The problem of optimization is to minimize or maximize a function, namely the objective function subject to certain conditions known as constraints. If the objective function is a linear function say from $\mathbb{R}^n$ to $\mathbb{R}$ and the constraints are also given by linear functions, then the problem is called a linear programming problem.

Given $\mathbf{c} \in \mathbb{R}^n$, a column vector with n components, $\mathbf{b} \in \mathbb{R}^m$, a column vector with m components, and an $A \in \mathbb{R}^{m\times n}$, a matrix with m rows and n columns.
A linear programming problem is given by :
Max or Min $\mathbf{c}^T\mathbf{x}$
subject to $A\mathbf{x} \leq \mathbf{b}$ (or $A\mathbf{x} \geq \mathbf{b}$),
$\mathbf{x} \geq \mathbf{0}$.
The function $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ is called the objective function, the constraints $\mathbf{x} \geq \mathbf{0}$ are called the nonnegativity constraints.
Note that the above problem can also be written as:
Max or Min $\mathbf{c}^T\mathbf{x}$
$\mathbf{a}_i^T\mathbf{x} \leq b_i$ for $i = 1, 2, \ldots, m$,
$x_j \geq 0$ for $j = 1, 2, \ldots, n$, or $-\mathbf{e}_j{}^T\mathbf{x} \leq 0$ for all $j = 1, 2, \ldots, n$,
where $\mathbf{a}_i^T$ is the i th row of the matrix A, and $\mathbf{e}_j$ is the j th column of the identity matrix of order n, I_n.

Note that each of the functions $\mathbf{a}_i{}^T\mathbf{x}$, for $i = 1, 2 \ldots, m$, $-\mathbf{e}_j{}^T\mathbf{x}$, for $j = 1, 2, \ldots, n$, and $\mathbf{c}^T\mathbf{x}$ are all linear functions (refer to (*)) from $\mathbb{R}^n :\to \mathbb{R}$ (check this), hence the name **linear programming problem**.
((*) A function $T : \mathbb{R}^n \to \mathbb{R}$ is called a linear map (linear function, linear transformation) if $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $T(\alpha\mathbf{x}) = \alpha T(\mathbf{x})$ for all $\alpha \in \mathbb{R}$ and all $\mathbf{x} \in \mathbb{R}^n$).
An $\mathbf{x} \geq \mathbf{0}$ satisfying the constraints $A\mathbf{x} \leq \mathbf{b}$ (or $A\mathbf{x} \geq \mathbf{b}$) is called a **feasible solution** of the linear programming problem (LPP). The set of all feasible solutions of a LPP is called the feasible solution set or the **feasible region** of the LPP.
Hence the feasible region of a LPP, denoted by Fea(LPP)is given by,
Fea(LPP)= $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0, A\mathbf{x} \leq \mathbf{b}\}$.
A feasible solution of an LPP is said to be an **optimal solution** if it minimizes or maximizes the objective function, depending on the nature of the problem. The set of all optimal solutions is called the optimal solution set of the LPP. If the LPP has an optimal solution, then the value of the objective function $\mathbf{c}^T\mathbf{x}$ when $\mathbf{x}$ is an optimal solution of the LPP is called the **optimal value** of the LPP.

Note that as discussed before, the feasible region can also be written as
Fea(LPP)= $\{\mathbf{x} \in \mathbb{R}^n : x_j \geq 0 \text{ for } j = 1, 2, \ldots, n, \quad \mathbf{a}_i^T\mathbf{x} \leq b_i \text{ for } i = 1, 2, \ldots, m\}$ or as
$= \{\mathbf{x} \in \mathbb{R}^n : -\mathbf{e}_j^T\mathbf{x} \leq 0 \text{ for } j = 1, \ldots, n, \quad \mathbf{a}_i^T\mathbf{x} \leq b_i \text{ for } i = 1, \ldots, m\}$,
where $\mathbf{a}_i^T$ is the i th row of the matrix A, and $\mathbf{e}_j$ is the jth standard unit vector, or the j th column of the identity matrix I_n.

Definition: A subset H of $\mathbb{R}^n$ is called a hyperplane if it can be written as:
H= $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = d\}$ for some $\mathbf{a} \in \mathbb{R}^n$ and $d \in \mathbb{R}$, or equivalently as
H= $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0) = 0\}$ for some $\mathbf{a} \in \mathbb{R}^n$, $d \in \mathbb{R}$, and $\mathbf{x}_0$ satisfying $\mathbf{a}^T\mathbf{x}_0 = d$. (**)
So geometrically a hyperplane in $\mathbb{R}$ is just an element of $\mathbb{R}$ (a single point), in $\mathbb{R}^2$ it is just a straight line, in $\mathbb{R}^3$ it is just the usual plane we are familiar with.
The vector $\mathbf{a}$ is called a **normal** to the hyperplane H, since it is orthogonal (or perpendicular) to each of the vectors on the hyperplane starting from (with tail at) $\mathbf{x}_0$ (for precision refer to (**)).
A collection of hyperplanes $H_1, \ldots, H_k$ in $\mathbb{R}^n$ are said to be **Linearly Independent (LI)** if the

corresponding normal vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent as vectors in $\mathbb{R}^n$. Otherwise the hyperplanes are said to be **Linearly Dependent (LD)**.
Hence any set of $(n+1)$ hyperplanes in $\mathbb{R}^n$ is LD.

Associated with the hyperplane H are two **closed half spaces** $H_1 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} \leq d\}$ and $H_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} \geq d\}$.
Note that the hyperplane H, and the two half spaces H_1, H_2 are all closed subsets of $\mathbb{R}^n$ (since each of these sets contains all its boundary points, the boundary points being $\mathbf{x} \in \mathbb{R}^n$ satisfying the condition $\mathbf{a}^T\mathbf{x} = d$).
Hence the feasible region of a LPP is just the intersection of a finite number of closed half spaces, which is called a **polyhedral set**.
Definition: A set which is the intersection of a finite number of closed half spaces is called a **polyhedral set**.
Hence the feasible region of a LPP is a **polyhedral set**.
Since the intersection of any collection of closed subsets of $\mathbb{R}^n$ is again a closed subset of $\mathbb{R}^n$ hence Fea(LPP) is always a **closed** subset of $\mathbb{R}^n$ (also geometrically you can see (or guess) that the feasible region of a LPP contains all its boundary points).

Let us consider two very familiar real life problems which can be formulated as a linear programming problem.

Problem 1: (The Diet Problem) Let there be m nutrients $N_1, N_2, ..., N_m$ and n food products, $F_1, F_2, ..., F_n$, available in the market which can supply these nutrients. For healthy survival a human being requires say, b_i units of the i th nutrient, $i = 1, 2, \ldots, m$, respectively. Let a_{ij} be the amount of the i th nutrient (N_i) present in unit amount of the j th food product (F_j), and let $c_j, j = 1, 2, ..., n$ be the cost of unit amount of F_j. So now the problem is to decide on a diet of minimum cost consisting of the n food products (in various quantities) so that one gets the required amount of each of the nutrients.
Let x_j be the amount of the jth food product to be used in the diet (so $x_j \geq 0$ for $j = 1, 2, .., n$,) and the problem reduces to (under certain simplifying assumptions):
Min $\sum_{i=1}^n c_i x_i = \mathbf{c}^T\mathbf{x}$
subject to
$\sum_{j=1}^n a_{ij}x_j \geq b_i$, for $i = 1, 2, ..., m$,
$x_j \geq 0$ for all $j = 1, 2, \ldots, n$.
or as
$A\mathbf{x} \geq \mathbf{b}$, (or alternatively as $-A\mathbf{x} \leq -\mathbf{b}$) $\mathbf{x} \geq \mathbf{0}$,
where A is an $m \times n$ matrix (a matrix with m rows and n columns), the (i, j) th entry of A is given by a_{ij}, $\mathbf{b} = [b_1, b_2, \ldots, b_m]^T$ and $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$.

Problem 2: (The Transportation Problem) Let there be m supply stations, $S_1, S_2, .., S_m$ for a particular product (P) and n destination stations, $D_1, D_2, ..., D_n$ where the product is to be transported. Let c_{ij} be the cost of transportation of unit amount of the product (P) from S_i to D_j. Let s_i be the amount of the product available at S_i and let d_j be the corresponding demand at D_j. The problem is to find x_{ij}, $i = 1, 2, .., m$, $j = 1, 2, ..., n$, where x_{ij} is the amount of the product to be transported from S_i to D_j such that the cost of transportation is minimum.
The problem can be modeled as (under certain simplifying assumptions)
Min $\sum_{i,j} c_{ij}x_{ij}$
subject to
$\sum_{j=1}^n x_{ij} \leq s_i$, for $i = 1, 2, ..., m$,

$\sum_{i=1}^{m} x_{ij} \geq d_j$, for $j = 1, 2, ..., n$,
$x_{ij} \geq 0$ for all $i = 1, 2, .., m, j = 1, 2, .., n$.
Note that the constraints of the above LPP can again be written as:
Min $\mathbf{c}^T\mathbf{x}$
subject to $A\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$,
where A is a matrix with $(m+n)$ rows and $(m \times n)$ columns, $\mathbf{x}$ is a vector with $m \times n$ components and $\mathbf{b} = [s_1, \ldots, s_m, d_1, \ldots, d_n]^T$.
For example the 1st row of A(the row corresponding to the first supply constraint) is given by $[1, 1, \ldots, 1, 0 \ldots, \ldots, 0 \ldots, 0 \ldots]^T$ that is 1 in the first n positions and 0's elsewhere.
The second row of A(the row corresponding to the second supply constraint) is given by $[0, \ldots, 0, 1, 1, \ldots, 1, 0, \ldots, 0 \ldots, 0 \ldots]^T$ that is 1 in the $(n+1)$ th position to the $2n$ th position and 0's elsewhere.
The mth row of A (the row corresponding to the m th supply constraint) is given by $[0, \ldots, 0, 0, \ldots, 0 \ldots, 0 \ldots, 1, 1, \ldots, 1]^T$ that is 1 in the $(m-1)n+1$ th position to the mn th position and 0's elsewhere.

The $(m+1)$ th row of A (the row corresponding to the first destination constraint) is given by $[-1, 0 \ldots, 0, -1, 0, \ldots, 0, \ldots, -1, 0, \ldots, 0]$, that is -1 at the first position, -1 at the (n+1)th position, -1 at the (2n+1)th position, ..., -1 at the ((m-1)n +1) th position, etc and 0's elsewhere.
The $(m+n)$ th row of A (the row corresponding to the nth (last) destination constraint) is given by
$[0, \ldots, -1, 0, \ldots, -1, 0, \ldots, -1, 0, \ldots, -1, \ldots, 0, \ldots, -1]$, that is -1 at the nth position, -1 at the 2n th position, -1 at the 3n th position, ..., -1 at the $(m \times n)$ th position, etc and 0's elsewhere.
Solution by graphical method of LPP's in two variables:

Example 1: Given the linear programming problem
Max $5x + 2y$
subject to
$3x + 2y \leq 6$
$x + 2y \leq 4$
$x \geq 0, y \geq 0$.
Since the problem involves only two variables, we can try to solve this problem graphically. (Unfortunately I cannot draw graphs in latex.)
We got the feasible region (the set of all feasible solutions) to be a closed and bounded region, the boundary of this region is given by the straight lines obtained from the constraints as, $3x + 2y = 6$, $x + 2y = 4$, $x = 0$ and $y = 0$. Let us call the points of intersection of two or more of these straight lines as **corner points**. In order to get the optimal solution we first draw the line $5x + 2y = c$ for some c such that the line intersects the feasible region. Consider all points on this line which are inside the feasible region. If we keep on moving this line away from the origin, parallel to itself (that is by keeping the slope of the line unchanged) so that the line always intersects the feasible region, then we would get feasible solutions which would give higher (better) and higher (better) values of the objective function. So we get a collection of lines of the form $5x + 2y = c_1$, $5x + 2y = c_2$,...etc, where $c < c_1 < c_2 <$ .... We keep on moving the line till we reach a stage, that if we move it further away from the origin, then the line no longer intersects the feasible region. Then all feasible solutions on this line will give optimal solutions to the above.
Using this procedure we got the optimal solution to be $x = 2$ and $y = 0$, which happened to be a corner point of the feasible region. Later we will again convince ourselves (by using more rigor) that this is indeed the optimal solution. The optimal value in this problem is 10.

Example 2: Consider another problem,
Min $-x + 2y$
subject to
$x + 2y \geq 1$
$-x + y \leq 1,$
$x \geq 0, y \geq 0.$
Again using the same technique, we can start with the line $-x + 2y = 1$ and then move the line parallel to itself and towards the origin, to get a series of lines of the form $-x + 2y = c_1$, $-x + 2y = c_2$,...., where $1 > 0 > c_1 > c_2 > ...$, always intersecting the feasible region. We will see that no matter how far we move the line, it will always intersect the feasible region, or in other words given any line $-x+2y = c$ which is intersecting the feasible region we will always find another $d < c$, such that $-x + 2y = d$ will again intersect the feasible region. So our conclusion is that the LPP has unbounded solution (this is just a terminology) and that the linear programming problem does not have an optimal solution.

Example 3: Note that in the above problem keeping the feasible region same, if we just change the objective function to Min $2x + y$, then the changed problem has a unique optimal solution.

Example 4: Also in **Example 2** if we change objective function to Min $x + 2y$
then the changed problem has infinitely many optimal solutions, although the set of optimal solutions is bounded.

Example 5: Note that in **Example 2** keeping the feasible region same, if we just change the objective function to Min y, then the changed problem has infinitely many optimal solutions and the optimal solution set is unbounded.

Example 6: Max $-x + 2y$
subject to
$x + 2y \leq 1$
$-x + y \geq 1,$
$x \geq 0,\ y \geq 0.$
Clearly the feasible region of this problem is the empty set. So this problem is called **infeasible**, and since this problem does not have a feasible solution it obviously does not have an optimal solution.

So we observed from the previous examples that a linear programming problem may not have a feasible solution (Example 6). Even if it has a feasible solution it may not have an optimal solution (Example 2). If it has an optimal solution then it can have a unique solution (Example 3)or it can also have infinitely many solutions(Example 4 and Example 5). Few more questions which naturally arises are the following:

Question 1: Can there be exactly 2 or 5, or say exactly 100 solutions of a LPP?

And do the set of optimal solutions have some nice geometric structure, that is if we have two optimal solutions then what about points in between and on the line segment joining these two solutions?

Question 2: If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two optimal solutions of a LPP, then are $\mathbf{y}$'s of the form $\mathbf{y} = \lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$, $0 \leq \lambda \leq 1$, also optimal solutions?

Sets satisfying the above property are called **convex sets**,
or in other words is the set of optimal solutions/ feasible solutions of a LPP necessarily a convex set?

Definition: A nonempty set $S \subseteq R^n$ is said to be a **convex set** if for all $\mathbf{x}_1, \mathbf{x}_2 \in S$, $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in S$, for all $0 \leq \lambda \leq 1$.
Here $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, $0 \leq \lambda \leq 1$ is called a **convex combination** of $\mathbf{x}_1$ and $\mathbf{x}_2$.
If in the above expression, $0 < \lambda < 1$, then the **convex combination** is said to be a **strict convex combination** of $\mathbf{x}_1$ and $\mathbf{x}_2$.

Let us first try to answer **Question 2**, since if the answer to this question is a **YES** then that would imply that if a LPP has more than one solution then it should have infinitely many solutions, so the answer to **Question 1** would be a **NO**.

So let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two optimal solutions of the LPP given by
Max $\mathbf{c}^T\mathbf{x}$
subject to $A\mathbf{x} \leq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}$.
Since an optimal solution is also a feasible solution of the LPP, so
$A\mathbf{x}_1 \leq \mathbf{b}$, $\mathbf{x}_1 \geq \mathbf{0}$ and $A\mathbf{x}_2 \leq \mathbf{b}$, $\mathbf{x}_2 \geq \mathbf{0}$.
If $\mathbf{y}$ is a point in between and on the line segment joining $\mathbf{x}_1, \mathbf{x}_2$, then note that $\mathbf{y}$ is of the form $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $0 \leq \lambda \leq 1$.
Hence $\mathbf{y} \geq \mathbf{0}$ and

$$A\mathbf{y} = A(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) = \lambda A\mathbf{x}_1 + (1-\lambda)A\mathbf{x}_2 \leq \lambda\mathbf{b} + (1-\lambda)\mathbf{b} = \mathbf{b}. \tag{1}$$

So $\mathbf{y}$ is a feasible solution of the linear Programming Problem.
(Note that the above argument implies that the feasible region of a LPP is a convex set).
Since both $\mathbf{x}_1$ and $\mathbf{x}_2$ are optimal solutions to the same LPP, so
$\mathbf{c}^T\mathbf{x}_1 = \mathbf{c}^T\mathbf{x}_2 = q$, for some $q \in \mathbb{R}$.
Hence $\mathbf{c}^T\mathbf{y} = \mathbf{c}^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) = \lambda\mathbf{c}^T\mathbf{x}_1 + (1-\lambda)\mathbf{c}^T\mathbf{x}_2) = q$,
so $\mathbf{y}$ is also an optimal solution of the LPP.
Hence the set of all optimal solutions of a LPP is a convex set.

Hence a LPP either has no optimal solution, a unique optimal solution or infinitely many optimal solutions.
We have seen that, Fea(LPP) and the set of all optimal solutions of a LPP are both convex sets.
The observations from the above examples raises the following questions:

Question 3: If the feasible region of a LPP is a nonempty, bounded set then does the LPP always has an optimal solution?
The answer to this question is yes, due to a result by Weierstrass, called the **Extreme Value Theorem** given below:
Extreme Value Theorem: If S is a nonempty, closed and bounded subset of $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ is a continuous function, then f attains both its minimum and maximum value in S.

For a given $\mathbf{c} \in \mathbb{R}^n$, the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ is a continuous function from $\mathbb{R}^n$ to $\mathbb{R}$.
(To see the above, note that, $|f(\mathbf{x}) - f(\mathbf{y})| = |\mathbf{c}^T\mathbf{x} - \mathbf{c}^T\mathbf{y}| = |\mathbf{c}^T(\mathbf{x}-\mathbf{y})| \leq \|\mathbf{c}\|\|(\mathbf{x}-\mathbf{y})\|$ (by Cauchy-Schwarz Inequality).)
Since $S = Fea(LPP)$ is a closed set, if S is nonempty and bounded, then from Weierstrass's theorem, the continuous function $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ always attains both its maximum and minimum value in S.

Question 4: Whenever a LPP has an optimal solution does there always exist at least one corner point (points lying at the point of intersection of at least two distinct lines in case $n = 2$), at which the optimal value is attained?

Definition: Given a linear programming problem with a nonempty feasible region, $Fea(LPP) = S \subset \mathbb{R}^n$, an element $\mathbf{x} \in S$ is said to be an **corner point** of S, if $\mathbf{x}$ lies at the point of intersection of n linearly independent hyperplanes defining S.

An $\mathbf{x} \in Fea(LPP)$ which does not lie in any of the defining hyperplanes of $Fea(LPP)$ will be called as an **interior point** of $Fea(LPP)$ and an $\mathbf{x} \in Fea(LPP)$ which lies in atleast one defining hyperplane of the $Fea(LPP)$ as a **boundary point** of LPP.

Note that the way we have written our feasible region S, it has $(m+n)$ defining hyperplanes. Also, note that the **corner points** of the feasible region of the LPP cannot be written as a strict convex combination of two distinct points of the feasible region, or in other words those are all **extreme points** of the feasible region.

Definition: Given a nonempty convex set, $S \subset \mathbb{R}^n$, an element $\mathbf{x} \in S$ is said to be an **extreme point** of the set S if $\mathbf{x}$ cannot be written as a strict convex combination of two distinct points of S. That is if
$\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, for some $0 < \lambda < 1$, and $\mathbf{x}_1, \mathbf{x}_2 \in S$, then $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$.

Theorem: If $S = Fea(LPP)$ is a nonempty set,
where $Fea(LPP) = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}_i^T\mathbf{x} \le b_i$ for all $i = 1, \ldots, m, \quad -\mathbf{e}_j^T\mathbf{x} \le 0$ for all $j = 1, \ldots, n\}$
then $\mathbf{x} \in S$ is a corner point of S if and only if it is an extreme point of S.

Proof: Let $\mathbf{x}_0 \in S = Fea(LPP)$ be a corner point of S, that is $\mathbf{x}_0$ lies at the point of intersection of n linearly independent (LI) hyperplanes out of the $(m+n)$ hyperplanes defining S.
Let us denote the defining hyperplanes of $Fea(LPP)$ as
$\tilde{\mathbf{a}}_i^T\mathbf{x} = \tilde{b}_i$, for $i = 1, 2, .., m+n$, where
$\tilde{\mathbf{a}}_i = \mathbf{a}_i$ for $i = 1, 2, \ldots, m$
$= -\mathbf{e}_{i-m}$ for $i = m+1, m+2, \ldots, m+n$.
Hence $\tilde{\mathbf{a}}_{m+1} = -\mathbf{e}_1, \ldots, \tilde{\mathbf{a}}_{m+n} = -\mathbf{e}_n$, etc, and
$\tilde{b}_i = b_i$ for $i = 1, 2, \ldots, m$
$= 0$ for $i = m+1, m+2, \ldots, m+n$.
Hence the constraints $A\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$ can be together written as:
$\tilde{A}\mathbf{x} \le \tilde{\mathbf{b}}$, where $\tilde{A}$ is the matrix whose i th row is given by $\tilde{\mathbf{a}_i}^T$, or
$\tilde{A} = \begin{bmatrix} A_{m\times n} \\ -I_{n\times n} \end{bmatrix}$
and $\tilde{\mathbf{b}} = [\tilde{b_1}, \ldots, \tilde{b}_{m+n}]^T$, where $\tilde{b_i}$'s are as defined before.
Hence $\mathbf{x}_0$ satisfies the following conditions,
$\tilde{\mathbf{a}_j}^T\mathbf{x}_0 = \tilde{b_j}$ for $j = 1, \ldots, n$ for some n LI vectors $\{\tilde{\mathbf{a}_1}, \ldots, \tilde{\mathbf{a}_n}\}$ (we have taken WLOG the first n), where $\tilde{\mathbf{a}_j}$ could be either one of the $\mathbf{a}_i$'s or one of the $-\mathbf{e}_i$'s and the $\tilde{b_j}$'s are either equal to one of the b_i's or is equal to 0.
Then to show that $\mathbf{x}_0$ cannot be written as a strict convex combination of two distinct points of S. Suppose if $\mathbf{x}_0 = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ for some $\mathbf{x}_1, \mathbf{x}_2 \in S$ and some λ, $0 < \lambda < 1$, then to show that $\mathbf{x}_0 = \mathbf{x}_1 = \mathbf{x}_2$.
Since $\mathbf{x}_1, \mathbf{x}_2 \in S$, then note that $\tilde{\mathbf{a}_j}^T\mathbf{x}_1 \le \tilde{b_j}$ and $\tilde{\mathbf{a}_j}^T\mathbf{x}_2 \le \tilde{b_j}$ for all $j = 1, 2, \ldots, n$, and
$\tilde{\mathbf{a}_j}^T\mathbf{x}_0 = \lambda\tilde{\mathbf{a}_j}^T\mathbf{x}_1 + (1-\lambda)\tilde{\mathbf{a}_j}^T\mathbf{x}_2$. (**)
Hence if for any $j = 1, 2, \ldots, n$,

$\tilde{\mathbf{a}_j}^T\mathbf{x}_1 < \tilde{b_j}$ or $\tilde{\mathbf{a}_j}^T\mathbf{x}_2 < \tilde{b_j}$, then since $0 < \lambda < 1$, $\tilde{\mathbf{a}_j}^T\mathbf{x}_0$ will also be strictly less that $\tilde{b_j}$ which is a contradiction, hence each of $\mathbf{x}_1$ and $\mathbf{x}_2$ must also lie in each of these n LI hyperplanes.
Hence each of $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ must satisfy the following system of equations:
$\tilde{\mathbf{a}_j}^T\mathbf{x} = \tilde{b_j}$, $j = 1, 2, .., n$.
OR
$B\mathbf{x} = \mathbf{p}$, where B is an $n \times n$ matrix with n linearly independent rows given by $\tilde{\mathbf{a}_j}^T$, $j = 1, 2, ..., n$ and $\mathbf{p}$ is the vector with n components given by $\tilde{b_j}$, $j = 1, 2, \ldots, n$.
Since B is a nonsingular matrix the system $B\mathbf{x} = \mathbf{p}$ has a unique solution given by $\mathbf{x} = B^{-1}\mathbf{p}$.
Hence $\mathbf{x}_0 = \mathbf{x}_1 = \mathbf{x}_2$, or in other words $\mathbf{x}_0$ is an extreme point of S.

To show the converse, that is if $\mathbf{x}$ is an extreme point of S (that is it cannot be written as a strict convex combination of 2 distinct points of S) then it should lie at the point of intersection of n LI hyperplanes defining the feasible region, we show that if $\mathbf{x}_0$ is an element of S which lies in **exactly** k, $0 \leq k < n$, LI hyperplanes defining the feasible region then $\mathbf{x}_0$ is not an extreme point of S.

If $\mathbf{x}_0 \in S$ is one such point then let us assume (without loss of generality) that
$\tilde{\mathbf{a}_j}^T\mathbf{x}_0 = \tilde{b_j}$ for $j = 1, 2, .., k$,
where $\tilde{\mathbf{a}_j}^T = \mathbf{a}_j^T$ for $j = 1, \ldots, m$ and $\tilde{\mathbf{a}_j}^T = -\mathbf{e}_{j-m}^T$ for $j = m+1, \ldots, m+n$.
Let as before, $\tilde{A}$ be the $(m+n) \times n$ matrix with the ith row equal to $\tilde{\mathbf{a}}_i^T$, for $i = 1, \ldots, m$, and equal to $-\mathbf{e}_{i-m}^T$, for $i = m+1, \ldots, n$.
Let B be the $k \times n$ matrix (which is a submatrix of $\tilde{A}$) whose rows are given by $\tilde{\mathbf{a}_i}^T, i = 1, .., k$.
Then $\mathbf{x}_0$ satisfies the equation $B_{k\times n}\mathbf{x} = \mathbf{p}$ where $\mathbf{p}$ is a column vector with k components obtained from $\tilde{\mathbf{b}}$ by removing the last $(m+n-k)$ components.
Since $rank(B) = k < n$, the columns of B, denoted by say, $\mathbf{b}_1', .., \mathbf{b}_n'$ are linearly dependent.
Hence there exists real numbers $d_1, ..., d_n$ not all zero's such that
$\sum_{i=1}^n d_i\mathbf{b}_i' = B\mathbf{d} = \mathbf{0}$,
where $\mathbf{d} \neq \mathbf{0}$ is a column vector with n components, $d_1, d_2, .., d_n$ and $\mathbf{0}$ is the column vector with k components, all of which are equal to zero.

$$\text{Since } \tilde{A}(\mathbf{x}_0 + \epsilon\mathbf{d}) = \begin{bmatrix} B \\ \tilde{\mathbf{a}}_{k+1}^T \\ \vdots \\ \tilde{\mathbf{a}}_{m+n}^T \end{bmatrix} (\mathbf{x}_0 + \epsilon\mathbf{d}) = \begin{bmatrix} B\mathbf{x} + \epsilon B\mathbf{d} \\ \tilde{\mathbf{a}}_{k+1}^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_{k+1}^T\mathbf{d} \\ \vdots \\ \tilde{\mathbf{a}}_{m+n}^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_{m+n}^T\mathbf{d} \end{bmatrix} = \begin{bmatrix} B\mathbf{x} \\ \tilde{\mathbf{a}}_{k+1}^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_{k+1}^T\mathbf{d} \\ \vdots \\ \tilde{\mathbf{a}}_{m+n}^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_{m+n}^T\mathbf{d} \end{bmatrix}.$$

Note that for any $i = k+1, \ldots, m+n$, if the set $\{\tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_k\}$ is LI then $\tilde{\mathbf{a}_i}^T\mathbf{x}_0 < \tilde{b_i}$, since by assumption $\mathbf{x}_0 \in S$ (so $\tilde{\mathbf{a}_i}^T\mathbf{x}_0 \leq \tilde{b_i}$), and $\mathbf{x}_0$ lies in exactly k LI hyperplanes defining S.

Hence we can choose an $\epsilon > 0$ sufficiently small, such that $\tilde{\mathbf{a}}_i^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_i^T\mathbf{d} < \tilde{b}_i$ for all such i's (since there are only finitely many($\leq m+n-k$) such i's).

If however for some $i = k+1, \ldots, m+n$, the set $\{\tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_k\}$ is linearly dependent, then since the set $\{\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_k\}$ is LI, it implies that $\tilde{\mathbf{a}}_i$ can be expressed as a linear combination of the vectors, $\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_k$,
that is there exists real numbers $u_1, u_2, .., u_k$, such that $\tilde{\mathbf{a}}_i = u_1\tilde{\mathbf{a}}_1 + u_2\tilde{\mathbf{a}}_2 \ldots + u_k\tilde{\mathbf{a}}_k$.
Hence $\tilde{\mathbf{a}}_i^T\mathbf{d} = u_1\tilde{\mathbf{a}}_1^T\mathbf{d} + u_2\tilde{\mathbf{a}}_2^T\mathbf{d} \ldots + u_k\tilde{\mathbf{a}}_k^T\mathbf{d} = 0$.
Hence $\tilde{\mathbf{a}}_i^T\mathbf{x}_0 + \epsilon\tilde{\mathbf{a}}_i^T\mathbf{d} = \tilde{\mathbf{a}}_i^T\mathbf{x}_0 \leq \tilde{b}_i$.
Hence $(\mathbf{x}_0 + \epsilon\mathbf{d}) \in S$. By choosing $\epsilon > 0$ sufficiently small, we can similarly make $(\mathbf{x}_0 - \epsilon\mathbf{d}) \in S$.
Hence for $\epsilon > 0$ sufficiently small, we have both, $(\mathbf{x}_0 + \epsilon\mathbf{d}) \in S$ and $(\mathbf{x}_0 - \epsilon\mathbf{d}) \in S$.
Since $\mathbf{d} \neq \mathbf{0}$, note that $(\mathbf{x}_0 + \epsilon\mathbf{d})$ and $(\mathbf{x}_0 - \epsilon\mathbf{d})$ are two distinct points of S.

But $\mathbf{x}_0 = \frac{1}{2}(\mathbf{x}_0 + \epsilon\mathbf{d}) + \frac{1}{2}(\mathbf{x}_0 - \epsilon\mathbf{d})$, implies that $\mathbf{x}_0$ is not an extreme point of $S = Fea(LPP)$.

Remark: From the above equivalent definition of extreme points of the feasible region of a LPP it is clear that the total number of extreme points of the feasible region is $\leq (m+n)_{C_n}$.

Exercise: Think of a LPP such that the number of extreme points of the $Fea(LPP)$, is equal to that given by the upper bound.

Exercise: If possible give an example of a LPP with $(m+n)$ constraints (including the non-negativity constraints) such that the number of extreme points of the $Fea(LPP)$ is strictly greater than $(m+n)$ and is strictly less than $\leq (m+n)_{C_n}$.

Exercise: If possible give an example of a LPP with $(m+n)$ constraints (including the non-negativity constraints) such that the number of extreme points of the $Fea(LPP)$ is strictly less than $(m+n)$.

Question: Does the feasible region of a LPP (where the feasible region is of the form given before) always have an extreme point?
Answer: **Yes**, we will see this later.

Exercise: Think of a convex set defined by only one hyperplane. Will it have any extreme point?