COT 5615 Math for Intelligent Systems, Fall 2021

Final

12/10/2021

1. (7 points) Use the Cauchy-Schwarz inequality to prove that

$$\frac{1}{n}\sum_{k=1}^{n} x_k \geq \left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k}\right)^{-1}$$

   for all n-vectors $\boldsymbol{x}$ with positive elements x_k. The left-hand side is the arithmetic mean (average) of the numbers x_k; the right-hand side is called the harmonic mean.

2. (10 points) *Circular convolution.* The circular convolution of two n-vectors $\boldsymbol{a}, \boldsymbol{b}$ is the n-vector $\boldsymbol{c}$ defined as

$$c_k = \sum_{\substack{\text{all } i \text{ and } j \text{ with} \\ (i+j) \bmod n = \\ k+1}} a_i b_j, \qquad k = 1, \ldots, n,$$

   where $(i+j) \bmod n$ is the remainder of $i+j$ after integer division by n. Therefore the sum is over all i, j with $i+j = k+1$ or $i+j = n+k+1$. For example, if $n = 4$,

$$\begin{aligned} c_1 &= a_1b_1 + a_4b_2 + a_3b_3 + a_2b_4 \\ c_2 &= a_2b_1 + a_1b_2 + a_4b_3 + a_3b_4 \\ c_3 &= a_3b_1 + a_2b_2 + a_1b_3 + a_4b_4 \\ c_4 &= a_4b_1 + a_3b_2 + a_2b_3 + a_1b_4. \end{aligned}$$

   We use the notation $\boldsymbol{c} = \boldsymbol{a} \circledast \boldsymbol{b}$ for circular convolution, to distinguish it from the standard convolution $\boldsymbol{c} = \boldsymbol{a} * \boldsymbol{b}$ defined in the textbook (p. 136).

   Suppose $\boldsymbol{a}$ is given. Show that $\boldsymbol{a} \circledast \boldsymbol{b}$ is a linear function of $\boldsymbol{b}$, by giving a matrix $\boldsymbol{T_c}(\boldsymbol{a})$ such that $\boldsymbol{a} \circledast \boldsymbol{b} = \boldsymbol{T_c}(\boldsymbol{a})\boldsymbol{b}$ for all $\boldsymbol{b}$.

3. (5 points) Suppose the columns of a matrix $\boldsymbol{A}$ are orthonormal, and we (attempt) to compute its QR factorization $\boldsymbol{A} = \boldsymbol{QR}$. Which of the following must be true?

   (a) The QR factorization will fail.
   (b) $\boldsymbol{R} = \boldsymbol{I}$
   (c) $\boldsymbol{R} = \boldsymbol{A}$
   (d) $\boldsymbol{Q} = \boldsymbol{I}$
   (e) $\boldsymbol{Q} = \boldsymbol{A}$

4. (10 points) The net present value (NPV) of a cash flow $\boldsymbol{c} = (c_1, c_2, c_3)$ with interest rate r is $c_1 + c_2/(1+r) + c_3/(1+r)^2$. With interest rate $r = 0$, the NPV is 1. With interest rate $r = 0.05$, the NPV is 0. With interest rate $r = 0.10$, the NPV is -1.

   Find the cash flow $\boldsymbol{c}$. You are welcome, indeed encouraged, to use a computer to solve any equations you might need to.

5. (10 points) For each of the situations described below, circle the most appropriate response from among the five choices. You can circle only one for each situation. You do not need to justify your responses.

   (a) An intern working for you develops several different models to predict the daily demand for a product. How should you choose which model is the best one, the one to put into production?

   Least squares   Magic   Regularization   Validation k-means

   (b) You have a collection of n documents represented as word histograms and you want to find a few topics from them. Which one is a possible approach?

   Least squares   Magic   Regularization   Validation k-means

   (c) As an intern you develop an auto-regressive model to predict tomorrow's sales volume. It works very well, making predictions that are typically within 5% of the actual sales volume. Your boss, who is not particularly interested in mathematical details, asks how do you know your predictor works. What do you respond?

   Least squares   Magic   Regularization   Validation k-means

   (d) A colleague needs to train a classification model, but she does not want the learned coefficients to take very large values. What do you suggest?

   Least squares   Magic   Regularization   Validation k-means

   (e) You want to solve a linear equation $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{A}$ is a tall matrix, and you find that there does not exist a solution. What would you do next?

   Least squares   Magic   Regularization   Validation k-means

6. (10 points) Formulate the following problem as a set of linear equations $\boldsymbol{Ax} = \boldsymbol{b}$. Find the two cubic polynomials

$$p(t) = x_1 + x_2 t + x_3 t^2 + x_4 t^3, \qquad q(t) = x_5 + x_6 t + x_7 t^2 + x_8 t^3$$

that satisfy the following eight conditions:

- $p(t_1) = y_1, p(t_2) = y_2, p(t_3) = y_3$.
- $\cancel{p(t_5) = y_5}, q(t_5) = y_5, q(t_6) = y_6, q(t_7) = y_7$.
- $p(t_4) = q(t_4), p'(t_4) = q'(t_4)$. This specifies that at $t = t_4$ the polynomials should have the same value and the same derivative.

The variables in the problem are the coefficients $x_1, \ldots, x_8$. The numbers t_i, y_i are given, with $t_1 < t_2 < t_3 < t_4 < t_5 < t_6 < t_7$.

Test the method in Julia (or any other language you like) on the following problem. We take the 7 points t_i equally spaced in the interval $[-0.75, 0.25]$, and

$$y_1 = 0, \quad y_2 = -0.1, \quad y_3 = 0.5, \quad y_5 = 1, \quad y_6 = 0.8, \quad y_7 = 0.5.$$

Calculate the two polynomials $p(t)$ and $q(t)$ (using the command `x=A\b` to solve the equation $\boldsymbol{Ax} = \boldsymbol{b}$), and plot them on the interval $[-0.75, 0.25]$.

7. (10 points) We consider the problem of localization from range measurements in 3-dimensional space. The 3-vector $\boldsymbol{y}$ represents the unknown location. We measure the distances of the location $\boldsymbol{y}$ to five points at known locations $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_5$. The five distance measurements $\rho_1, \ldots, \rho_5$ are exact, except for an unknown systematic error or offset z (for example, due to a clock offset). We therefore have five equations

$$\|\boldsymbol{y} - \boldsymbol{c}_k\| + z = \rho_k, \qquad k = 1, \ldots, 5,$$

with four unknowns y_1, y_2, y_3, z. We assume that the following matrix

$$\begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \boldsymbol{c}_3 & \boldsymbol{c}_4 & \boldsymbol{c}_5 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & \rho_5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

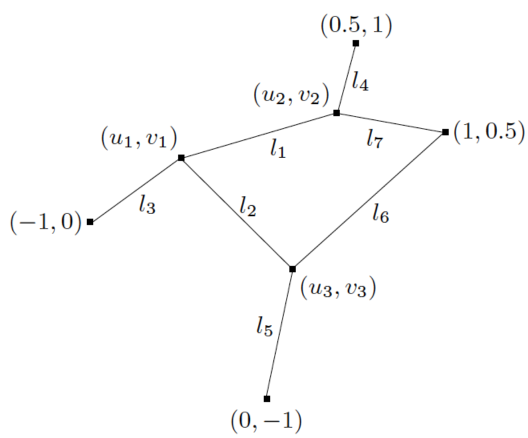
is invertible.

Write a set of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, with a nonsingular matrix $\boldsymbol{A}$, from which the variable $\boldsymbol{x} = (y_1, y_2, y_3, z)$ can be determined. Explain why $\boldsymbol{A}$ is nonsingular.

8. (15 points) We have n points in $\mathbb{R}^2$, and a list of pairs of points that must be connected by links. The positions of some of the n points are fixed; our task is to determine the positions of the remaining points. The objectiveis to place the points so that some measure of the total interconnection length of the links is minimized. As an example application, we can think of the points as locations of plants or warehouses, and the links as the routes over which goods must be shipped. The goal is to find locations that minimize the total transportation cost. In another application, the points represent the position of modules or cells on an integrated circuit, and the links represent wires that connect pairs of cells. Here the goal might be to place the cells in such a way that the total length of wire used to interconnect the cells is minimized.

The problem can be described in terms of a graph with n nodes, representing the n points. With each free node we associate a variable $(u_i, v_i) \in \mathbb{R}^2$, which represent its location or position.

In this problem we will consider the example shown in the figure below. Here we have 3 free points with



coordinates $(u_1, v_1), (u_2, v_2), (u_3, v_3)$. We have 4 fixed points, with coordinates $(-1, 0), (0.5, 1), (0, -1), (1, 0.5)$. There are 7 links, with lengths $l_1, \ldots, l_7$. We are interested in finding the coordinates $(u_1, v_1), (u_2, v_2), (u_3, v_3)$ that minimizes the total square length

$$l_1^2 + l_2^2 + l_3^2 + l_4^2 + l_5^2 + l_6^2 + l_7^2.$$

(a) Formulate this problem as a least squares problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$$

where the 6-vector $\boldsymbol{x}$ contain the six variables $u_1, u_2, u_3, v_1, v_2, v_3$. Give the coefficient matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$.

(b) Show that you can also obtain the optimal coordinates by solving two smaller least squares problems

$$\underset{\boldsymbol{u}}{\text{minimize}} \quad \|\overline{\boldsymbol{A}}\boldsymbol{u} - \overline{\boldsymbol{b}}\|^2, \qquad \underset{\boldsymbol{v}}{\text{minimize}} \quad \|\widetilde{\boldsymbol{A}}\boldsymbol{u} - \widetilde{\boldsymbol{b}}\|^2,$$

where $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$. Give the coefficient matrices $\overline{\boldsymbol{A}}, \widetilde{\boldsymbol{A}}$ and the vectors $\overline{\boldsymbol{b}}, \widetilde{\boldsymbol{b}}$.

(c) Solve the least squares problems derived in part (a) or (b) using Julia (or any language you like).

Hint. We went over a similar exercise 12.12 from the textbook during a lecture.

9. (8 points) We wish to estimate a 4-vector $\boldsymbol{x}$, given noisy measurements of all pairs of its entries, i.e., $y_{ij} \approx x_i + x_j$ for $1 \leq i < j \leq 4$. (The numbers y_{ij} are given.) You will estimate $\boldsymbol{x}$ by minimizing the sum of squares of the residuals $x_i + x_j - y_{ij}$ over all such pairs. Express this objective in the form $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ for an appropriate matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$. You must state the dimensions of $\boldsymbol{A}$ and $\boldsymbol{b}$, and given their entries.

   Hint. There are six pairs of entries.

10. (8 points) Suppose that the $m \times n$ matrix $\boldsymbol{Q}$ has orthonormal columns, i.e., $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}$, and $\boldsymbol{b}$ is an m-vector. Find an expression for the solution $\widehat{\boldsymbol{x}}$ of the following least squares problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}\|^2.$$

   What is the complexity of computing $\widehat{\boldsymbol{x}}$, given $\boldsymbol{Q}$ and $\boldsymbol{b}$, and how does it compare to the complexity of a general least squares problem with an $m \times n$ matrix?

11. (7 points) We have two $n \times n$ circulant matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (c.f. Homework 9 Question 3). Let the eigenvalues of $\boldsymbol{A}$ be $\lambda_1, \ldots, \lambda_n$ and the eigenvalues of $\boldsymbol{B}$ be $\rho_1, \ldots, \rho_n$. What are the eigenvalues of their product $\boldsymbol{A}\boldsymbol{B}$, and why?