Note: Your TA probably will not cover all the problems. This is totally fine, the discussion worksheets are deliberately made long so they can serve as a resource you can use to practice, reinforce, and build upon concepts discussed in lecture, readings, and the homework.

1 LP Basics

Linear Program. A *linear program* is an optimization problem that seeks the best value of a linear objective over linear constraints. Let $x \in \mathbb{R}^n$ be the variables, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. One standard canonical form is

$$\begin{aligned} \text{maximize } & \langle c, x \rangle \\ \text{subject to } & Ax \leq b \\ & x \geq 0. \end{aligned}$$

Here, $\langle c, x \rangle = \sum_{i=1}^{n} c_i x_i$ is the inner product of c and x.

Example. Consider the LP

$$\begin{aligned} \text{maximize } & 3x_1 + 2x_2 \\ \text{subject to } & x_1 + x_2 \leq 4 \\ & 2x_1 + x_2 \leq 5 \\ & x_1, x_2 \geq 0. \end{aligned}$$

This same LP can be written in matrix form as

$$\begin{aligned} \text{maximize } & \left\langle \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle \\ \text{subject to } & \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 4 \\ 5 \end{pmatrix} \\ & \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

So in this example,

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \qquad c = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Let's check that many LPs can still be converted into this canonical form:

(i) What if the objective is minimization instead of maximization?

(ii) What if one of the constraints is $a^\top x \geq d$?

(iii) What if one of the constraints is $a^\top x = d$?

(iv) What if one of the variables is unrestricted in sign, so it can be positive or negative?

(v) What if a variable is constrained by $x \geq -2$?

2 LP Meets Linear Regression

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:

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

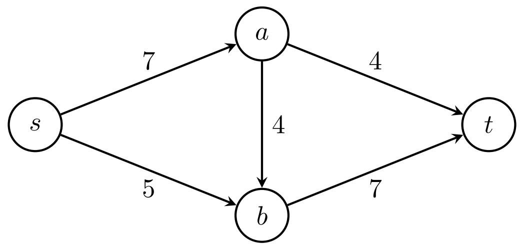
Write a linear program with variables a, b to solve this problem.

Hint: Create new variables z_i and new constraints to help represent $|y_i - (a + bx_i)|$ in a linear program.

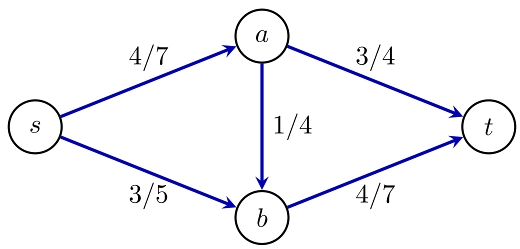
3 Max Flow = Min Cut via LP Duality

In lecture, we discussed both max flow and LP duality. In this problem, you will connect them and derive the max-flow min-cut theorem from strong duality.

Consider the following flow network:



Here is one *feasible* flow on this graph; each edge is labeled as *flow/capacity*. This picture is only meant to make the edge variables concrete.



We will use one flow variable per edge:

$$f_{sa},\ f_{sb},\ f_{ab},\ f_{at},\ f_{bt}.$$

An LP for the max-flow problem is

$$\begin{aligned} \text{maximize } & f_{at} + f_{bt} \\ \text{subject to } & f_{sa} - f_{ab} - f_{at} = 0 \qquad \text{(flow conservation at } a\text{)} \\ & f_{sb} + f_{ab} - f_{bt} = 0 \qquad \text{(flow conservation at } b\text{)} \\ & f_{sa} \le 7, \quad f_{sb} \le 5, \quad f_{ab} \le 4, \quad f_{at} \le 4, \quad f_{bt} \le 7 \\ & f_{sa}, f_{sb}, f_{ab}, f_{at}, f_{bt} \ge 0. \end{aligned}$$

(a) The LP above came from this specific graph. Now generalize it. Let the source be s, sink t, and capacity c_e on each edge $e \in E$. Assume for simplicity that no edges enter s and no edges leave t. Use one variable f_{uv} for each edge $(u, v) \in E$.

(b) Write the dual of your general max-flow LP. Use a nonnegative dual variable z_{uv} for each capacity constraint, and a free dual variable p_v for each flow-conservation constraint.

(c) Return to the example graph, and consider the cut (S, T) where

$$S = \{s, a, b\}, \qquad T = \{t\}.$$

Draw this cut on the graph and try to assign values to the dual variables in the most natural possible way:

- vertices on the s-side should get one label,
- vertices on the t-side should get another,

- edges crossing the cut should be treated differently from edges that do not.

Use your assignment to produce a feasible dual solution, and compute its objective value.

(d) Generalize the previous part to an arbitrary s-t cut. Explain why the dual objective becomes exactly the cut capacity. Then, assuming strong duality, explain why solving the dual is the same as solving the min-cut problem.

4 Cal Day Hosting

For Cal Day, Berkeley students volunteer to host visiting prospective students for the day. Each Berkeley student can host at most one visitor, and each visitor can be hosted by at most one Berkeley student.

Not every pair is compatible. For example:

- some visitors only want hosts from a particular major,
- some hosts can only accommodate visitors with compatible dietary or allergy needs,
- some visitors want hosts from a specific dorm or campus community,
- some pairs have schedule conflicts and cannot be matched.

Suppose there are m Berkeley hosts and n visiting students, and for each pair (i, j) we know whether host i is allowed to host visitor j.

Design a polynomial-time algorithm that matches as many visitors to hosts as possible.

