CS 170 Homework 7

Due **3/14/2022, at 10:00 pm (grace period until 11:59pm)**

1 Study Group

List the names and SIDs of the members in your study group. If you have no collaborators, you must explicitly write "none".

2 Modeling: Tricks of the Trade

One of the most important problems in the field of *statistics* is the *linear regression problem*. Roughly speaking, this problem involves fitting a straight line to statistical data represented by points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ on a graph. Denoting the line by $y = a + bx$, the objective is to choose the constants a and b to provide the "best" fit according to some criterion. The criterion usually used is the *method of least squares*, but there are other interesting criteria where linear programming can be used to solve for the optimal values of a and b.

Suppose instead we wish to minimize the sum of the absolute deviations of the data from the line, that is,

$$\min \sum_{i=1}^{n} |y_i - (a + bx_i)|$$

Write a linear program with variables a, b to solve this problem.

Hint: Create a new variable z_i that will equal $|y_i - (a + bx_i)|$ in the optimal solution.

Solution:

Note that the smallest value of z that satisfies $z \geq x, z \geq -x$ is $z = |x|$.

Now, consider the following linear programming problem:

$$\min \sum_{i=1}^{n} z_i$$
$$\text{subject to} \begin{cases} y_i - (a + bx_i) \leq z_i & \text{for } 1 \leq i \leq n \\ (a + bx_i) - y_i \leq z_i & \text{for } 1 \leq i \leq n \end{cases}.$$

Since $\sum_i z_i$ is minimized, z_i will be set to the $\max(y_i - (a + bx_i), (a + bx_i) - y_i)$. Note that $\max(y_i - (a + bx_i), (a + bx_i) - y_i)$ is, in fact, $|y_i - (a + bx_i)|$.

If for some solution we have that $z_i > |y_i - (a + bx_i)|$, then by setting $z_i = |y_i - (a + bx_i)|$ we will get a solution with a smaller value of the objective function, therefore the initial solution was not optimal. Hence, the constraints requires that the optimal solution will set $z_i = |y_i - (a + bx_i)|$, so the new problem is indeed equivalent to the original problem. However, now it is a linear programming problem.

3 Jeweler

You are a jeweler who sells necklaces and rings. Each necklace takes 4 ounces of gold and 2 diamonds to produce, each ring takes 1 ounce of gold and 3 diamonds to produce. You have 80 ounces of gold and 90 diamonds. You make a profit of 60 dollars per necklace you sell and 30 dollars per ring you sell, and want to figure out how many necklaces and rings to produce to maximize your profits.

(a) Formulate this problem as a linear programming problem. Draw the feasible region, and find the solution (state the cost function, linear constraints, and all vertices except for the origin).

(b) Suppose instead that the profit per necklace is C dollars and the profit per ring remains at 30 dollars. For each vertex you listed in the previous part, give the range of C values for which that vertex is the optimal solution.

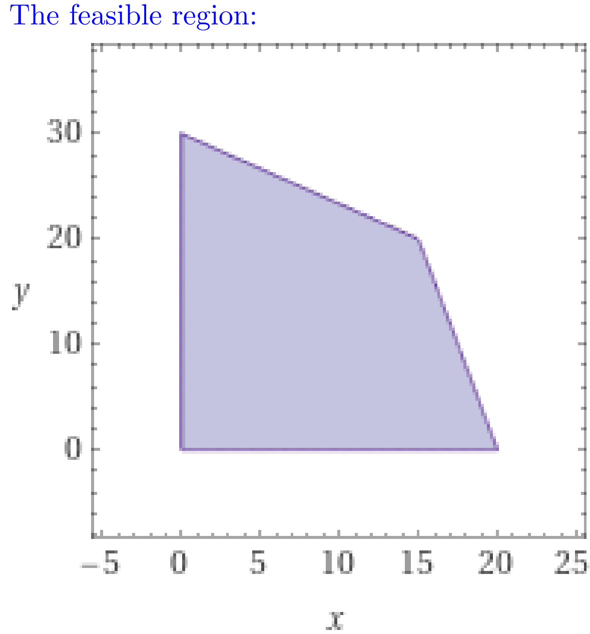
Solution:

(a) x = number of necklaces
y = number of engagement rings

Maximize: $60x + 30y$

Linear Constraints:
$4x + y \leq 80$
$2x + 3y \leq 90$
$x \geq 0$
$y \geq 0$

The feasible region:



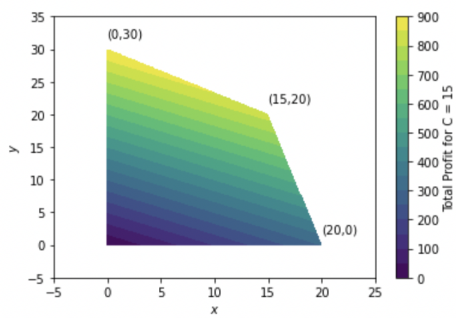
The vertices are $(x = 20, y = 0), (x = 15, y = 20), (x = 0, y = 30)$, and the objective is maximized at $(x = 15, y = 20)$, where $60x + 30y = 1500$.

(b) There are lots of ways to solve this part. The most straightforward is to write and solve a system of inequalities checking when the objective of one vertex is at least as large as the objective of the other vertices. For example, for $(x = 15, y = 20)$ the system of inequalities would be $C \cdot 15 + 30 \cdot 20 \geq C \cdot 20$ and $C \cdot 15 + 30 \cdot 20 \geq 30 \cdot 30$. Doing this for each vertex gives the following solution:

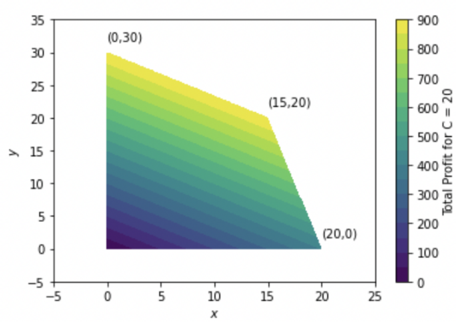
$(x = 0, y = 30) : C \leq 20$
$(x = 15, y = 20) : 20 \leq C \leq 120$
$(x = 20, y = 0) : 120 \leq C$

One should note that there is a nice geometric interpretation for this solution: Looking at the graph of the feasible region, as C increases, the vector $(C, 30)$ starts pointing closer to the x-axis. The objective says to find the point furthest in the direction of this vector, so the optimal solution also moves closer to the x-axis as C increases. When $C = 20$ or $C = 120$, the vector $(C, 30)$ is perpendicular to one of the constraints, and there are multiple optimal solutions all lying on that constraint, which are all equally far in the direction $(C, 30)$.

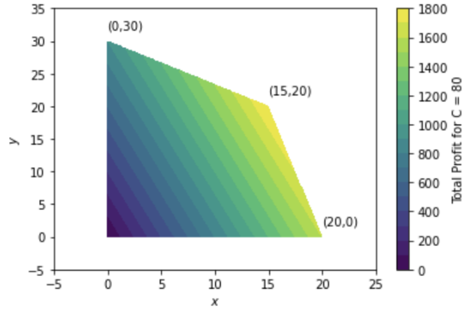
Below are some graphics to help build up your intuition about this problem.
C = 15 : (x = 0, y = 30) optimal



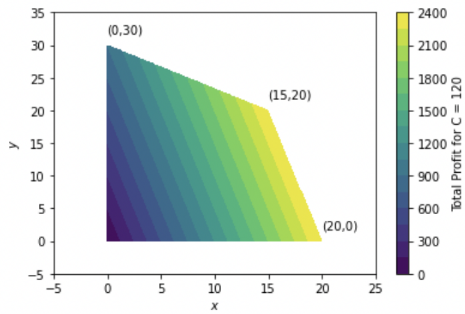
C = 20 : (x = 0, y = 30) and (x = 15, y = 20) optimal



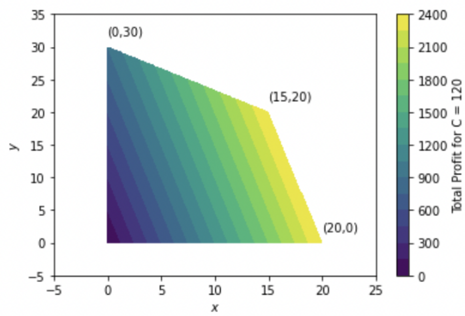
C = 80 : (x = 0, y = 30) optimal



C = 120: (x = 15, y = 20) and (x = 20, y = 0) optimal



C = 200 : (x = 20, y = 0) optimal



4 Standard Form LP

Recall that any Linear Program can be reduced to a more constrained *standard form* where all variables are nonnegative, the constraints are given by equations and the objective is that of minimizing a cost function.

More formally, our variables are x_i. Our objective is $\min c^\top x = \sum_i c_i x_i$ for some constants c_i. The jth constraint is $\sum_i a_{ij} x_i = b_j$ for some constants a_{ij}, b_j. Finally, we also have the constraints $x_i \geq 0$.

An example standard form LP:

$$\text{minimize } 5x_1 + 3x_2$$
$$s.t. \quad x_1 + x_2 - x_3 = 1, \quad -(x_1 + x_2 - x_3) = -1, \quad -x_1 + 2x_2 + x_4 = 0, \quad -(-x_1 + 2x_2 + x_4) = 0, \quad x_1, x_2, x_3, x_4 \geq 0$$

For each of the subparts, what system of variables, constraints, and objectives would be equivalent to the following:

(a) Max Objective: $\max \sum_i c_i x_i$

(b) Upper Bound on Variable: $x_1 \leq b_1$

(c) Lower Bound on Variable: $x_2 \geq b_2$

(d) Bounded Variable: $b_2 \leq x_3 \leq b_1$

(e) Inequality Constraint: $x_1 + x_2 + x_3 \leq b_3$

(f) Min Max Objective: $\min \max(y_1, y_2)$

(g) Unbounded Variable: $x_4 \in R$

Solution:

(a) $\min -\sum_i c_i x_i$

(b) $x_1 + s_1 = b_1 \,, \quad s_1 \geq 0$

(c) $-x_2 + s_2 = -b_2 \quad s_2 \geq 0$

(d) Break it into two inequalities $x_3 \leq b_1$ and $x_3 \geq b_2$ and use the parts above

(e) $x_1 + x_2 + x_3 + s_1 = b_3 \,, \quad s_1 \geq 0$

(f) $\min t \,, \quad x \leq t \,, \quad y \leq t$

(g) Replace x_4 by $x^+ - x^-$ along with $x^+ \geq 0 \,, \quad x^- \geq 0$

5 Special Points

Let $G = (V, E)$ be a simple, connected, undirected graph. We say a vertex v is a *special point* if, after we delete v and all edges incident to it, G becomes disconnected. Write an algorithm to identify all special points in G in $O(|V| + |E|)$ time.

This is a coding problem, so please write code at `https://judge.cs170.org/problem/4294967302`. It is **highly recommended** that you design the whole algorithm first and convince yourself it is correct before writing any code.

The sample code handles all input and output formatting, converting the graph into an adjacency-list represented as a Python dictionary. In addition, it performs a recursive depth-first-search traversal of the graph. We suggest that you modify this traversal for your purposes. However, you are free to ignore/edit the sample code as you see fit, so long as the test cases pass.

To run your code and see output/errors, click "Validate" on the online judge. However, to receive full credit, you must click "Submit" and pass the hidden test cases (for which you will not see any output). You can also run your code locally, by calling the `solution` function and passing in a graph, represented as an adjacency list using a Python dictionary.

Please report any technical issues on Piazza.