Recall the formulation of the problem:

\[
\begin{align*}
\text{max } w \\
&c_1 + p_1 - e_1 = 150 \\
&c_2 + p_2 + 1.003e_1 - 1.01c_1 - e_2 = 100 \\
&c_3 + p_3 + 1.003e_2 - 1.01c_2 - e_3 = -200 \\
&c_4 - 1.02p_1 - 1.01c_3 + 1.003e_3 - e_4 = 200 \\
&c_5 - 1.02p_2 - 1.01c_4 + 1.003e_4 - e_5 = -50 \\
&- 1.02p_3 - 1.01c_5 + 1.003e_5 - w = -300 \\
0 \leq c_i \leq 100, i \leq i \leq 5 \\
p_i > 0, 1 \leq i \leq 3 \\
e_i \geq 0, 1 \leq i \leq 5
\end{align*}
\]

This is a *Linear Programming (LP)* problem:

- the variables are *continuous*;
- the constraints are *linear functions* of the variables;
- the objective function is *linear* in the variable.
Linear Programming

Linear Programming is arguably the best known and most frequently solved class of optimization problems.

<table>
<thead>
<tr>
<th>Definition (Linear Program)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A generic LP problem has the following form:</td>
</tr>
</tbody>
</table>
| \[
\begin{align*}
\text{min } & \quad c^T x \\
\text{subject to } & \quad a^T x = b \text{ for } (a, b) \in E \\
& \quad a^T x \geq b \text{ for } (a, b) \in I
\end{align*}
\] |

where

- \( x \in \mathbb{R}^n \) is the vector of decision variables;
- \( c \in \mathbb{R}^n \), the costs vector, defines the objective function;
- \( E \) and \( I \) are sets of pairs \((a, b)\) where \( a \in \mathbb{R}^n \) and \( b \in \mathbb{R} \).

Any assignment of values to \( x \) is called a solution. A feasible solution satisfies the constraints. The optimal solution is feasible and minimize the objective function.
Graphical solution to a linear optimization problem

Let’s solve:

\[ \text{max } 500x_1 + 450x_2 \]
\[ x_1 + \frac{5}{6}x_2 \leq 10 \]
\[ x_1 + 2x_2 \leq 15 \]
\[ x_1 \leq 8 \]
\[ x_1, x_2 \geq 0 \]
More often, we will express LP in the *standard form*

<table>
<thead>
<tr>
<th>Standard Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \min c^T x ]</td>
</tr>
<tr>
<td>[ Ax = b ]</td>
</tr>
<tr>
<td>[ x \geq 0 ]</td>
</tr>
</tbody>
</table>

where \( A \) is a \( n \times n \) matrix, and \( b \in \mathbb{R}^n \).

The standard form is *not restrictive*:

- we can rewrite inequality constraints as equalities by introducing *slack variables*;
- maximization problems can be written as minimization problems by multiplying the objective function by \(-1\);
- variables that are unrestricted in sign can be expressed as the difference of two new non-negative variables.
**Standard form and slack variables (cont.)**

How do we express the following problem in standard form?

<table>
<thead>
<tr>
<th>Non-standard form</th>
</tr>
</thead>
<tbody>
<tr>
<td>max $x_1 + x_2$</td>
</tr>
<tr>
<td>$2x_1 + x_2 \leq 12$</td>
</tr>
<tr>
<td>$x_1 + 2x_2 \leq 9$</td>
</tr>
<tr>
<td>$x_1 \geq 0$, $x_2 \geq 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard form</th>
</tr>
</thead>
<tbody>
<tr>
<td>min $-x_1 - x_2$</td>
</tr>
<tr>
<td>$2x_1 + x_2 + z_1 = 12$</td>
</tr>
<tr>
<td>$x_1 + 2x_2 + z_2 = 9$</td>
</tr>
<tr>
<td>$x_1 \geq 0$, $x_2 \geq 0$, $z_1 \geq 0$, $z_2 \geq 0$</td>
</tr>
</tbody>
</table>

The $z_i$ are **slack variables**: they appear in the constraints, but not in the objective function.
Lower bounds to the optimal objective value

We would like to compute an optimal solution to a LP…

But let’s start with an easier question:

**Question**

How can we find a *lower bound* to the objective function value for a feasible solution?

**Intuition for the answer**

The constraints provide some bounds on the value of the objective function.
Consider

\[ \min - x_1 - x_2 \]
\[ 2x_1 + x_2 \leq 12 \]
\[ x_1 + 2x_2 \leq 9 \]
\[ x_1 \geq 0, x_2 \geq 0 \]

**Question**

*Can we use the first constraint to give a lower bound to \(-x_1 - x_2\) for any feasible solution?*

**Answer (Yes!)*

*Any feasible solution must be such that*

\[ -x_1 - x_2 \geq -2x_1 - x_2 \geq -12 \, . \]

We leverage the fact that the coefficients in the constraint (multiplied by -1) are lower bounds to the coefficients in the objective function.
The second constraint gives us a better bound:

\[-x_1 - x_2 \geq -x_1 - 2x_2 \geq -9\]

**Question**

*Can we do even better?*

Yes, with a linear combination of the constraints:

\[-x_1 - x_2 \geq -\frac{1}{3}(2x_1 + x_2) - \frac{1}{3}(x_1 + 2x_2) \geq -\frac{1}{3}(12 + 9) \geq -7\]

No feasible solution can have an objective value smaller than -7.
Lower bounds (cont.)

**General strategy**

Find a linear combination of the constraints such that

- the resulting coefficient for each variable is no larger than the corresponding coefficient in the objective function; and
- the resulting lower bound to the optimal objective value is *maximized*.

In other words, we want to find the coefficients of the linear combinations of constraints that satisfy the new constraints on the coefficients of the variables, and maximize the lower bound.

This is a new optimization problem: the *dual problem*. The original problem is known as the *primal problem*. 
**Dual problem**

**Question**

What is the dual problem in our example?

**Answer**

\[
\begin{align*}
\text{max } & \quad 12y_1 + 9y_2 \\
2y_1 + y_2 & \leq -1 \\
y_1 + 2y_2 & \leq -1 \\
y_1, y_2 & \leq 0
\end{align*}
\]

In general:

<table>
<thead>
<tr>
<th>Primal problem</th>
<th>Dual problem</th>
<th>Dual problem (min. form)</th>
</tr>
</thead>
</table>
| \[ \begin{align*}
\text{min } & \quad c^T x \\
Ax & \leq b \\
x & \geq 0
\end{align*} \] | \[ \begin{align*}
\text{max } & \quad b^T y \\
A^T y & \leq c \\
y & \leq 0
\end{align*} \] | \[ \begin{align*}
\text{min } & \quad -b^T y \\
A^T y & \leq c \\
y & \leq 0
\end{align*} \] |
Weak duality

**Theorem (Weak duality)**

Let $x$ be any feasible solution to the primal LP and $y$ be any feasible solution to the dual LP. Then

$$c^T x \geq b^T y .$$

**Proof.**

We have

$$c^T x - b^T y = c^T x - y^T b = c^T x - y^T Ax = (c - A^T y)^T x > 0$$

where the last step follows from $x \geq 0$ and $(c - A^T y) \geq 0$ (why?).

**Definition (Duality gap)**

The quantity $c^T x - b^T y$ is known as the *duality gap*. 
Corollary

If:
• $x$ is feasible for the primal LP;
• $y$ is feasible for the dual LP; and
• $c^T x = b^T y$;

then $x$ must be optimal for the primal LP and $y$ must be optimal for the dual LP.

The above gives a sufficient condition for optimality of a primal-dual pair of feasible solutions. This condition is also necessary.

Theorem (Strong duality)

If the primal (resp. dual) problem has an optimal solution $x$ (resp. $y$), then the dual (resp. primal) has an optimal solution $y$ (resp. $x$) such that $c^T x = b^T y$. 
Complementary slackness

How can we obtain an optimal solution to the dual problem from an optimal solution to the primal (and vice versa)?

**Theorem (Complementary slackness)**

Let $x$ be an optimal solution for the primal LP and $y$ be an optimal solution to the dual LP. Then

$$y^T(Ax - b) = 0 \text{ and } (c^T - y^T A)x = 0$$

**Proof.**

Exercise.
Due to the great relevance of LP in many fields, a number of algorithms have been developed to solve LP problems.

**Simplex Algorithm:** Due to G. Dantzig in 1947, it was a breakthrough and it is considered “one of the top 10 algorithms of the twentieth century.” It is an *exponential time* algorithm, but extremely efficient in practice also for large problems.

**Ellipsoid Method:** by Yudin and Nemirovski in 1976, it was the first *polynomial time* algorithm for LP. In practice, the performance is so bad that the algorithm is only of theoretical interest, and even so, only for historical purposes.

**Interior-Point Method:** by Karmarkar in 1984, it is the first polynomial time algorithm for LP that can solve some real-world LPs faster than the simplex.

We now present and analyze the Simplex Algorithm, and will discuss the Interior-Point Method later, in the context of Quadratic Programming.
Roadmap

1. Look at the geometry of the LP feasible region
2. Prove the existence of an optimal solution that satisfy a specific geometrical condition
3. Study what solutions that satisfy the condition look like
4. Discuss how to move between solutions that satisfy the condition
5. Use these ingredients to develop an algorithm
6. Analyze correctness and running time complexity
Convex polyhedra

Definition (Convex polyhedron)

A *convex polyhedron* $\mathcal{P}$ is the solution set of a system of $m$ linear inequalities:

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

$A$ is $m \times n$, $b$ is $m \times 1$.

Fact

*The feasible region of an LP is a convex polyhedron.*

Definition (Polyhedron in standard form)

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

$A$ is $m \times n$, $b$ is $m \times 1$. 
Extreme points and their optimality

**Definition (Extreme point)**

$x \in \mathcal{P}$ is an *extreme point* of $\mathcal{P}$ iff there exist no distinct $y, z \in \mathcal{P}$, $\lambda \in (0, 1)$ s.t. $x = \lambda y + (1 - \lambda)z$.

**Theorem (Optimality of extreme points)**

Let $\mathcal{P} \subseteq \mathbb{R}^n$ be a polyhedron and consider the problem $\min_{x \in \mathcal{P}} c^T x$ for a given $c \in \mathbb{R}^n$. If $\mathcal{P}$ has at least one extreme point and there exists an optimal solution, then there exists an optimal solution that is an extreme point.

**Proof.**

Coming up.
Proof of optimality of extreme points

$v$: optimal objective value
$Q$: set of optimal solutions,

\[ Q = \{ x \in \mathbb{R}^n : Ax = b, x \geq 0, c^T x = v \} \subseteq \mathcal{P} \]

Fact

$Q$ is a convex polyhedron.

Fact

Since $Q \subseteq \mathcal{P}$ and $\mathcal{P}$ has an extreme point, $Q$ must have an extreme point.

Let $x^*$ be an extreme point of $Q$.
We now show that $x^*$ is an extreme point in $\mathcal{P}$. 
Proof of optimality of extreme points (cont.)

Let’s show that $x^*$ (extreme point of $Q$) is an extreme point of $P$.

- Assume $x^*$ is not an extreme point of $P$, i.e., 
  $\exists y, z \in P, y \neq x^*, z \neq x^*, \lambda \in (0, 1)$ s.t. $\lambda y + (1 - \lambda)z = x^*$.

- How can we write $c^T x^*$? $c^T x^* = \lambda c^T y + (1 - \lambda)c^T z$.

- Let’s compare $c^T x^*$ and $c^T y$ and $c^T x^*$ and $c^T z$. It must be $c^T y \geq c^T x^*$ or $c^T z \geq c^T x^*$. Suppose $c^T y \geq c^T x^*$.

- It must also be $c^T y = c^T x^*$. Why? Because $x^*$ is an optimal solution: $c^T y = c^T x^* = v$.

- But then what about $c^T z$? $c^T z = v$.

- What does $c^T y = c^T z = v$ imply? They must belong to $Q$.

- Does this contradict the hypothesis? We found $y$ and $z$ in $Q$ s.t. $x^* = \lambda y + (1 - \lambda)z$, so $x^*$ is not an extreme point of $Q$.

- We reach a contradiction, thus $x^*$ is an extreme point in $P$.

QED
Theorem

Every convex polyhedron in standard form has an extreme point.

Corollary

Every LP with a non-empty feasible region $\mathcal{P}$
- either has optimal value $-\infty$;
- or there exists an optimal solution at one of the extreme points of $\mathcal{P}$. 
Algorithmic idea and questions

Algorithmic idea

Find an initial extreme point \( x \).
While \( x \) is not optimal:
    \( x \leftarrow \) a different extreme point.
Output \( x \)

There may be many extreme points, so not the best algorithm?

We must answer the following questions:

- What is the algebraic form of extreme points?
- How do we move among extreme points?
- How do we avoid revisiting the same extreme point / ensure that we find all of them if necessary?
- How do we know that an extreme point is optimal?
- How do we find the first extreme point?
Basic solutions

Assume from now on that the \( m \) rows of \( A \) are \textit{linearly independent}. Consider a set of \( m \) columns of \( A = [a_1 \cdots a_n] \) that are linearly independent. Let \( r_1, \ldots, r_m \) be their indices. Let \( B = [a_{r_1} \cdots a_{r_m}] \). Dimensions of \( B \) are \( m \times m \).

Permute the columns of \( A \) to obtain \([B \quad N]\). Consider the same permutation for the rows of \( x \) to obtain \([x^T_B \quad x^T_N]^T\).

\[
Ax = b \iff [B \quad N] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = b
\]

\[
\iff Bx_B + Nx_N = b
\]

\[
\iff x_B = B^{-1}b - B^{-1}Nx_N
\]

Why is \( B \) invertible? It's a square matrix with full rank.

What’s an easy solution to the last equation? \( x_N = 0 \) and \( x_B = B^{-1}b \).

A solution obtained this way is known as a \textit{basic solution}.
# Basic solutions

## To obtain a basic solution:

1. Choose $m$ linearly independent columns of $A$ to obtain $B$. The columns in $B$ are known as the basic columns.
2. Split the variables $x$ into basic variables $x_B$ and non-basic variables $x_N$.
3. Set the non-basic variables $x_N$ to zero.
4. Set $x_B = B^{-1}b$.

Is a basic solution $x$ always feasible? No. It must also be $x \geq 0$! (i.e., $x_B \geq 0$).

If $x$ is basic and feasible, is a Basic Feasible Solution (BFS).
Basic feasible solutions and extreme points

Theorem

Each BFS corresponds to one and only one extreme point of $\mathcal{P}$.

Proof: (only that a BFS $x$ must be an extreme point of $\mathcal{P}$).

- Assume that $x$ is not an extreme point of $\mathcal{P}$. What does it mean? $\exists$ distinct $y, z \in \mathcal{P}$ and $\lambda \in (0, 1)$ s.t. $x = \lambda y + (1 - \lambda)z$.

- What about $x_B$ and $y_B$, $z_B$, and $x_N$ and $y_N$, $z_N$?

  
  
  $0 = x_N = \lambda y_N + (1 - \lambda)z_N$

  It must be... $y_N = z_N = 0$, because $y, z \in \mathcal{P}$, so $y, z \geq 0$.

- What about $x_B$ and $y_B$ and $z_B$?

  
  
  $x_B = B^{-1}b$ but also must be $By_B = b$ and $Bz_B = bi$

  so it must be $x_B = y_B = z_B$.

- What’s the contradiction? We did not find distinct $y$ and $z$ to express $x$. 

  $\square$
Corollary

\textit{If } P \textit{ is not empty,}

\begin{itemize}
  \item \textit{either the optimal value is } -\infty
  \item \textit{or there is a BFS … that is optimal.}
\end{itemize}

BFSs are “algebraic” representations of the “geometric” extreme points.

Algorithmic idea

Find an initial BFS \( x \).
While \( x \) is not optimal:
\[ x \leftarrow \text{a different BFS.} \]
Output \( x \)
Adjacent BFS and directions

Definition (Adjacent BFSs)

Two BFSs are adjacent iff their basic matrices differ in only one basic column.

We want to move between adjacent BFSs because they look “similar”. How do we move among adjacent bfs?

Definition (Feasible direction)

Let $x \in \mathcal{P}$. A vector $d \in \mathbb{R}^n$ is a feasible direction at $x$ if there exists $\theta > 0$ such that $x + \theta d \in \mathcal{P}$.

Definition (Improving direction)

Consider the LP $\min_{x \in \mathcal{P}} c^T x$. A vector $d \in \mathbb{R}^n$ is a improving direction if $c^Td < 0$.

Starting from a BFS, we want to find a feasible improving direction towards an adjacent BFS.
Basic (feasible) directions

**Fact**

*Any direction that is feasible at a BFS $x$ must have a strictly positive value in at least one of the components corresponding to non-basic variables of $x$ (why?).*

We move in a *basic direction* $d = [d_B \quad d_N]$ that has a strictly positive value in *exactly* one of the components corresponding to non-basic variables, e.g., $x_j$.

We fix $d_N$ so that

$$d_j = 1$$

$$d_i = 0 \text{ for every non-basic index } i \neq j$$

And

$$d_B = -B^{-1}a_j .$$

The choice for $d_B$ comes from the requirement $Ad = 0$ for any feasible direction $d$ (why?).
How do we know if the chosen basic feasible direction is improving?

**Definition (Reduced cost)**

Let \( x \) be a basic solution with basis matrix \( B \), and let \( c_B \) be the vector of the costs of the basic variables. For each \( 1 \leq i \leq n \), the *reduced cost* \( \hat{c}_i \) of variable \( x_i \) is

\[
\hat{c}_i = c_i - c_B^T B^{-1} a_i .
\]

**Fact**

The basic direction associated with \( x_j \) is improving iff \( \hat{c}_j < 0 \).

Why? (Hint: \( \hat{c}_B = 0 \) (why?))
**Definition (Degenerate bfs)**

Let $x$ be a bfs. $x$ is said to be **degenerate** iff $|\{ j : x_j > 0 \}| < m$.

**Theorem**

Let $x$ be a BFS and let $\hat{c}$ be the corresponding vector of reduced costs.

- if $\hat{c} \geq 0$, then $x$ is optimal.
- if $x$ is optimal and nondegenerate, then $\hat{c} \geq 0$. 
### Step size

- Let \( d \) be a basic, feasible, improving direction from the current BFS \( x \), and let \( B \) be the basis matrix for \( x \).
- We want to find move along \( d \) to find a BFS \( x' \) adjacent to \( x \) that can be expressed as \( x' = x + \theta d \) for some value \( \theta \in \mathbb{R} \).

#### How to compute the step size \( \theta \)?

We know that at \( x' \), one of the basic variables in \( x \) will be 0. We want to find the maximum \( \theta \) such that

\[
x_B + \theta d_B \geq 0.
\]

Let \( u = -d_B = B^{-1}a_j \). The value of a basic variable \( x_{r_i} \) changes, along \( d \), at a “rate” of \( x_{r_i}/u_{r_i} \). Hence we want

\[
\theta = \min_{r_i : u_{r_i} > 0} \frac{x_{r_i}}{u_i}.
\]

The basic variable \( x_{r_i} \) that attains the minimum is the one that leaves the BFS (it is a non-basic variable at \( x' \)).
Bland’s rule

- What non-basic variable should enter the BFS?
- If more than one basic variable could leave (i.e., more than one basic variable attains the minimum that gives $\theta$), which one should leave?
- Why does all this matter?
Without an accurate choice of the entering/exiting variables, the algorithm may cycle forever at the same degenerate BFS.

Bland’s anticycling rule (eliminates the risk of cycling)

- Among all non-basic variables that can enter the basis, choose the one with the minimum index.
- Among all basic variables that can exit the basis, choose the one with the minimum index.
The Simplex algorithm

1. Start from a bfs $x$, with basic variables $x_{r1}, \ldots, x_{rm}$ and basis matrix $B$.

2. Compute the reduced costs $\hat{c}_j = c_j - c_B^T B^{-1} a_j$ for $1 \leq j \leq n$. If $\hat{c} \geq 0$, then $x$ is optimal.

3. Use Bland’s rule to select the entering variable $x_j$ and obtain $u = -d_B = B^{-1} a_j$ by solving the system $Bu = a_j$. If $u \leq 0$, then the LP is unbounded.

4. Compute the step size $\theta = \min_{r_i} : u_{r_i} > 0 \frac{x_{r_i}}{u_{r_i}}$.

5. Determine the new solution and the leaving variable $x_{r_i}$, using Bland’s rule.

6. Replace $x_{r_i}$ with $x_j$ in the list of basic variables.

7. Go to step 1.
### Fact

There are at most \( \binom{n}{m} \) basic (feasible) solutions.

**Why?**

There are artificial LPs where the simplex visits all BFS.

Runtime is not polynomial.

With distribution assumptions on the input, the simplex performs \( O(n^4) \) iterations, each taking time \( O(nm) \).

For real-world instances, the average is \( \leq 3(n - m) \) steps.
Finding the first BFS

The simplex algorithm needs a BFS to start. How do we find it?

How to find a feasible solution s.t. $x_i = 0$?

Set the objective function to be just $x_i$, and find the optimal solution to this problem.

If the optimal value is 0 then we found a feasible solution with $x_i = 0$, otherwise there is no such feasible solution.

If we want multiple variables set to 0, set the objective function to be to the sum of the variables that should be 0.
Two-phases simplex algorithm

1. introduce *artificial variables* (not the same as slack variables);
2. easily find initial BFS involving the artificial variables;
3. Set objective function to be the sum of the artificial variables;
4. Run the simplex algorithm on the modified LP;
5. If the optimal value is 0, we obtain a BFS for the original LP. Otherwise, the original LP has no feasible solution;
6. If we found an initial BFS for the original problem, run the simplex algorithm on it.

This procedure is known as the *Two-phases simplex algorithm*, with steps 1-5 being *Phase 1* and step 6 being *Phase 2*. 
Artificial variables and initial BFS

Introducing the artificial variables

- Start with a LP in *standard* form.
- Add to the LHS of each constraint an *artificial variable* with sign equal to the sign of the RHS.

Initial BFS (for the artificial LP)

A BFS with the artificial variables being the basic variables, with values equal to the RHS of their constraints (non-artificial variables are (non-basic) and set to 0).

We can now solve the artificial LP and find a BFS for the original LP, if it exists.
Dual Simplex

Through strong duality, we can compute the dual optimal solution after having solved the primal optimal solution with the simplex algorithm.

Observation

Solve the primal by solving the dual with the simplex, and then obtain a primal optimal solution from the dual optimal solution.

The advantage is that if $c \geq 0$, then $y = 0$ is dual feasible, so we can avoid Phase 1, and go directly to finding the optimal solution.