COT 5615 Math for Intelligent Systems, Fall 2021

Final Solution

1. Define n-vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ with their elements

$$a_k = \sqrt{\frac{x_k}{n}}, \quad b_k = \sqrt{\frac{1}{x_k n}}, \quad k = 1, \ldots, n.$$

Notice that elements of $\boldsymbol{x}$ are positive so these elements are well-defined. The Cauchy-Schwarz inequality implies

$$(\boldsymbol{a}^\top \boldsymbol{b})^2 \leq \|\boldsymbol{a}\|^2 \|\boldsymbol{b}\|^2.$$

The left-hand side equals

$$(\boldsymbol{a}^\top \boldsymbol{b})^2 = \left(\sum_{k=1}^{n} \sqrt{\frac{x_k}{n}} \sqrt{\frac{1}{x_k n}} \right)^2 = 1.$$

On the right-hand side we have

$$\|\boldsymbol{a}\|^2 = \frac{1}{n} \sum_{k=1}^{n} x_k, \qquad \|\boldsymbol{b}\|^2 = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{x_k}.$$

Therefore

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

Dividing both sides by $\|\boldsymbol{b}\|^2$ gives the arithmetic-harmonic mean inequality

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

2. The matrix $\boldsymbol{T_c}(\boldsymbol{a})$ would be a $n \times n$ circulant matrix with $\boldsymbol{a}$ on its first column

$$\boldsymbol{T_c}(\boldsymbol{a}) = \begin{bmatrix} a_1 & a_n & \cdots & a_3 & a_2 \\ a_2 & a_1 & \cdots & a_4 & a_3 \\ a_3 & a_2 & \cdots & a_5 & a_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_n & a_{n-1} & \cdots & a_2 & a_1 \end{bmatrix}.$$

3. (b) and (e) must be true.

4. The problem description gives the following equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1/1.05 & 1/1.05^2 \\ 1 & 1/1.1 & 1/1.1^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Solving this equation in Julia gives us (approximately)

$$c_1 = -42.0, \qquad c_2 = 66.1, \qquad c_3 = -23.1.$$

5. (a) Validation
   (b) k-means
   (c) Validation
   (d) Regularization
   (e) Least squares

6. The conditions describe a linear equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with

$$\boldsymbol{A} = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 & 0 & 0 & 0 & 0 \\ 1 & t_2 & t_2^2 & t_2^3 & 0 & 0 & 0 & 0 \\ 1 & t_3 & t_3^2 & t_3^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & t_5 & t_5^2 & t_5^3 \\ 0 & 0 & 0 & 0 & 1 & t_6 & t_6^2 & t_6^3 \\ 0 & 0 & 0 & 0 & 1 & t_7 & t_7^2 & t_7^3 \\ 1 & t_4 & t_4^2 & t_4^3 & -1 & -t_4 & -t_4^2 & -t_4^3 \\ 0 & 1 & 2t_4 & 3t_4^2 & 0 & -1 & -2t_4 & -3t_4^2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_5 \\ y_6 \\ y_7 \\ 0 \\ 0 \end{bmatrix}.$$

Plugging in the values in Julia we obtain

```
8-element VectorFloat64:
-0.27698863636363313
-12.887499999999964
-38.65909090909083
-29.290909090909036
0.9196022727272727
-1.2284090909090912
-2.822727272727278
4.090909090909108
```

7. We rewrite the five equations as follows:

$$\begin{aligned} & \|\boldsymbol{y} - \boldsymbol{c}_k\| + z = \rho_k \\ \iff \quad & \|\boldsymbol{y} - \boldsymbol{c}_k\|^2 = (\rho_k - z)^2 \\ \iff \quad & \|\boldsymbol{y}\|^2 + \|\boldsymbol{c}_k\|^2 - 2\boldsymbol{c}_k^\top \boldsymbol{y} = \rho_k^2 + z^2 - 2\rho_k z \\ \iff \quad & \|\boldsymbol{y}\|^2 + z^2 = 2\boldsymbol{c}_k^\top \boldsymbol{y} - 2\rho_k z + \rho_k^2 - \|\boldsymbol{c}_k\|^2. \end{aligned}$$

These are not exactly linear equations with respect to $\boldsymbol{y}$ and z, but we notice that the left-hand sides are the same, so they are equivalent to four linear equations

$$2\boldsymbol{c}_k^\top \boldsymbol{y} - 2\rho_k z + \rho_k^2 - \|\boldsymbol{c}_k\|^2 = 2\boldsymbol{c}_5^\top \boldsymbol{y} - 2\rho_5 z + \rho_5^2 - \|\boldsymbol{c}_5\|^2, \quad k = 1, 2, 3, 4.$$

This can be written as a set of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with

$$\boldsymbol{A} = \begin{bmatrix} (\boldsymbol{c}_1 - \boldsymbol{c}_5)^\top & \rho_1 - \rho_5 \\ (\boldsymbol{c}_2 - \boldsymbol{c}_5)^\top & \rho_2 - \rho_5 \\ (\boldsymbol{c}_3 - \boldsymbol{c}_5)^\top & \rho_3 - \rho_5 \\ (\boldsymbol{c}_4 - \boldsymbol{c}_5)^\top & \rho_4 - \rho_5 \end{bmatrix}, \qquad \boldsymbol{b} = \frac{1}{2} \begin{bmatrix} \rho_5^2 - \|\boldsymbol{c}_5\|^2 - \rho_1^2 + \|\boldsymbol{c}_1\|^2 \\ \rho_5^2 - \|\boldsymbol{c}_5\|^2 - \rho_2^2 + \|\boldsymbol{c}_2\|^2 \\ \rho_5^2 - \|\boldsymbol{c}_5\|^2 - \rho_3^2 + \|\boldsymbol{c}_3\|^2 \\ \rho_5^2 - \|\boldsymbol{c}_5\|^2 - \rho_4^2 + \|\boldsymbol{c}_4\|^2 \end{bmatrix}.$$

Since 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, matrix $\boldsymbol{A}$ is also invertible. (see solution of Q1, Midterm 2 for details)

8. Denote $(u_1, v_1), (u_2, v_2), (u_3, v_3)$ as node 1,2,3 and $(-1, 0), (0.5, 1), (0, -1), (1, 0.5)$ as node 4,5,6,7, and using the edge numbering $l_1, \ldots, l_7$, we construct the graph incidence matrix

$$\boldsymbol{B} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix},$$

where $\boldsymbol{B}_1$ is the first 3 rows of $\boldsymbol{B}$ and $\boldsymbol{B}_2$ is the last 4 rows of $\boldsymbol{B}$. Define 7-vector $\boldsymbol{p} = (u_1, u_2, u_3, -1, 0.5, 0, 1)$ that collects all the first coordinates of the 7 points, and 7-vector $\boldsymbol{q} = (v_1, v_2, v_3, 0, 1, -1, 0.5)$ collecting all the second coordinates of them, we have

$$\sum_{k=1}^{7} l_k^2 = \|\boldsymbol{B}^\top \boldsymbol{p}\|^2 + \|\boldsymbol{B}^\top \boldsymbol{q}\|^2 = \left\| \begin{bmatrix} \boldsymbol{B}^\top & 0 \\ 0 & \boldsymbol{B}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix} \right\|^2$$

(a) We note that

$$\begin{bmatrix} \boldsymbol{B}^\top & 0 \\ 0 & \boldsymbol{B}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_1^\top & 0 \\ 0 & \boldsymbol{B}_1^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} + \begin{bmatrix} \boldsymbol{B}_2^\top & 0 \\ 0 & \boldsymbol{B}_2^\top \end{bmatrix} \begin{bmatrix} -1 \\ 0.5 \\ 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ 0.5 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} 1 \\ 1 \\ 1 \\ 1 \end{matrix}}\right\} \bar{\boldsymbol{p}} \\ \left.\vphantom{\begin{matrix} 1 \\ 1 \\ 1 \\ 1 \end{matrix}}\right\} \tilde{\boldsymbol{q}} \end{matrix}$$

So for the least squares problem with variable $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{v})$, the matrix $\boldsymbol{A}$ and $\boldsymbol{b}$ aare

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}_1^\top & 0 \\ 0 & \boldsymbol{B}_1^\top \end{bmatrix}, \qquad \boldsymbol{b} = -\begin{bmatrix} \boldsymbol{B}_2^\top & 0 \\ 0 & \boldsymbol{B}_2^\top \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{p}} \\ \tilde{\boldsymbol{q}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_2^\top \bar{\boldsymbol{p}} \\ \boldsymbol{B}_2^\top \tilde{\boldsymbol{q}} \end{bmatrix}$$

(b)

$$\overline{\boldsymbol{A}} = \widetilde{\boldsymbol{A}} = \boldsymbol{B}_1^\top, \qquad \bar{\boldsymbol{b}} = -\boldsymbol{B}_2^\top \bar{\boldsymbol{p}}, \quad \tilde{\boldsymbol{b}} = -\boldsymbol{B}_2^\top \tilde{\boldsymbol{q}}.$$

(c) Solving (b) gives

```
u =
-0.0714285714285715
0.476190476190476
0.30952380952380953
v =
0.14285714285714282
0.5476190476190474
-0.11904761904761903
```

9. Matrix $\boldsymbol{A}$ is a 6×4 matrix and $\boldsymbol{b}$ is a 6-vector

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} y_{12} \\ y_{13} \\ y_{14} \\ y_{23} \\ y_{24} \\ y_{34} \end{bmatrix}$$

10. Since $\boldsymbol{Q}$ has orthonormal columns,

$$\widehat{\boldsymbol{x}} = \boldsymbol{Q}^{\dagger}\boldsymbol{b} = (\boldsymbol{Q}^{\top}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\top}\boldsymbol{b} = \boldsymbol{Q}^{\top}\boldsymbol{b}.$$

The complexity is the cost of a matrix-vector multiplication, so $2mn$ flops. This is much cheaper than solving a general least squares problem with $2mn^2$ flops.

11. As explained in Question 3 of Homework 9, rows of the DFT matrix are eigenvectors of a circulant matrix. The question does not mention how the eigenvalues are ordered. If we have that

$$\lambda_k = a_0 + a_1\omega^{k+1} + \cdots + a_{n-1}\omega^{(k+1)(n-1)}$$
$$\rho_k = b_0 + b_1\omega^{k+1} + \cdots + b_{n-1}\omega^{(k+1)(n-1)},$$

where $(a_0, a_1, \ldots, a_{n-1})$ and $(b_0, b_1, \ldots, b_{n-1})$ are the first columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The eigen-decomposition of $\boldsymbol{A}$ is $\boldsymbol{F}^*\boldsymbol{\Lambda}\boldsymbol{F}$ and that of $\boldsymbol{B}$ is $\boldsymbol{F}^*\boldsymbol{P}\boldsymbol{F}$, so $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{F}^*\boldsymbol{\Lambda}\boldsymbol{P}\boldsymbol{F}$, which implies that

- $\boldsymbol{A}\boldsymbol{B}$ can be diagonalized by the DFT matrix, so it is also circulant;
- its eigenvalues are $\lambda_1\rho_1, \ldots, \lambda_n\rho_n$.