Optimization Breakouts

Zachary Schutzman, Sarah Cannon, Austin Buchanan, Justin Solomon

A brief, hands-on introduction to four topics in optimization.

Contents

Version 1.0 | Voting Rights Data Institute, June 2019

1 Day 3: *Linear and Integer Programming*

1.1 The Diet Problem, following Chvátal (1983)

Polly wonders how much money she must spend on food in order to get all the energy (2000 kcal), protein (55 g), and calcium (800 mg) that she needs every day. (For other nutrients, she will depend on pills. Nutritionists would disapprove, but the introductory example ought to be simple.) She chooses 6 foods that seem to be cheap sources of the nutrients, collected in the table below.

Food	Serving Size	Energy (kcal)	Protein (g)	Calcium (mg)	Price/serving (cents)
Oatmeal	28 g	110	4	2	3
Chicken	100 g	205	32	12	24
Eggs	2 large	160	13	54	13
Whole milk	237 cc	160	8	285	9
Cherry pie	170 g	420	4	22	20
Pork w/ beans	260 g	260	14	80	19

If she chose to eat only pork w/ beans, how many servings would she need? How costly is this?

She determines that this would be too much to stomach, and decides to limit herself to:

Oatmeal	At most 4 servings per day
Chicken	At most 3 servings per day
Eggs	At most 2 servings per day
Milk	At most 8 servings per day
Cherry pie	At most 2 servings per day
Pork w/ beans	At most 2 servings per day

After taking another look at the data, Polly sees that 8 servings of milk and 2 servings of cherry pie will satisfy the requirements at a cost of:

She could cut down a little on the milk or the pie or perhaps try a different combination. But there are so many combinations that one could go on and on trying them out. Trial and error isn't very helpful here.

To be systematic, we may speculate about some as yet unspecified menu consisting of x_1 servings of oatmeal, x_2 servings of chicken, x_3 servings of eggs, and so on. In this case, what lower and upper bounds must the menu satisfy?

To meet the requirements for energy, protein, and calcium, the menu must satisfy:

If some numbers x_1, x_2, x_3, x_4, x_5, x_6 satisfy these inequalities, then they describe a satisfactory menu. Such a menu will cost how much (in cents) per day?

To design the most economical menu, Polly wants to find numbers x_1, x_2, x_3, x_4, x_5, x_6 that satisfy the inequalities above and make the cost expression as small as possible.

A mathematical optimizer would say she wants to solve the *linear program* (LP):

$$
\begin{aligned}
\text{minimize} \quad & 3x_1 + 24x_2 + 13x_3 + 9x_4 + 20x_5 + 19x_6 \\
\text{subject to} \quad & 110x_1 + 205x_2 + 160x_3 + 160x_4 + 420x_5 + 260x_6 \geq 2000 \\
& 4x_1 + 32x_2 + 13x_3 + 8x_4 + 4x_5 + 14x_6 \geq 55 \\
& 2x_1 + 12x_2 + 54x_3 + 285x_4 + 22x_5 + 80x_6 \geq 800 \\
& 0 \leq x_1 \leq 4 \\
& 0 \leq x_2 \leq 3 \\
& 0 \leq x_3 \leq 2 \\
& 0 \leq x_4 \leq 8 \\
& 0 \leq x_5 \leq 2 \\
& 0 \leq x_6 \leq 2.
\end{aligned}
$$

We have seen a menu with cost \$1.12. If we solve the LP from above, we find a different menu: 4 servings of oatmeal, 4.5 servings of milk, and 2 servings of cherry pie.

This menu $x^* = (4, 0, 0, 4.5, 2, 0)^T$ is *feasible*, and its *objective value* is 92.5 cents.

Remark. Polly might not want her menu to have fractional serving sizes. (Can you buy $x_5 = 5.38$ cherry pies?) This can be handled by requiring that x take integer values, yielding an *integer program*. If only x_5 were required to be an integer, this would yield a *mixed integer program*.

Exercise. Solve Chvátal's Diet Problem with CVXPY and confirm that 92.5 is the optimal cost.

1.2 Some notes about LPs

1. LPs can have a max or min objective.
2. Constraints can be $\leq$, $\geq$, or $=$.
3. The expressions in the objective and constraints must be linear functions of x.
4. By some tricks, any LP can be converted to this form: $\min\{c^T x \mid Ax \geq b, x \geq 0\}$, where A is a matrix and x, c, and b are column vectors of appropriate size.
5. A is the *constraint matrix*; b is the *right-hand-side*; x contains the *decision variables*, and c contains the *objective coefficients*.
6. In practice, A is usually very sparse (i.e., contains mostly zeros).
7. LPs can have 0, 1, or infinitely many feasible/optimal solutions. (Why?)

1.3 The 0-1 Knapsack Problem

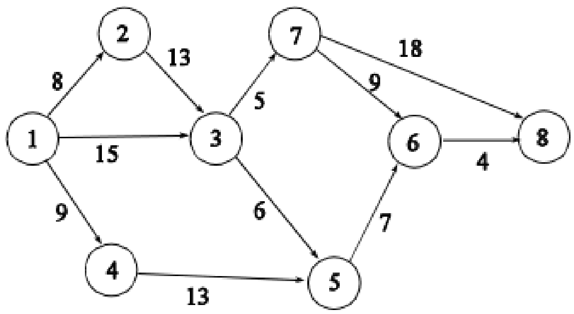
After a long day at Mike E. Mozzarella's, you have collected b tickets which you would like to spend on toys. There are n different toys (which you have numbered $1, 2, \ldots, n$). Suppose that toy i gives you u_i "utils" of enjoyment and costs c_i tickets. Supposing that you purchase at most one of each toy, write an integer program to maximize the number of utils obtained while not exceeding your ticket budget. (Suppose that a subset $S \subseteq [n]$ of toys gives you $\sum_{i \in S} u_i$ utils of enjoyment.)

Identify constraints that could be added to your IP to ensure that:

- toy 1 cannot be chosen if toy 3 is chosen;
- toy 4 can be chosen only if toy 2 is chosen;
- you must choose both toys 1 and 5 or neither of them;
- if you choose toy 1, then you must choose toy 6 or toy 7 (or both).

1.4 The Shortest Path Problem

Formulate an integer program to find a shortest path from vertex 1 to vertex 8.



Remark 1. The constraint matrix for this shortest path formulation is *totally unimodular* and the right-hand-sides are integer. So, by Hoffman and Kruskal (1956), the associated linear programming relaxation has integer extreme points, implying that it suffices to solve the LP.

Exercise. Solve the shortest path problem above with CVXPY. Confirm that the answer does not change when you solve the linear programming relaxation.

Exercise. Formulate an integer program to find a shortest path from s to t in an arbitrary directed graph $G = (V, A)$. You can suppose that the length c_{ij} of each arc $(i, j) \in A$ is positive. What can go wrong when the arc lengths can be negative?

1.5 Further Exercises

Consider the following abstract from the paper "Locating the representational baseline: Republicans in Massachusetts" by Duchin et al. (2018).

> Abstract. Republican candidates often receive between 30% and 40% of the two-way vote share in statewide elections in Massachusetts. For the last three Census cycles, MA has held 9-10 seats in the House of Representatives, which means that a district can be won with as little as 6% of the statewide vote. Putting these two facts together, one may be surprised to learn that a Massachusetts Republican has not won a seat in the U.S. House of Representatives since 1994. We argue that the underperformance of Republicans in Massachusetts is not attributable to gerrymandering, nor to the failure of Republicans to field House candidates, but is a structural mathematical feature of the distribution of votes. For several of the elections studied here, there are more ways of building a valid districting plan than there are particles in the galaxy, and every one of them will produce a 9-0 Democratic delegation.

This motivates the following problem. A state has n indivisible units numbered $1, 2, \ldots, n$. Unit i contains p_i people in it, of which r_i reliably vote Republican and d_i reliably vote Democratic. (Here, $r_i + d_i \leq p_i$ and equality probably does not hold.). A district requires between L and U people in it. For simplicity, suppose that no other constraints (e.g., contiguity) are imposed. Of the possible districts $S \subset [n]$ that satisfy the population bounds $L \leq \sum_{i \in S} p_i \leq U$, which would likely exhibit the strongest Republican performance, i.e., maximize $\sum_{i \in S}(r_i - d_i)$?

1. Write an integer program for this problem. If you solve it, how will you know if a majority-Republican district is possible?
2. Implement your integer program in CVXPY and solve one of the instances from `https://github.com/gerrymandr/Massachusetts_underperformance`.
3. Write an integer program for this problem if contiguity is required.