Branch & Bound and Knapsack Lab

Objectives

- Preform the branch and bound algorithm
- Apply branch and bound to the knapsack problem
- Understand the geometry of the branch and bound algorithm

Brief description: In this lab, we will try solving an example of a knapsack problem with the branch-and-bound algorithm. We will also see how adding a cutting plane helps in reducing the computation time and effort of the algorithm. Lastly, we will explore the geometry of the branch and bound algorithm.

```
In [1]: # imports -- don't forget to run this cell
import pandas as pd
import gilp
from gilp.visualize import feasible_integer_pts
from ortools.linear_solver import pywraplp as OR
```

Part 1: Branch and Bound Algorithm

Recall that the branch and bound algorithm (in addition to the simplex method) allows us to solve integer programs. Before applying the branch and bound algorithm to the knapsack problem, we will begin by reviewing some core ideas. Furthermore, we will identify a helpful property that will make branch and bound terminate quicker later in the lab!

Q1: What are the different ways a node can be fathomed during the branch and bound algorithm? Describe each.

A: (1) Integer solution: If one of the new nodes has integer solution, its bound is compared to the bounds of other such nodes. If it does not have the best value - it is fathomed. If it has the best value it is fathomed and it is our current best solution. (2) Bound value. If any of the new nodes has a bound smaller than currently the best bound - fathom the node. (3) Infeasibility. If LP at any of the new nodes has no solution (not feasible) - fathom the node.

If the optimal solution to the linear programming relaxation is not an integer solution, but its optimal value is greater than or equal to the values of the best integer solution found already we will not find a better integer solution so we must fathom the node.

If we know that the solution to an LP is c, then the optimal solution must be at least the next integer rounded up from c. If we already know an integer solution of the next integer rounded up from c then we can fathom the search at this node.

Q2: Suppose you have a maximization integer program and you solve its linear program relaxation. What does the LP-relaxation optimal value tell you about the IP optimal value? What if it is a minimization problem?

A:

MINIMIZATION PROBLEM: the optimal value of the linear program relax ation is a lower bound of the IP optimal value.

MAXIMIZATION PROBLEM: the optimal value of the linear program relax ation is an upper bound of the IP optimal value.

Q3: Assume you have a maximization integer program with all integral coefficients in the objective function. Now, suppose you are running the branch and bound algorithm and come across a node with an optimal value of 44.5. The current incumbent is 44. Can you fathom this node? Why or why not?

A:

Yes because it is the currently the best integer solution.

Q4: If the optimal solution to the LP relaxation of the original program is integer, then you have found an optimal solution to your integer program. Explain why this is true.

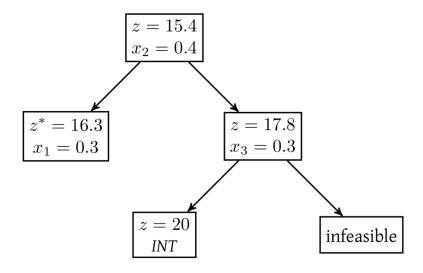
A:

Q5: If the LP is infeasible, then the IP is infeasible. Explain why this is true.

A:

If the LP is infeasible then there cannot exist any IP solution that is feasible because IP solutions are a subset of LP solutions which don't exist.

The next questions ask about the following branch and bound tree. If the solution was not integral, the fractional x_i that was used to branch is given. If the solution was integral, it is denoted *INT*. In the current iteration of branch and bound, you are looking at the node with the ******.



Q6: Can you determine if the integer program this branch and bound tree is for is a minimization or maximixation problem? If so, which is it?

A: It's a minimization problem.

Hint: For **Q7-8**, you can assume integral coefficients in the objective function.

Q7: Is the current node (marked z^*) fathomed? Why or why not? If not, what additional constraints should be imposed for each of the next two nodes?

A:

No it is not fathomed. Each of the next two nodes should each have one of these constraints: x1>=1 or x1<=0

Q8: Consider the nodes under the current node (where z=16.3). What do you know about the optimal value of these nodes? Why?

A:

Optimal values of the node must be at least 16.3.

Part 2: The Knapsack Problem

In this lab, you will solve an integer program by branch and bound. The integer program to be solved will be a knapsack problem.

Knapsack Problem: We are given a collection of n items, where each item $i=1,\ldots,n$ has a weight w_i and a value v_i . In addition, there is a given capacity W, and the aim is to select a maximum value subset of items that has a total weight at most W. Note that each item can be brought at most once.

$$\max \sum_{i=1}^{n} v_{i} x_{i}$$
s.t.
$$\sum_{i=1}^{n} w_{i} x_{i} \leq W$$

$$0 \leq x_{i} \leq 1, \text{ integer, } i = 1, \dots, n$$

Consider the following data which we import from a CSV file:

Out[2]:

item		
1	50	10
2	30	12
3	24	10
4	14	7
5	12	6
6	10	7
7	40	30

value weight

and W = 18.

Q9: Are there any items we can remove from our input to simplify this problem? Why? If so, replace index with the item number that can be removed in the code below. Hint: how many of each item could we possibly take?

A:

Yes, we can already remove input item #7 as it is (on its own) heav ier than the maximum amount of weight we can carry which is 18.

```
In [3]: #TODO: replace index
data = data.drop(7)
```

Q10: If we remove item 7 from the knapsack, it does not change the optimal solution to the integer program. Explain why.

A:

Because item 7 is heavier than maximum weight we can carry on its o wn, so it can never be selected.

Q11: Consider removing items i such that $w_i > W$ from a knapsack input. How does the LP relaxation's optimal value change?

A:

The IP optimal solution shouldn't change at all but the LP relaxati on will.

In **Q10-11**, you should have found that removing these items removes feasible solutions from the linear program but does not change the integer program. This is desirable as the gap between the optimal IP and LP values can become smaller. By adding this step, branch and bound may terminate sooner.

Recall that a branch and bound node can be fathomed if its bound is no better than the value of the best feasible integer solution found thus far. Hence, it helps to have a good feasible integer solution as quickly as possible (so that we stop needless work). To do this, we can first try to construct a good feasible integer solution by a reasonable heuristic algorithm before starting to run the branch and bound procedure.

In designing a heuristic for the knapsack problem, it is helpful to think about the value per unit weight for each item. We compute this value in the table below.

```
In [4]: data['value per unit weight'] = (data['value'] / data['weight']).round(2)
data
```

Out[4]:

item			
1	50	10	5.00
2	30	12	2.50
3	24	10	2.40
4	14	7	2.00
5	12	6	2.00

7

value weight value per unit weight

Q12: Design a reasonable heuristic for the knapsack problem. Note a heuristic aims to find a decent solution to the problem (but is not necessarily optimal).

1.43

A:

6

10

Select the item with the greatest value per unit weight, if including this item surpasses maximum weight limit then skip to the next. Continue until the target weight is exactly reached or all items have been 'processed'.

Q13: Run your heuristic on the data above to compute a good feasible integer solution. Your heuristic should generate a feasible solution with a value of 64 or better. If it does not, try a different heuristic (or talk to your TA!)

A:

```
The heuristic selects items 1 and 4 which give a good feasible integer solution of value 64
```

We will now use the branch and bound algorithm to solve this knapsack problem! First, let us define a mathematical model for the linear relaxation of the knapsack problem.

Q14: Complete the model below.

```
In [5]: def Knapsack(table, capacity, integer = False):
             """Model for solving the Knapsack problem.
            Args:
                 table (pd.DataFrame): A table indexd by items with a column for val
                 capcity (int): An integer-capacity for the knapsack
                 integer (bool): True if the variables should be integer. False other
             ITEMS = list(table.index)
                                              # set of items
            v = table.to_dict()['value']  # value for each item
w = table.to_dict()['weight']  # weight for each item
            W = capacity
                                                # capacity of the knapsack
             # define model
            m = OR.Solver('knapsack', OR.Solver.CBC MIXED INTEGER PROGRAMMING)
             # decision variables
             x = \{\}
             for i in ITEMS:
                 if integer:
                     x[i] = m.IntVar(0, 1, 'x %d' % (i))
                 else:
                     x[i] = m.NumVar(0, 1, 'x %d' % (i))
             # define objective function here
            m.Maximize(sum(v[i]*x[i] for i in ITEMS))
            m.Add((sum(w[i]*x[i] for i in x)) \le W)
             # TODO: Add a constraint that enforces that weight must not exceed capa
             # recall that we add constraints to the model using m.Add()
             return (m, x) # return the model and the decision variables
```

```
In [6]: # You do not need to do anything with this cell but make sure you run it!
def solve(m):
    """Used to solve a model m."""
    m.Solve()

    print('Objective =', m.Objective().Value())
    print('iterations :', m.iterations())
    print('branch-and-bound nodes :',m.nodes())

    return ({var.name() : var.solution_value() for var in m.variables()})
```

We can now create a linear relaxation of our knapsack problem. Now, m represents our model and x represents our decision variables.

```
In [7]: m, x = Knapsack(data, 18)
```

We can use the next line to solve the model and output the solution

Q15: How does this optimal value compare to the value you found using the heuristic integer solution?

A:

This optimal value has non integer inputs. And its optimal value is a bit better than what we obtained in the heuristic integer solutio $n\ (70)$

Q16: Should this node be fathomed? If not, what variable should be branched on and what additional constraints should be imposed for each of the next two nodes?

A:

yes this node should be fathomed and the additional constraints should be added to each branch: x2 <= 0, x2 >= 1

After constructing the linear relaxation model using Knapsack(data1, 18) we can add additional constraints. For example, we can add the constraint $x_2 \le 0$ and solve it as follows:

```
In [9]: m, x = Knapsack(data, 18)
    m.Add(x[2] <= 0)
    solve(m)

Objective = 69.2
    iterations : 0
    branch-and-bound nodes : 0

Out[9]: {'x_1': 1.0, 'x_2': 0.0, 'x_3': 0.8, 'x_4': 0.0, 'x_5': 0.0, 'x_6': 0.0}</pre>
```

NOTE: The line m, x = Knapsack(data1, 18) resets the model m to the LP relaxation. All constraints from branching have to be added each time.

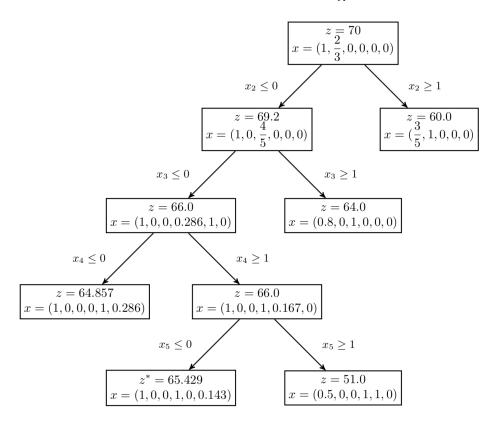
Q17: Use the following cell to compute the optimal value for the other node you found in Q16.

Q18: What was the optimal value? Can this node be fathomed? Why? (Hint: In **Q13**, you found a feasible integer solution with value 64.)

A:

This node should be fathomed as it has a value of 60 that is less t han previously obtained of 64.

If we continue running the branch and bound algorithm, we will eventually reach the branch and bound tree below where the z^* indictes the current node we are looking at.



Q19: The node with z=64.857 was fathomed. Why are we allowed to fathom this node? (Hint: think back to **Q3**)

A:

Because it is more than the best current optimal integer solution (64)

Q20: Finish running branch and bound to find the optimal integer solution. Use a separate cell for each node you solve and indicate if the node was fathomed with a comment. (Hint: Don't forget to include the constraints further up in the branch and bound tree.)

```
In [11]: m, x = Knapsack(data, 18)
         m.Add(x[2] \le 0)
         m.Add(x[3] \le 0)
         m.Add(x[4] >= 1)
         m.Add(x[5] \le 0)
         solve(m)
         Objective = 65.42857142857143
         iterations : 0
         branch-and-bound nodes : 0
Out[11]: {'x 1': 1.0,
           'x_2': 0.0,
          'x 3': 0.0,
          'x_4': 1.0,
          'x 5': 0.0,
          'x 6': 0.14285714285714302}
In [12]: m, x = Knapsack(data, 18)
         m.Add(x[2] \le 0)
         m.Add(x[3] \le 0)
         m.Add(x[4] >= 1)
         m.Add(x[5] \le 0)
         m.Add(x[6] \le 0)
         solve(m)
         Objective = 64.0
         iterations : 0
         branch-and-bound nodes: 0
Out[12]: {'x_1': 1.0, 'x_2': 0.0, 'x_3': 0.0, 'x_4': 1.0, 'x_5': 0.0, 'x_6': 0.0}
In [13]: m, x = Knapsack(data, 18)
         m.Add(x[2] \le 0)
         m.Add(x[3] \le 0)
         m.Add(x[4] >= 1)
         m.Add(x[5] \le 0)
         m.Add(x[6] >= 1)
         solve(m)
         Objective = 44.0
         iterations : 0
         branch-and-bound nodes: 0
Out[13]: {'x_1': 0.4, 'x_2': 0.0, 'x_3': 0.0, 'x_4': 1.0, 'x_5': 0.0, 'x_6': 1.0}
         A:
```

Q21: How many nodes did you have to explore while running the branch and bound algorithm?

Α

Just two additional nodes.

In the next section, we will think about additional constraints we can add to make running branch and bound quicker.

Part 3: Cutting Planes

In general, a cutting plane is an additional constraint we can add to an integer program's linear relaxation that removes feasible linear solutions but does not remove any integer feasible solutions. This is very useful when solving integer programs! Recall many of the problems we have learned in class have something we call the "integrality property". This is useful because it allows us to ignore the integrality constraint since we are garunteed to reach an integral solution. By cleverly adding cutting planes, we strive to remove feasible linear solutions (without removing any integer feasible solutions) such that the optimal solution to the linear relaxation is integral!

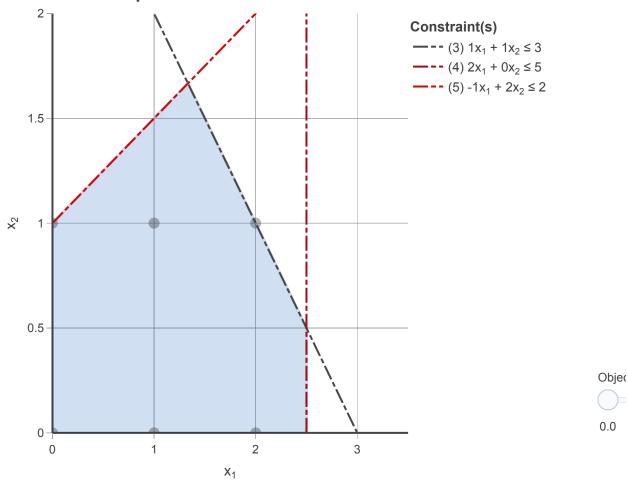
Conisder an integer program whose linear program releaxation is

max
$$2x_1 + x_2$$

s.t. $x_1 + x_2 \le 3$
 $2x_1 \le 5$
 $-x_1 + 2x_2 \le 2$
 $x_1, x_2 \ge 0$

We can define this linear program and then visualize its feasible region. The integer points have been highlighted.

Geometric Interpretation of LPs



Q22: List every feasible solution to the integer program.

A:

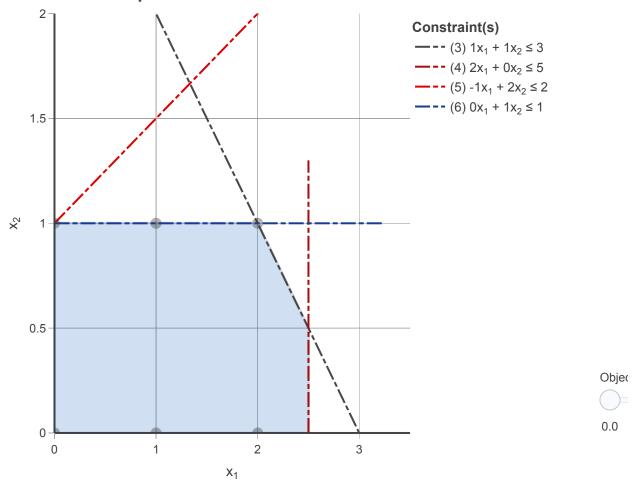
Q23: Is the constraint $x_2 \le 1$ a cutting plane? Why? (Hint: Would any feasible integer points become infeasible? What about feasible linear points?)

Δ.

Yes it's a cutting plane because no feaslible integer solution would become infeasible.

Let's add this cutting plane to the LP relaxation!

Geometric Interpretation of LPs



Q24: Is the constraint $x_1 \le 3$ a cutting plane? Why?

A:

No, because it doesn't restrict any LP relaxation solutions whilst at the same time not restrict any IP feasible solutions.

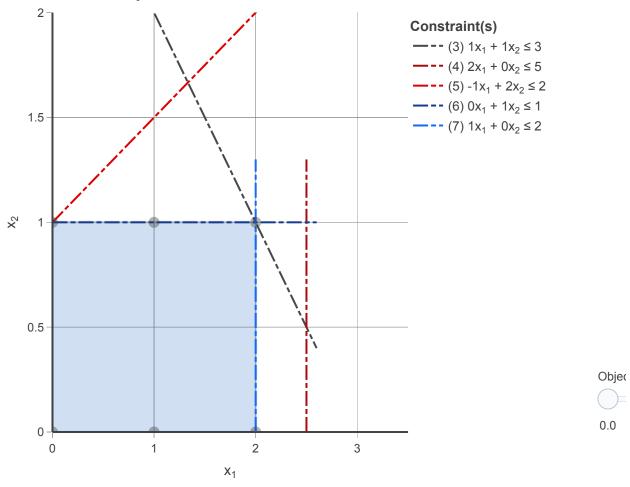
Q25: Can you provide another cutting plane? If so, what is it?

A:

x1 <= 2

Let's look at the feasible region after adding the cutting plane from **Q23** and one of the possible answers from **Q25**. Notice the optimal solution to the LP relaxation is now integral!

Geometric Interpretation of LPs



Let's try applying what we know about cutting planes to the knapsack problem! Again, recall our input was W=18 and:

Out[17]:

value	weight	value per	unit weight
-------	--------	-----------	-------------

item			
1	50	10	5.00
2	30	12	2.50
3	24	10	2.40
4	14	7	2.00
5	12	6	2.00
6	10	7	1.43

Q26: Look at items 1, 2, and 3. How many of these items can we take simultaneously? Can you write a new constraint to capture this? If so, please provide it.

A:

We can only take one of these items simultaneously:
$$x(1)+x(2)+x(3) \le 1$$

Q27: Is the constraint you found in **Q26** a cutting plane? If so, provide a feasible solution to the linear program relaxation that is no longer feasible (i.e. a point the constraint *cuts off*).

A:

Yes,
$$z = 69.2 x1 = 1, x2 = 4/5$$

Q28: Provide another cutting plane involving items 4,5 and 6 for this integer program. Explain how you derived it.

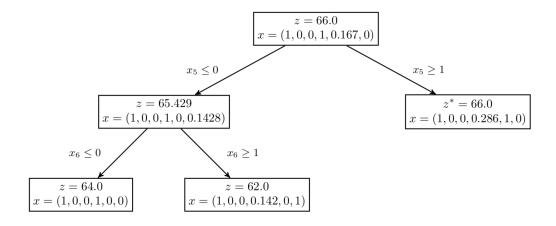
A:

$$x(4)+x(5)+x(6) <= 2$$

Q29: Add the cutting planes from **Q26** and **Q28** to the model and solve it. You should get a solution in which we take items 1 and 4 and $\frac{1}{6}$ of item 5 with an objective value of 66.

Let's take a moment to pause and reflect on what we are doing. Recall from **Q9-11** that we dropped item 7 becuase its weight was greater than the capcity of the knapsack. Essentially we added the constraint $x_7 \le 0$. This constraint was a cutting plane! It eliminated some linear feasible solutions but no integer ones. By adding these two new cutting planes, we can get branch and bound to terminate earlier yet again! So far, we have generated cutting planes by inspection. However, there are more algorithmic ways to identify them (which we will ignore for now).

If we continue running the branch and bound algorithm, we will eventually reach the branch and bound tree below where the z^* indictes the current node we are looking at.



NOTE: Do not forget about the feasible integer solution our heuristic gave us with value 64.

Q30 Finish running branch and bound to find the optimal integer solution. Use a separate cell for each node you solve and indicate if the node was fathomed with a comment. Hint: Don't forget the cutting plane constraints should be included in every node of the branch and bound tree.

```
In [20]: # Template
         m, x = Knapsack(data, 18)
         m.Add(x[1] + x[2] + x[3] \le 1)
         m.Add(x[4] + x[5] + x[6] \le 2)
         m.Add(x[5] \le 0)
         m.Add(x[6] >= 1)
         solve(m)
         # fathomed?
         Objective = 62.0
         iterations : 0
         branch-and-bound nodes : 0
Out[20]: {'x 1': 1.0,
          'x_2': 0.0,
          'x 3': 0.0,
          'x 4': 0.14285714285714293,
          'x 5': 0.0,
          'x 6': 1.0}
In [22]: m, x = Knapsack(data, 18)
         m.Add(x[1] + x[2] + x[3] \le 1)
         m.Add(x[4] + x[5] + x[6] \le 2)
         m.Add(x[5] \le 0)
         m.Add(x[6] >= 1)
         m.Add(x[4] >= 1)
         solve(m)
         Objective = 44.0
         iterations : 0
         branch-and-bound nodes : 0
Out[22]: {'x 1': 0.4, 'x 2': 0.0, 'x 3': 0.0, 'x 4': 1.0, 'x 5': 0.0, 'x 6': 1.0}
In [21]: m, x = Knapsack(data, 18)
         m.Add(x[1] + x[2] + x[3] \le 1)
         m.Add(x[4] + x[5] + x[6] \le 2)
         m.Add(x[5] \le 0)
         m.Add(x[6] >= 1)
         m.Add(x[4] \le 0)
         solve(m)
         Objective = 60.0
         iterations : 0
         branch-and-bound nodes: 0
Out[21]: {'x 1': 1.0, 'x 2': 0.0, 'x 3': 0.0, 'x 4': 0.0, 'x 5': 0.0, 'x 6': 1.0}
         A:
```

Q31: Did you find the same optimal solution? How many nodes did you explore? How did this compare to the number you explored previously?

A:

Yes, the same optimal solution was obtained (64) and two nodes wher e explored.

Part 4: Geometry of Branch and Bound

Previously, we used the gilp package to viusualize the simplex algorithm but it also has the functionality to visualize branch and bound. We will give a quick overview of the tool. Similar to lp_visual and simplex_visual, the function bnb_visual takes an LP and returns a visualization. It is assumed that every decision variable is constrained to be integer. Unlike previous visualizations, bnb_visual returns a series of figures for each node of the branch and bound tree. Let's look at a small 2D example:

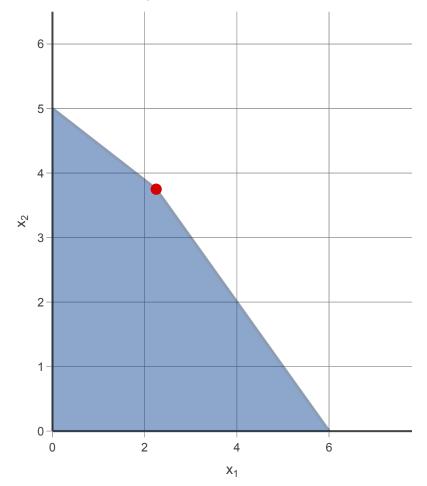
max
$$5x_1 + 8x_2$$

s.t. $x_1 + x_2 \le 6$
 $5x_1 + 9x_2 \le 45$
 $x_1, x_2 \ge 0$, integral

In [23]: nodes = gilp.bnb_visual(gilp.examples.STANDARD_2D_IP)

In [26]: nodes[0].show()

Geometric Interpretation of LPs





Run the cells above to generate a figure for each node and view the first node. At first, you will see the LP relaxation on the left and the root of the branch and bound tree on the right. The simplex path and isoprofit slider are also present.

Q32: Recall the root of a branch and bound tree is the unaltered LP relaxation. What is the optimal solution? (Hint: Use the objective slider and hover over extreme points).

A:

41.25

Q33: Assume that we always choose the variable with the minimum index to branch on if there are multiple options. Write down (in full) each of the LPs we get after branching off the root node.

A:

ON PAPER

Q34: Draw the feasible region to each of the LPs from Q33 on the same picture.

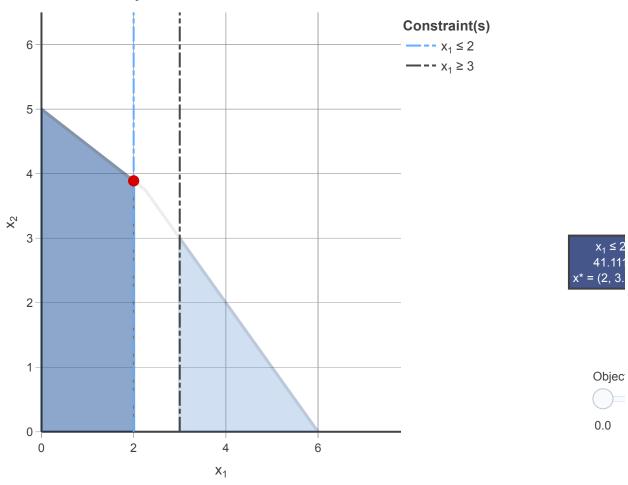
A:

ON PAPER

Run the following cell to see if the picture you drew in Q34 was correct.

In [27]: nodes[1].show()

Geometric Interpretation of LPs



The outline of the original LP relaxation is still shown on the left. Now that we have eliminated some of the fractional feasible solutions, we now have 2 feasible regions to consider. The darker one is the feasible region associated with the current node which is also shaded darker in the branch and bound tree. The unexplored nodes in the branch and bound tree are not shaded in.

Q35: Which feasible solutions to the LP relaxation are removed by this branch?

A:

All the solution in the range 2<=x1

Q36: At the current (dark) node, what constraints will we add? How many feasible regions will the original LP relaxation be broken into?

A:

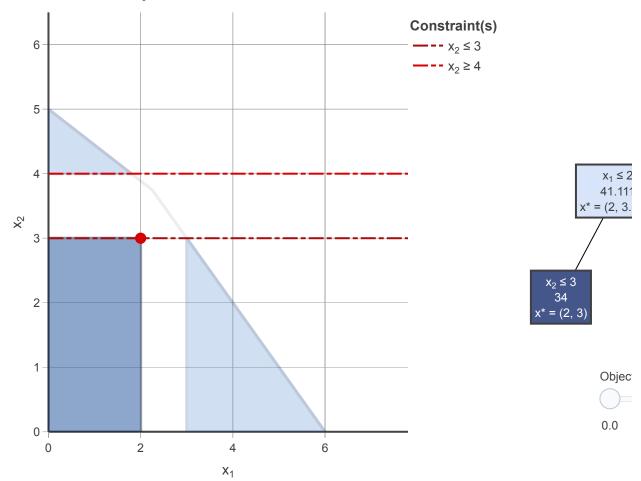
x2 <= 3

or

x2 > = 4

In [28]: nodes[2].show()

Geometric Interpretation of LPs



Q37: What is the optimal solution at the current (dark) node? Do we have to further explore this branch? Explain.

A:

Optimal solition at the current dart node is 34. No further explora tion can be done in the branch as this is already the optimal solut ion which happens to be an integer.

Q38: Recall shaded nodes have been explored and the node shaded darker (and feasible region shaded darker) correspond to the current node and its feasible region. Nodes not shaded have not been explored. How many nodes have not yet been explored?

A:

There are 2 node that have been unexplore so far $x^2 = 4$ and $x^1 > 3$

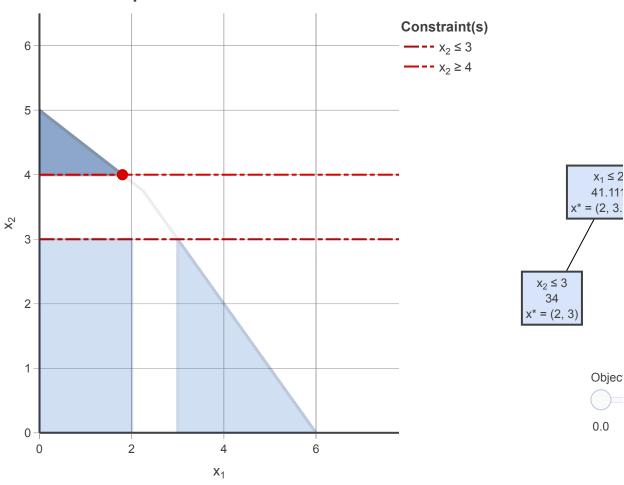
Q39: How many nodes have a degree of one in the branch and bound tree? (That is, they are only connected to one edge). These nodes are called leaf nodes. What is the relationship between the leaf nodes and the remaining feasible region?

A:

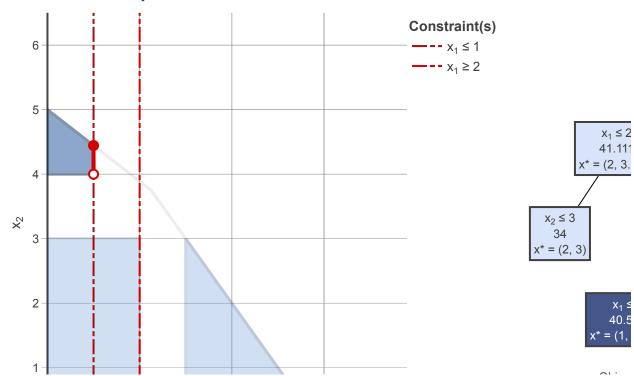
Only two nodes have a degree of 1 and they are the regions which have yet to be explored.

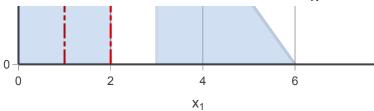
In [29]: # Show the next two iterations of the branch and bound algorithm
nodes[3].show()
nodes[4].show()

Geometric Interpretation of LPs



Geometric Interpretation of LPs







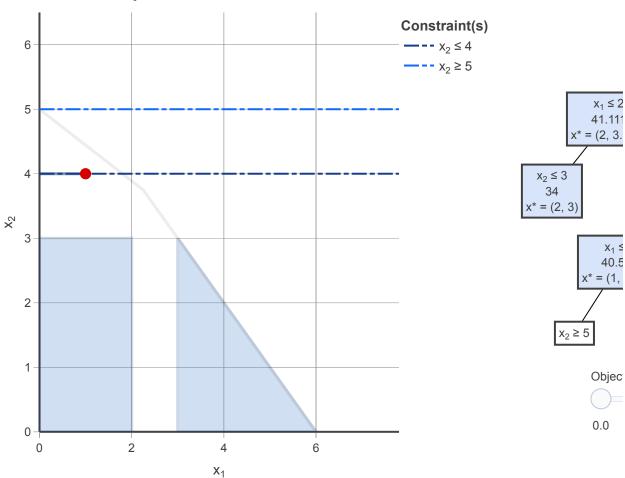
Q40: At the current (dark) node, we added the constraint $x_1 \le 1$. Why were the fractional solutions $1 < x_1 < 2$ not eliminated for $x_2 <= 3$?

A:

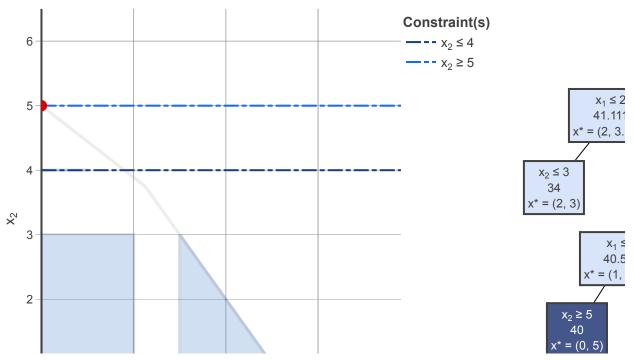
Because they are still within the region of $x2 \le 3$

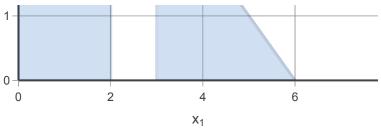
In [30]: # Show the next three iterations of the branch and bound algorithm
 nodes[5].show()
 nodes[6].show()
 nodes[7].show()

Geometric Interpretation of LPs



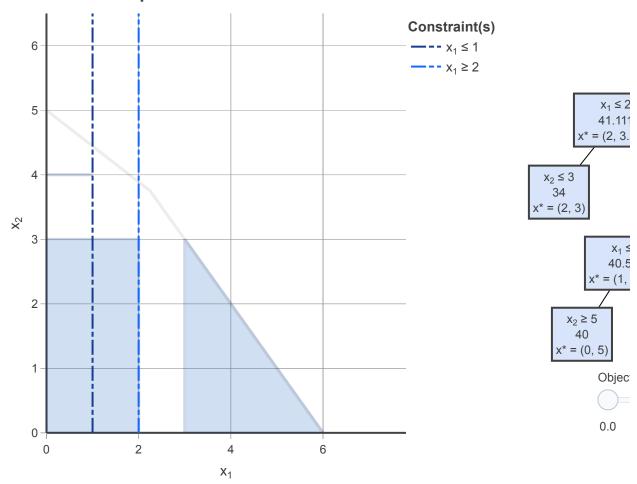
Geometric Interpretation of LPs







Geometric Interpretation of LPs



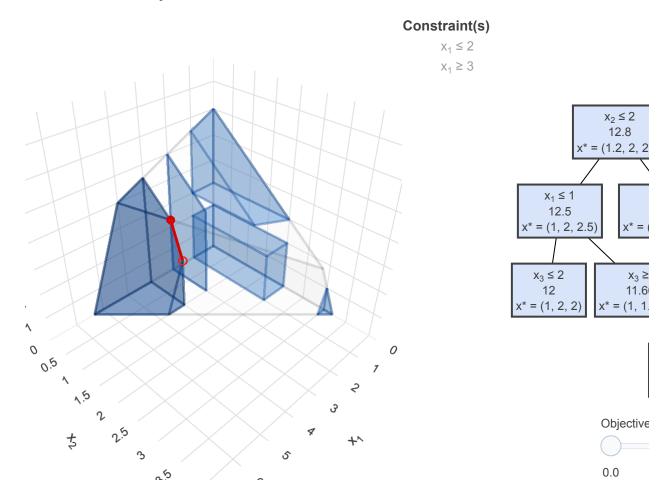
Q41: What constraints are enforced at the current (dark) node? Why are there no feasible solutions at this node?

A:

Because we have already 'explored' that region

In [35]: nodes[8].show()

Geometric Interpretation of LPs



Q42: Are we done? If so, what nodes are fathomed and what is the optimal solution? Explain.

A:

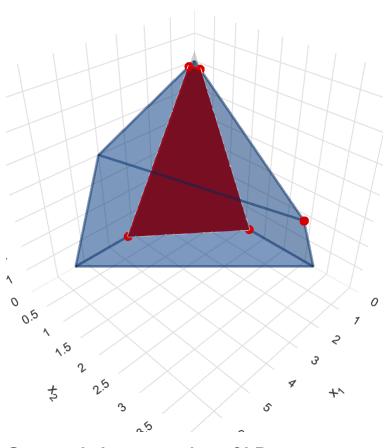
Optimal solution is 13.

Let's look at branch and bound visualization for an integer program with 3 decision variables!

In [37]: nodes = gilp.bnb_visual(gilp.examples.VARIED_BRANCHING_3D_IP)

```
In [33]: # Look at the first 3 iterations
nodes[0].show()
nodes[1].show()
nodes[2].show()
```

Geometric Interpretation of LPs



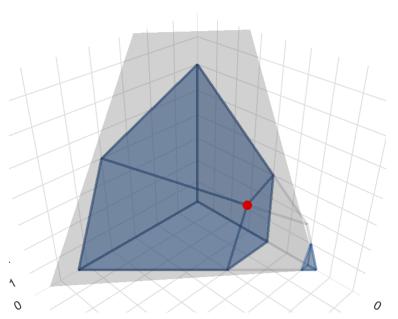
Objective

0.0

Geometric Interpretation of LPs

Constraint(s)

 $x_2 \le 2$ $x_2 \ge 3$



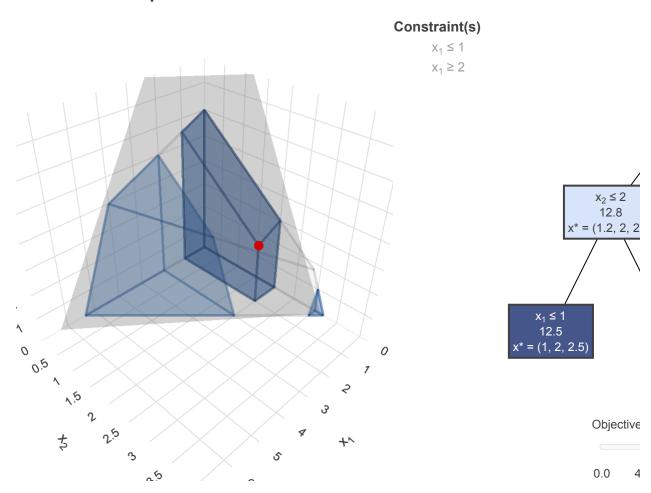




Objective

0.0 4

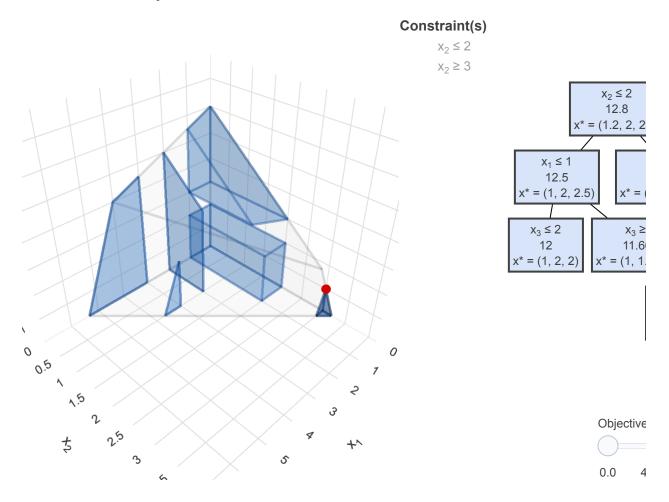
Geometric Interpretation of LPs



Let's fast-forward to the final iteration of the branch and bound algorithm.

In [34]: nodes[-1].show()

Geometric Interpretation of LPs



Q43: Consider the feasible region that looks like a rectangular box with one corner point at the origin. What node does it correspond to in the tree? What is the optimal solution at that node?

A:

corresponds to solution on 0

Q44: How many branch and bound nodes did we explore? What was the optimal solution? How many branch and bound nodes would we have explored if we knew the value of the optimal solution before starting branch and bound?

A:

We could have drastically reduced the number of nodes explored if w e knew from the start that the optimal solution was 13.

Bonus: Branch and Bound for Knapsack

Consider the following example:

item	value	weight
1	2	1
2	9	3
3	6	2

The linear program formulation will be:

max
$$2x_1 + 9x_2 + 6x_3$$

s.t. $1x_1 + 3x_2 + 2x_3 \le 10$
 $x_1, x_2, x_3 \ge 0$, integer

In gilp, we can define this lp as follows: