \( x^{(0)} \) to another corner point \( x^{(1)} \) of \( S \). As in (24), we automatically have \( y^{(1)} = f(x^{(1)}) \), and according to (31) with \( i = 0 \),

\[
y^{(1)} = y^{(0)} - \lambda_0 x_r^{(0)} = y^{(0)} + \frac{c_k^{(0)} x_r^{(0)}}{a_{rk}^{(0)}} > y^{(0)}.
\]

After a finite number of similar iterations, we eventually come to one of cases (I) or (II).

**Example 3.4** in the textbook. We can take \( x^{(0)} = [0, 0, 0, 1000, 300, 625]^T \) as the initial corner point for

\[
T = T^{(0)} = \left[ \begin{array}{c|c|c|c|c|c|c|c} 
1 & -400 & -200 & -250 & 0 & 0 & 0 & 0 \\
0 & 3 & 1 & 1.5 & 1 & 0 & 0 & 1000 \\
0 & 0.8 & 0.2 & 0.3 & 0 & 1 & 0 & 300 \\
0 & 1 & 1 & 1 & 0 & 0 & 1 & 625 \\
\end{array} \right].
\]

Choose \( k = 1 \). According to (P), from the relation

\[
\min \{1000/3, 300/0.8, 625\} = 1000/3,
\]

we get \( r = 1 \), so that the pivot element is \( a_{11} = 3 \). Applying the ERO, we obtain

\[
T^{(1)} = T^{(0)} = \left[ \begin{array}{c|c|c|c|c|c|c|c} 
1 & -400 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4/5 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\end{array} \right] \cdot T^{(0)} = \left[ \begin{array}{c|c|c|c|c|c|c|c} 
1 & 400/3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -4/15 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & -1/3 & 0 & 1 & 0 & 0 & 0 & 0 \\
\end{array} \right] \cdot T^{(0)}
\]

\[
= \left[ \begin{array}{c|c|c|c|c|c|c|c} 
1 & 0 & -200/3 & -50 & 400/3 & 0 & 0 & 400 \cdot 000/3 \\
0 & 1 & 1/3 & 1/2 & 1/3 & 0 & 0 & 1000/3 \\
0 & 0 & -1/15 & -1/10 & -4/15 & 1 & 0 & 100/3 \\
0 & 0 & 2/3 & 1/2 & -1/3 & 0 & 1 & 875/3 \\
\end{array} \right].
\]

12
with the new corner point \( x^{(1)} = [1000/3, 0, 0, 0, 1000/3, 875/3]^T \). Here the matrix
\[
M^{-1} = \begin{bmatrix}
1 & -400 & 0 & 0 \\
0 & 3 & 0 & 0 \\
0 & 4/5 & 1 & 0 \\
0 & 1 & 0 & 1
\end{bmatrix}^{-1} = \begin{bmatrix}
1 & 400/3 & 0 & 0 \\
0 & 1/3 & 0 & 0 \\
0 & -4/15 & 1 & 0 \\
0 & -1/3 & 0 & 1
\end{bmatrix}
\]
is ”responsible” for the following sequence of ERO applied to \( T^{(0)} \) in order to get \( T^{(1)} \):
\[
\begin{align*}
(1) \quad & R_1 \quad \rightarrow \quad 1 \cdot R_1 + \frac{400}{3} \cdot R_2; \\
(2) \quad & R_2 \quad \rightarrow \quad \frac{1}{3} \cdot R_2; \\
(3) \quad & R_3 \quad \rightarrow \quad -\frac{4}{15} \cdot R_2 + R_3; \\
(4) \quad & R_4 \quad \rightarrow \quad -\frac{1}{3} \cdot R_2 + R_4.
\end{align*}
\]

Now we can take \( k = 2 \). Since
\[
\min \{1000, 875/2\} = 875/2,
\]
we get \( r = 3 \), with the pivot element \( a_{32} = 2/3 \). The last iteration is
\[
T^{(2)} = \begin{bmatrix}
1 & 0 & -200/3 & 0 \\
0 & 1 & 1/3 & 0 \\
0 & 0 & -1/15 & 1 \\
0 & 0 & 2/3 & 0
\end{bmatrix}^{-1} \cdot T^{(1)} = \begin{bmatrix}
1 & 0 & 0 & 100 \\
0 & 1 & 0 & -1/2 \\
0 & 0 & 0 & 3/2 \\
0 & 0 & 1 & 1/10
\end{bmatrix}. T^{(1)}
\]
\[
= \begin{bmatrix}
1 & 0 & 0 & 0 & 100 & 0 & 100 & 162500 \\
0 & 1 & 0 & 1/4 & 1/2 & 0 & -1/2 & 375/2 \\
0 & 0 & 1 & 3/4 & -1/2 & 0 & 3/2 & 875/2 \\
0 & 0 & 0 & -1/20 & -3/10 & 1 & 1/10 & 125/2
\end{bmatrix}
\]
\[
= \begin{bmatrix}
1 & 0 & 0 & 0 & 100 & 0 & 100 & 162500 \\
0 & 1 & 0 & 0.25 & 0.5 & 0 & -0.5 & 187.5 \\
0 & 0 & 1 & 0.75 & -0.5 & 0 & 1.5 & 437.5 \\
0 & 0 & 0 & -0.05 & -0.3 & 1 & 0.1 & 62.5
\end{bmatrix}
\]
with the corner point \( x^{(2)} := [187.5, 437.5, 0, 0, 62.5, 0]^T \). Since in the last table we have \(-c^T \geq 0\), this is the case (I), which guarantees that
\[
x_1 = 187.5, \quad x_2 = 437.5, \quad x_3 = 0
\]
is the optimal solution with the maximal profit of $162,500, with the slack reserve \( x_5 = 62.5 \).

These data match with results on p.79 in the textbook.

**Remark 5.** Similarly to Step 5 in Example 5 (page 8), in calculation ”by hand” in order to avoid operations with fractions, it is convenient to use instead of vector \( e_i \) in (30) the vectors \( f_i \) with large enough interders \( d_i \) in place of 1. The solutions remain the same, because the ratios in \((P)\) contain the elements from same row, so that the \( i^{th} \) ratio does not change after dividing \( i^{th} \) row by \( d_i \) in order to reduce it to the case \( d_i = 1 \).