EECS 16A Designing Information Devices and Systems I
Spring 2018 Homework 4

This homework is due February 19, 2018, at 23:59.
Self-grades are due February 22, 2018, at 23:59.

Submission Format
Your homework submission should consist of **one** file.

- `hw4.pdf`: A single PDF file that contains all of your answers (any handwritten answers should be scanned).

Submit the file to the appropriate assignment on Gradescope.

1. **Finding Null Spaces**

   (a) Consider the column vectors of any 3×5 matrix. What is the maximum possible number of linearly independent vectors you can pick from these column vectors?

   (b) Suppose we have the following 3×5 matrix after row reduction:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & -2 & 3 \\ 0 & 0 & 2 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

   What is the minimum number of vectors spanning the column space of $\mathbf{A}$? Find a set of such vectors.

   (c) Recall that for every vector $\vec{x}$ in the null space of $\mathbf{A}$, $\mathbf{A}\vec{x} = \vec{0}$. The dimension of a the null space is the minimum number of vectors needed to span it. Find vectors that span the null space of $\mathbf{A}$ (the matrix in the previous part). What is the dimension of the null space of $\mathbf{A}$?

   (d) Find vector(s) that span the null space of the following matrix:

$$\mathbf{B} = \begin{bmatrix} 2 & -4 & 4 & 8 \\ 1 & -2 & 3 & 6 \\ 2 & -4 & 5 & 10 \\ 3 & -6 & 7 & 14 \end{bmatrix}$$

2. **Traffic Flows**

   Your goal is to measure the flow rates of vehicles along roads in a town. However, it is prohibitively expensive to place a traffic sensor along every road. You realize, however, that the number of cars flowing into an intersection must equal the number of cars flowing out. You can use this "flow conservation" to determine the traffic along all roads in a network by only measuring flow along only some roads. In this problem, we will explore this concept.

   (a) Let's begin with a network with three intersections, A, B and C. Define the flows t_1 as the rate of cars (cars/hour) on the road between B and A, t_2 as the rate on the road between C and B and t_3 as the rate on the road between C and A.

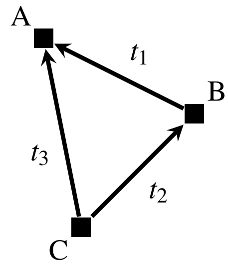


Figure 1: A simple road network.

(Note: The directions of the arrows in the figure are only the way that we define the flow by convention. If there were 100 cars per hour traveling from A to C, then $t_3 = -100$.)

We assume the "flow conservation" constraints: the total number of cars per hour flowing into each intersection is zero. For example at intersection B, we have the constraint $t_2 - t_1 = 0$. The full set of constraints (one per intersection) is:

$$\begin{cases} t_1 + t_3 = 0 \\ t_2 - t_1 = 0 \\ -t_3 - t_2 = 0 \end{cases}$$

As mentioned earlier, we can place sensors on a road to measure the flow through it, but we have a limited budget, and we would like to determine all of the flows with the smallest possible number of sensors.

Suppose for the network above we have one sensor reading, $t_1 = 10$. Can we figure out the flows along the other roads? (That is, the values of t_2 and t_3).

(b) Now suppose we have a larger network, as shown in Figure 2.

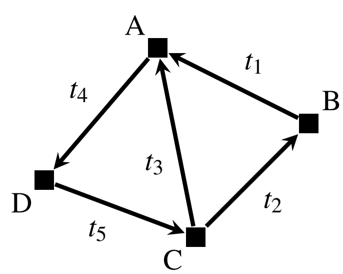


Figure 2: A larger road network.

We would again like to determine the traffic flows on all roads, using measurements from some sensors. A Berkeley student claims that we need two sensors placed on the roads AD and BA. A Stanford student claims that we need two sensors placed on the roads CB and BA. Is it possible to determine all traffic flows with the Berkeley student's suggestion? How about the Stanford student's suggestion?

(c) We would like a more general way of determining the possible traffic flows in a network. Suppose we write the traffic flow on all roads as a vector $\vec{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix}$. As a first step, let us try to write all the flow conservation constraints (one per intersection) as a matrix equation.

Find a 4×5 matrix $\mathbf{B}$ such that the equation $\mathbf{B}\vec{t} = \vec{0}$

$$\begin{bmatrix} & & \\ & \mathbf{B} & \\ & & \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

represents the flow conservation constraints for the network in Figure 2.

Hint: Each row is the constraint of an intersection. You can construct $\mathbf{B}$ using only $0, 1$, and -1 entries. This matrix is called the **incidence matrix.** What constraint does each column of $\mathbf{B}$ represent?

(d) Again, suppose we write the traffic flow on all roads as a vector $\vec{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix}$. Then, determine the subspace of traffic flows for the network of Figure 2. Specifically, express this space as the span of two linearly independent vectors.

Hint: Use the claim of the correct student in part (b).

(e) Notice that the set of all vectors $\vec{t}$ that satisfy $\mathbf{B}\vec{t} = \vec{0}$ is exactly the null space of the matrix $\mathbf{B}$. That is, we can find all valid traffic flows by computing the null space of $\mathbf{B}$. Use Gaussian elimination to determine the dimension of the null space of $\mathbf{B}$ and compute a basis for the null space. Does this match your answer to part (d)? Can you interpret the dimension of the null space of $\mathbf{B}$ for the road networks of Figure 1 and Figure 2?

(f) Now let us analyze more general road networks. Say there is a road network graph G, with incidence matrix $\mathbf{B}_G$. If $\mathbf{B}_G$ has a k-dimensional null space, does this mean measuring the flows along *any* k roads is always sufficient to recover the exact flows? Prove or give a counterexample.

Hint: Consider the Stanford student.

(g) Let G be a network of n roads with the incidence matrix $\mathbf{B}_G$, which has a k-dimensional null space. We would like to characterize exactly when measuring the flows along a set of k roads is sufficient to recover the exact flow along all roads. To do this, it will help to generalize the problem and consider measuring *linear combinations* of flows. Let t_i be the flow on one road. We measure some linear combination of t_i's or $m_0 \cdot t_0 + m_1 \cdot t_1 + \cdots + m_n \cdot t_n$. Now we measure many of these linear combinations, which we will represent using matrix vector multiplication. Then, making k measurements is equivalent to observing the vector $\mathbf{M}\vec{t}$ for some $k \times n$ "measurement matrix" $\mathbf{M}$.

For example, for the network of Figure 2, the measurement matrix corresponding to measuring t_1 and t_4 (as the Berkeley student suggests) is:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Similarly, the measurement matrix corresponding to measuring t_1 and t_2 (as the Stanford student suggests) is:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

For general networks G and measurements M, give a condition for when the exact traffic flows can be recovered in terms of the null space of $\mathbf{M}$ and the null space of $\mathbf{B}_G$.

Hint: Recovery will fail iff there are two valid flows with the same measurements, that is, there exist distinct $\vec{t}_1$ and $\vec{t}_2$, such that $\mathbf{M}\vec{t}_1 = \mathbf{M}\vec{t}_2$. Can you express this in terms of the null spaces of $\mathbf{M}$ and $\mathbf{B}_G$?

(h) Express the condition of the previous part in a way that can be checked computationally. For example, suppose we are given a huge road network G of all roads in Berkeley, and we want to find if our measurements M are sufficient to recover the flows.

Hint: Consider a matrix $\mathbf{U}$ whose columns form a basis of the null space of $\mathbf{B}_G$. Then $\{\mathbf{U}\vec{x} \mid \vec{x} \in \mathbb{R}^k\}$ is exactly the set of all possible traffic flows, that is, every valid flow $\vec{t}$ can be represented as $\vec{t} = \mathbf{U}\vec{x}$ for some $\vec{x}$. How can we represent measurements on these flows?

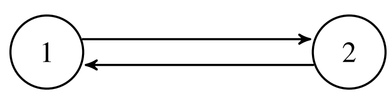
(i) If the incidence matrix $\mathbf{B}_G$ has a k-dimensional null space, does this mean we can always pick a set of k roads such that measuring the flows along these roads is sufficient to recover the exact flows? Prove or give a counterexample.

3. Counting The Paths of a Random Surfer

In class, we discussed the behavior of a random web-surfer who jumps from webpage to webpage. We would like to know how many possible paths there are for a random surfer to get from a page to another page. To do this, we represent the webpages as a graph. If page 1 has a link to page 2, we have a directed edge from page 1 to page 2. This graph can further be represented by what is known as an "adjacency matrix", $\mathbf{A}$, with elements a_{ij}. We define $a_{ji} = 1$ if there is link from page i to page j. Matrix operations on the adjacency matrix make it very easy to compute the number of paths to get from a particular webpage i to webpage j.

This path counting aspect actually is an implicit part of the how the "importance scores" for each webpage are described. Recall that the "importance score" of a website is the steady-state frequency of the fraction of people on that website.

Consider the following graphs.

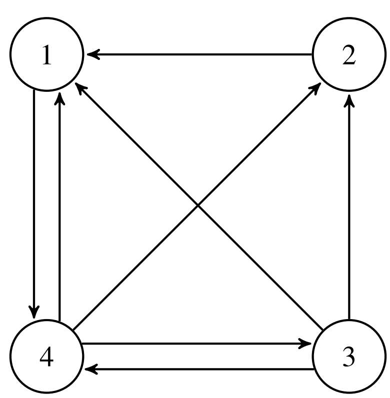


Graph A

(a) Write out the adjacency matrix for graph A.

(b) For graph A: How many one-hop paths are there from webpage 1 to webpage 2? How many two-hop paths are there from webpage 1 to webpage 2? How about three-hop paths?

(c) For graph A: What are the importance scores of the two webpages?

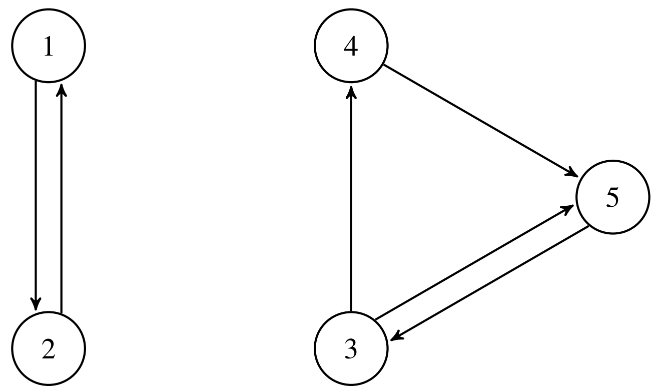


Graph B

(d) Write out the adjacency matrix for graph B.

(e) For graph B: How many two-hop paths are there from webpage 1 to webpage 3? How many three-hop paths are there from webpage 1 to webpage 2?

(f) For graph B: What are the importance scores of the webpages? You may use your IPython notebook for this.



Graph C

(g) Write out the adjacency matrix for graph C.

(h) For graph C: How many paths are there from webpage 1 to webpage 3?

(i) For graph C: What are the importance scores of the webpages? How is graph (c) different from graph (b), and how does this relate to the importance scores and eigenvalues and eigenvectors you found?

4. Homework Process and Study Group

Who else did you work with on this homework? List names and student ID's. (In case of homework party, you can also just describe the group.) How did you work on this homework?