EE263 final exam, December 2022

- This is a 24-hour take-home exam. Please turn it in on Gradescope. Be aware that you *must* turn it in within 24 hours of downloading it. After that, Gradescope will not let you turn it in and we cannot accept it.
- You may use any books, notes, or computer programs. You may not discuss the exam or course material with others, or work in a group.
- The exam should not be discussed at all until 12/16 after everyone has taken it.
- If you have a question, please submit a private question on Ed, or email the staff mailing list. We have tried very hard to make the exam unambiguous and clear, so unless there is a mistake on the exam we're unlikely to say much.
- We expect your solutions to be legible, neat, and clear. Do not hand in your rough notes, and please try to simplify your solutions as much as you can. We will deduct points from solutions that are technically correct, but much more complicated than they need to be.
- Please check your email during the exam, just in case we need to send out a clarification or other announcement.
- Start each question on a new page. Correctly assign pages to problems in gradescope. We may take off points if a submission does not do so.
- Please show the work you do, as it especially helps us give partial credit.
- When a problem involves some computation (say, using Julia), we do not want just the final answers. We want a clear discussion and justification of exactly what you did as well as the final numerical result.
- Because this is an exam, **you must turn in your code**. Include the code in your pdf submission. We reserve the right to deduct points for missing code.
- In the portion of your solutions where you explain the mathematical approach, you *cannot* refer to Julia operators, such as the backslash operator. (You can, of course, refer to inverses of matrices, or any other standard mathematical constructs.)
- Some of the problems require you to download data or other files. These files can be found at the URL

  `http://ee263.stanford.edu/ultimate22.html`
- Good luck!

1. **Foot geometry reconstruction using pressure measurements.** In order to measure the shape of a person's foot, a technique called *plantar pressure cartography* can be used. The idea is that the person stands on a soft foam pad, under which is an array of pressure sensors. The pressure measured at a sensor depends on the compression of the foam directly above the sensor, which is determined by the shape of the foot. We will look at a simplified version of this problem in one dimension.

   The base of the foot is shown as the red line in the figure below. We would like to make multiple measurements of the pressure, in order to determine the profile of the foot, that is h as a function of x. Here the blue line is a reference, since the foot may be held at an angle θ to the ground by the ankle joint.

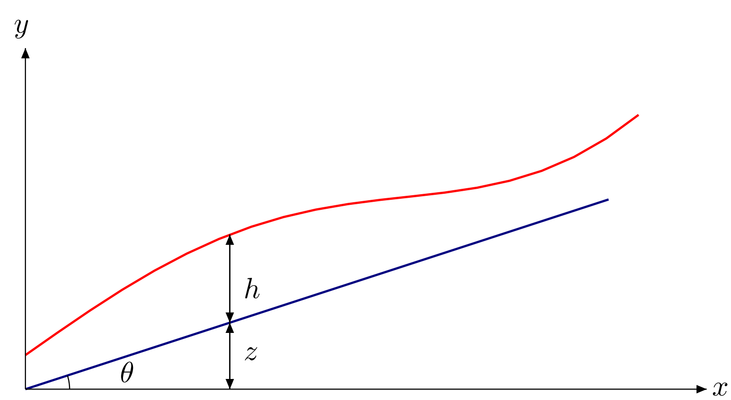
   In order to reduce errors due to noise in the measurements, we ask the subject to stand on the foam pad, and make multiple measurements of the pressure. Since the pressure sensors are independent, these translate into multiple measurements of $y = h + z$. However typically during this process the subject continually adjusts their position, so the repeated measurements correspond to different values for θ, which we cannot measure.

   The figure shows a large value of θ for clarity, but in practice θ is very small, and so we can use a small angle approximation $z \approx \theta x$. We will take measurements at positions $x_1, \dots, x_q$, and take p sets of such measurements. The ith set of measurements is measured at angle θ_i. We use the notation Y_{ij} to denote measurement j at position i, and so we have

$$Y_{ij} = x_i \theta_j + h_i + w_{ij}$$

   where w_{ij} is sensor noise. In practice, each measurement is a function of Y_{ij}, given by the pressure, but for simplicity here we assume that the sensor returns a position measurement.

   We order the sensor locations so that the sequence $x_1, \dots, x_q$ is increasing, and so that $x_1 > 0$.



   a) The $q \times p$ matrix Y is given by its columns

$$Y = \begin{bmatrix} c_1 & c_2 & \dots & c_p \end{bmatrix}$$

Define the pq length vector

$$y = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

If we ignore the noise w, then there exist matrices B and C such that

$$y = C\theta + Bh$$

Give expressions for B and C. You may either give expressions for their entries, or use matrix notation.

b) We are given x, and measure Y. We would like to use least squares squares to find an estimate of h. If h^{est} is an estimate of the foot profile, then if we knew θ, then the error of this estimate is

$$E(h, \theta) = \sum_{i=1}^{q} \sum_{j=1}^{p} (Y_{ij} - x_i\theta_j - h_i)^2$$

We will estimate *both* h and θ together, by finding h^{est} and θ^{est} which mimimize $E(h, \theta)$.

Show that the solution to this problem is not unique.

c) Because this is a least squares problem, we can give a parameterization of all solutions. Specifically, there is a matrix P and vector r such that every optimal h^{est} and θ^{est} is given by

$$\begin{bmatrix} \theta^{\text{est}} \\ h^{\text{est}} \end{bmatrix} = Pv + r$$

for some vector v. Explain how to find P and r given Y and x. Your method must give a matrix P of full column rank.

d) Give a brief interpretation of the nonuniqueness, in terms of the experimental setup.

e) To remove the nonuniqueness, we choose a *representative* solution. Let $\mathcal{H}$ be the set of all optimal solutions

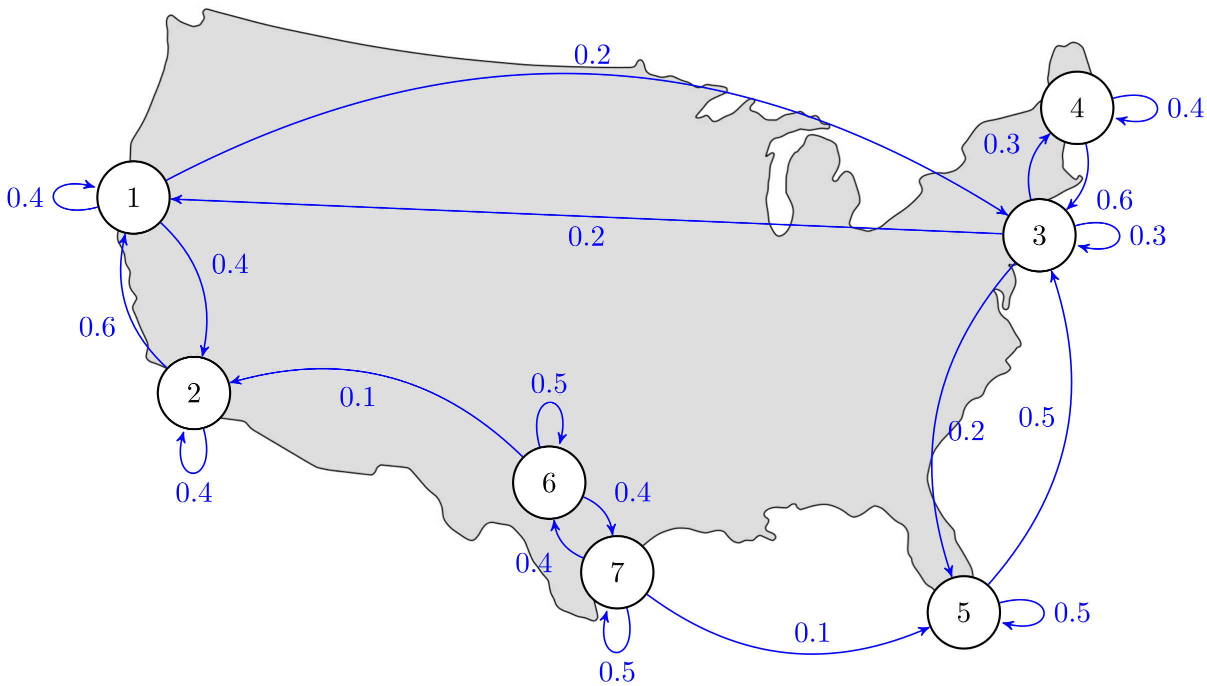
$$\mathcal{H} = \{Pv + r \mid v \in \mathbb{R}^k\}$$

then the representative h^{rep} is the h-component (last q elements) of a vector $\begin{bmatrix} \theta \\ h \end{bmatrix} \in \mathcal{H}$ such that $h_i \geq 0$ for all i and $\sum_i h_i$ is minimized. The P you found above has a particular form which makes the representative easy to find. Give an algorithm to find it, and a brief interpretation.

f) The file `foot.json` contains Y and x. Find the h-component of an optimal least-squares solution h^{est}. Plot h_i^{est} versus i.

g) For the representative solution h^{rep}. Plot h_i^{rep} versus i.

2. The great master of linear algebra. A great master of linear algebra has recently decided to visit the US for the first time! He will be staying in 7 major cities, and since he easily gets bored of staying in one place, after one week of staying in a city, he will either choose his next destination or decide to stay in the same place for another week. The seven cities are San Francisco, Los Angeles, New York, Boston, Miami, Austin, and Houston, which we will enumerate from 1 to 7. Since the master really loves linear dynamical systems, he chooses one of the most famous models for planning his trip, the Markov chain (with which we are familiar from the lecture slides). This means that if he is in the ith city in week t, then he will choose his destination for week $t+1$ based on the *transition probabilities*. These probabilities are shown in the Figure below. If there is no edge from city i to j, it means that the probability of visiting the jth city directly after city i is zero.



Let $z(t) \in \{1, 2, 3, 4, 5, 6, 7\}$ be the location of the master at week t, and define $x(t) \in \mathbb{R}^7$ as the vector of the probability distribution of $z(t)$, *i.e.*

$$x(t) = \begin{bmatrix} \text{Prob}(z(t) = 1) \\ \vdots \\ \text{Prob}(z(t) = 7) \end{bmatrix},$$

where t is any non-negative integer.

a) We know from the lecture slides that $x(t)$ can be expressed as an autonomous linear dynamical system, *i.e.*, there exists $A \in \mathbb{R}^{7\times 7}$ such that

$$x(t+1) = Ax(t).$$

Using the figure, find the matrix A.

b) Next, we would like to study how the probability of finding the master in different cities changes with time. We know that, for some reason, he has chosen Houston as his port of entry to the US, which means that $x(0) = e_7 = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}^\mathsf{T}$. Plot $x_i(t)$ for $i = 1, \cdots, 7$ and $t = 0, 1, \cdots, T$, where $T = 50$. You should generate one figure with all 7 lines on one plot, where the ith plot shows $x_i(t)$ versus t. We call this the *graph of probabilities*, and we will be plotting it in the next sections as well.

   How do the probabilities behave for large t? Which cities have the property that for large enough t, we know for sure that the master will never be there? Use the probabilities in the figure to briefly explain why this happens.

c) Show that $w_1^\mathsf{T} = \mathbf{1}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$ is a left eigenvector of A. What is the corresponding eigenvalue, namely λ_1? Now, let v_1 be the right eigenvector corresponding to λ_1. Explain how v_1 is related to the values observed on the plot in part (**??**).

d) We have been informed that in 50 weeks from now, there will be an international online linear algebra contest, and since we have taken EE263, we are really hopeful that we can win the contest. If at the time of the competition, i.e., at $t = T = 50$ the great master happens to be in San Francisco, we can ask him to be our team leader, and that will guarantee our win! However, from parts (b) and (c), we know that there is no way to make sure he will be in the Bay Area at that time, and there is a non-zero probability that he will not be here. He starts in Houston. Fortunately, we have been able to access the server where he runs his Markov chain (while we don't really know why he runs it on a server since it can be done on any personal computer), and at time t, we can directly modify his location probabilities as we wish. In other words, we get to choose a vector $u(t) \in \mathbb{R}^7$ and perturb the master's Markov chain as

$$x(t+1) = Ax(t) + u(t).$$

   However, the challenge is that it is super-expensive to perturb the Markov chain, and at each time step t, it costs $\|u(t)\|^2 \times \$100K$ to apply the input $u(t)$. So, we would like to make sure that he will visit San Francisco at $t = T$, while we incur the minimum cost, *i.e.*, minimize $\sum_{t=0}^{T} \|u(t)\|^2$. Note that making sure he will be here at $t = T$ means that the probability vector at $t = T$ is given by $x(T) = e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}$. Formulate the problem of minimizing the cost subject to the constraint $x(T) = e_1$ in the form of

$$\begin{array}{ll} \text{minimize} & \|A_1 u - b_1\|^2 \\ \text{subject to} & C_1 u = d_1, \end{array}$$

   where $u = \begin{bmatrix} u(0) \\ \vdots \\ u(T) \end{bmatrix}$. Express A_1, b_1, C_1, and d_1 explicitly. Then solve this problem for the given A with $T = 50$ and plot the graph of probabilities and report the cost. You must observe that the resulting vectors $x(t)$ are not actual probability vectors, *i.e.*, they may have negative values and may also sum to a value other than 1. Confirm that they may have negative values by referring to your graph of probabilities. Also, confirm that they may sum to a value other than one by plotting $s(t) = \sum_{i=1}^{7} x_i(t)$ versus t.

e) In part (**??**), we observed that our method may result in values of $x(t)$ that are not actual probability vectors. If the master notices this, he will get very angry and probably leave the US for good! So, we need to solve this issue. First, we concentrate on the constraint that for each t, we must have $s(t) = \sum_{i=1}^{7} x_i(t) = 1$. Formulate the problem of minimizing the cost subject to the constraint $x(T) = e_1$ and $s(t) = 1$ for all t in the form of

$$\begin{array}{ll} \text{minimize} & \|A_2 u - b_2\|^2 \\ \text{subject to} & C_2 u = d_2. \end{array}$$

Express A_2, b_2, C_2, and d_2 explicitly. Then solve this problem for the given A with $T = 50$, plot the graph of probabilities, and report the cost. Confirm that the resulting vectors $x(t)$ may have negative values by referring to your graph of probabilities. Also, confirm that the issue of their summation has been resolved by plotting $s(t)$ versus t. In addition, plot $u_1(t), \cdots, u_7(t)$ versus t on the same figure. This will be a figure containing 7 plots, each corresponding to one input. We will call this the *graph of inputs*. You must observe that $u_i(t)$ is zero for small t. Intuitively explain why this happens. A brief qualitative description will suffice.

f) Finally, we need to address the last issue, which is to enforce the non-negativity constraint for the probability values, *i.e.*, $x_i(t) \geq 0$. If we knew how to deal with inequality constraints in optimization problems, we could easily solve this issue. However, since that is not covered in this course, we will take an alternative approach. Let J_1 be the objective that we minimized in part (**??**). We now define

$$J_2 = \sum_{t=0}^{T} \sum_{i=1}^{7} (x_i(t) - p)^2,$$

where $p = \frac{1}{7}$, and focus on a multi-objective minimization problem by minimizing $J_1 + \mu J_2$. Explain why adding the term μJ_2 to our objective from part (**??**) can solve the issue of having negative $x_i(t)$'s, and why the choice of $p = \frac{1}{7}$ might be a reasonable choice. Formulate the problem of minimizing $J_1 + \mu J_2$ with the same constraints as the last part in the form of

$$\begin{array}{ll} \text{minimize} & \|A_3 u - b_3\|^2 \\ \text{subject to} & C_3 u = d_3. \end{array}$$

Express A_3, b_3, C_3, and d_3 explicitly. Then solve this problem for the given A with $T = 50$ for a range of values of μ. Plot the trade-off curve of J_1 and J_2. Plot the graph of probabilities for $\mu = 100$ and justify what you observe. Next, by inspection, find the smallest value of μ for which the constraint $x_i(t) \geq 0$ holds for all i and t. Report this value, namely μ^*, along with the corresponding value of J_1. Plot the corresponding graph of probabilities and the graph of inputs. Compare the latter with your results from part (**??**) and justify what you observe.

g) Based on your graph of inputs from the last part, you may notice that we seem to be incurring some unnecessary costs for perturbing the Markov chain for small values of t.

It seems reasonable that we shouldn't have to perturb the Markov chain at early stages if we only need to modify the location probabilities for some distant $t = T$. Use your intuition to modify the way we defined J_2 to address this issue and achieve a lower cost compared to part (**??**) while we maintain the condition $x_i(t) \geq 0$. You don't need to implement your new method, but briefly justify why it will work. Note that the answer to this part is not unique.

3. Pricing interchangeable goods. Bytes cafe is introducing two new seasonal coffee beverages, the peppermint cappuccino and Christmas tree mocha. The final prices of these beverages have not yet been determined. Initial polling reveals the two beverages are partially interchangeable. If the price of a peppermint cappuccino is low enough, customers who prefer the Christmas tree mocha will buy a peppermint cappuccino and vice versa. Let p_1 be the price of a peppermint cappuccino and p_2 be the price of a Christmas tree mocha. The prices of the two beverages will change according to demand, but the prices of the two beverages have a certain degree of "stickiness", which reduces the changes in price due to demand.

As such, we get the following dynamical system for the prices where $k_1, k_2 \geq 0$ are real scalars representing the relative price elasticities of the two goods and $c_1, c_2 \geq 0$ are real scalars representing the stickiness of the prices:

$$\ddot{p}_1 = -c_1\dot{p}_1 - k_1p_1 + k_2p_2 \tag{1}$$

$$\ddot{p}_2 = -c_2\dot{p}_2 + k_1p_1 - k_2p_2. \tag{2}$$

a) Choose an appropriate state vector x for the system described in equations (**??**) and (**??**). Construct a matrix A such that the linear dynamical system $\dot{x} = Ax$ describes all of the information in equations (**??**) and (**??**).

b) We will now use the forward Euler discretization $x(t+h) = (I + hA)x(t)$ of this system. Let $h = 0.1$, $k_1 = 5$, $k_2 = 4$, $c_1 = 1.5$, and $c_2 = 3$. Bytes cafe initially prices the beverages such that $p_1(0) = 3$ and $p_2(0) = 5.5$. Additionally, $\dot{p}_1(0) = 0 = \dot{p}_2(0)$. Plot the components of x from $t = 0$ to $T = 10$.

c) We determine that equilibrium is reached at time t if $\|x(t) - x(t-h)\| \leq 0.01$. Is equilibrium reached in this system, and why? Please formally justify your answer based on some property of A. If so, what is the time t at which it is reached? What are the prices of the beverages at equilibrium?

d) We continue to consider the forward Euler discretization from part (**??**), but we no longer know the values of k_1, k_2, c_1, c_2. The market researchers at Bytes cafe have determined that $c_1 = 1$ and $c_2 = 3$. We have taken some noisy measurements (noise has mean zero) of $p_1, p_2, \dot{p}_1, \dot{p}_2$ from $t = 0$ to $t = 10$ with $h = 0.1$ in `prices.json`. Based on these measurements and the given values of c_1, c_2, describe a method to determine the values of k_1 and k_2 that minimize the expression

$$\sum_{t=0}^{10/h-1} \|(I + hA)x(ht) - x(ht+h)\|_2^2.$$

Hint: Try to express the given expression in terms of the vector $[k_1, k_2]$.

e) Implement the method you described in part (**??**) to determine the the values of k_1 and k_2, and construct the matrix A.

f) Plot $p_1, p_2, \dot{p}_1, \dot{p}_2$ from $t = 0$ to $t = 10$ according to the forward Euler discretization $x(t+h) = (I + hA)x(t)$ with $h = 0.1$ and the matrix A determined in part (**??**). The initial value of p_1 is 4, the initial value of p_2 is 6, and both $\dot{p}_1$ and $\dot{p}_2$ start as 0. Compare this to the plot of the measurements in `prices.json`.

4. Some true or false questions. For each of the statements below, state whether it is true or false. If true, give a brief one-sentence explanation why. If false, give a counterexample.

a) If $A \in \mathbb{R}^{n\times n}$, $A^2 = 0$, and λ is an eigenvalue of A, then $\lambda = 0$

b) If $A \in \mathbb{R}^{n\times n}$ and $A^2 = 0$, then rank of A is at most 2.

c) If $A \in \mathbb{R}^{n\times n}$ and $A^2 = A$, then $A + I$ is invertible.

d) Suppose $n > 2$. If $A \in \mathbb{R}^{n\times n}$ and $A^2 = A$, then A can have n distinct eigenvalues.

e) For a square matrix A, $\operatorname{rank}(A^2) = \operatorname{rank}(A)$ if and only if range $A \cap$ null $A = \{0\}$.

f) If $A, B, P \in \mathbb{R}^{n\times n}$ and $B = P^{-1}AP$, then A and B have the same eigenvalues.

g) If $A, B, P \in \mathbb{R}^{n\times n}$ and $B = P^{-1}AP$, then A and B have the same singular values.

h) If $A \in \mathbb{R}^{n\times n}$ has a repeated eigenvalue, then A is not diagonalizable.

i) If $A \in \mathbb{R}^{n\times n}$ is symmetric and $A^k = 0$ for some $k \in \mathbb{N}$, then $A = 0$.

j) If $A \in \mathbb{R}^{n\times n}$ is symmetric and $A^k = I$ for some $k \in \mathbb{N}$, then $A = I$.

k) If A, B are real symmetric matrices and $A \geq B$, then $\lambda_{\max}(A) \geq \lambda_{\max}(B)$.

l) If A, B are real symmetric matrices and $A, B \geq 0$, then for any $x \in \mathbb{R}^n$, $x^T ABx \geq 0$.

m) If $A \in \mathbb{R}^{n\times n}$, symmetric, and $A > 0$, then $A + A^{-1} \geq 2I$.

n) If A is a real symmetric positive definite matrix, then $\lambda_i(A) = \sigma_i(A)$.

o) If $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$, then $\sigma_{\max}(AB) \leq \max\{\sigma_{\max}(A), \sigma_{\max}(B)\}$.

p) If $A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ is the QR decomposition of A, then A and R_1 have the same singular values.

5. **A layout problem.** In this question, there are m sources with locations $s_j \in \mathbb{R}^2$ for $j = 1, \ldots, m$ and n destinations with locations $d_i \in \mathbb{R}^2$ for $i = 1, \ldots, n$. We have K links, each of which connects a single source to a single destination. The k'th link connects source $s_{\mathrm{src}(k)}$ to destination $d_{\mathrm{dst}(k)}$. The source locations are fixed, but we would like to decide the positions of the destinations.

   You are given an incidence matrix $B \in \mathbb{R}^{K\times m}$ which specifies which source is connected to each link. Specifically

$$B_{kj} = \begin{cases} 1 & \text{if } j = \mathrm{src}(k) \\ 0 & \text{otherwise} \end{cases}$$

   Similarly the matrix $C \in \mathbb{R}^{K\times n}$ specifies which destination is connected to each link, via

$$C_{ki} = \begin{cases} 1 & \text{if } i = \mathrm{dst}(k) \\ 0 & \text{otherwise} \end{cases}$$

   We also define the matrices $S \in \mathbb{R}^{2\times m}$ and $D \in \mathbb{R}^{2\times n}$ by

$$S = \begin{bmatrix} s_1 & s_2 & \ldots & s_m \end{bmatrix} \qquad D = \begin{bmatrix} d_1 & d_2 & \ldots & d_n \end{bmatrix}$$

   a) Let $A = C^{\mathsf{T}} B$. Give an interpretation for A_{ij}.

   b) Each link k has an operational cost, given by

$$c_k = w_k \|s_{\mathrm{src}(k)} - d_{\mathrm{dst}(k)}\|^2$$

      Here w_k is a positive constant associated with link k. The total cost is

$$J_1 = \sum_{k=1}^{K} c_k$$

      Express J_1 in terms of S, D, C, B and w.

   c) As above, with a source and port connected via link k, the cost is the weight w_k multiplied by the corresponding distance squared. To reduce the cost of links, we would like to minimize the total weighted distance between all the sources and ports. The file `layout.json` contains B, C, w, and S. Find the set of destination locations D that minimizes J. Make a plot showing the sources and the optimal destinations. (Use one color for sources and another for destinations.)

   d) In this problem, we are also going to connect each destination i to its neighbors. For convenience, define the wrapping function

$$\mathrm{wrap}(i) = \begin{cases} i+n & \text{if i} < 1 \\ i-n & \text{if i} > \text{n} \\ i & \text{otherwise} \end{cases}$$

      for $-1 \le i \le n+1$. We define the smoothness of the set of destinations by $q_1, q_2, \ldots, q_n \in \mathbb{R}^2$ given by

$$q_i = d_{\mathrm{wrap}(i-1)} - 2d_i + d_{\mathrm{wrap}(i+1)}$$

Let the matrix Q be

$$Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$$

Find the matrix H such that $Q = DH$.

e) We prefer not to have sharp corners on the path through the destinations. So we have a secondary objective

$$J_2 = \sum_{i=1}^{n} \|q_i\|^2$$

We would like to find the destination positions D that minimize

$$J_1 + \mu J_2$$

Explain how you would do this.

f) Find the optimal destination positions when $\mu = 1$. Plot the sources and destinations.

g) Find the optimal destination positions when $\mu = 10$. Plot the sources and destinations.