

You are *not* allowed to work with your peers. Show all work and derivations to receive credit.
Write clearly and legibly for your own benefit.

Name : Varun Venkatesh

SID : 1939023

Attention!!: You will have to upload the final Python notebook to Canvas in the midterm exam assignment to get credit in addition to attaching a print out pdf to the exam you submit!

Problem 1 (Linear Independence, Matrices) [6pts].

Consider three vectors $u, v, w \in \mathbb{R}^3$ that are linearly independent, and let $a, b, c \in \mathbb{R}^3$ be defined as

$$a = u + v, \quad b = v + w, \quad c = u + w, \quad d = u - v.$$

- Suppose vectors u and v have unit length. What is the angle between the vectors a and d ?
- Again suppose vectors u and v have unit length. Give an expression for a left-inverse of the matrix $A = [a \ d]$ (your expression can depend on a and d).
- Show that nullspace of the matrix $B = [a \ b \ c]$ contains only the zero vector $\mathbf{0} \in \mathbb{R}^3$.

Solution.

$$a) \cos(\text{angle bw } a, d) = \frac{a \cdot d}{\|a\| \|d\|} = \frac{(u+v) \cdot (u-v)}{\|u+v\| \|u-v\|} = \frac{u \cdot u - u \cdot v + v \cdot u - v \cdot v}{\|u+v\| \|u-v\|} =$$

u, v are $L\perp$, so $\text{mag}(a)$ and $(d) = \sqrt{1+1} = \sqrt{2}$, so $\angle(a, d) = \frac{u^2 - v^2}{2}$

$$b) A = \begin{bmatrix} u & v \\ u & w \\ u & w \end{bmatrix}, \text{ dim} = 3 \times 2, \text{ so needs to produce } 2 \times 2 \text{ I matrix. Then}$$

for $XA = I_2$, need $X = \text{dim} = 2 \times 3$. Let $X = \begin{bmatrix} i^T \\ j^T \end{bmatrix}$ where $i, j \in \mathbb{R}^3$,

then $XA = \begin{bmatrix} i^T u_1 + i^T v_1 + i^T w_1 & i^T u_2 + i^T v_2 + i^T w_2 \\ j^T u_1 + j^T v_1 + j^T w_1 & j^T u_2 + j^T v_2 + j^T w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$c) B = [a \ b \ c] = [u+v \ v+w \ u+w], \text{ since } u, v, w \text{ are } L\perp, \text{ then}$$

$u+v, v+w, u+w$ are not linear combinations of each other meaning

for matrix B $\text{dim} = 3 \times 3$, $\text{rank}(B) = 3$ because the columns are $L\perp$,

and by rank-nullity, $\text{nullity} = 0$, so null space only contains $\mathbf{0} \in \mathbb{R}^3$

Problem 2 (Matrix inverse & properties) [6pts]. Consider the following $n \times n$ matrix,

$$S = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

What does the matrix S do when applied to a vector? Find the inverse of S : Solve $SX = I$ for the unknown matrix X by writing the linear equations each column of X should satisfy. What is the interpretation of S^{-1} ?

For a vector $w \in \mathbb{R}^n$, $w = [w_1, w_2, \dots, w_n]$, $Sw = z$, $z \in \mathbb{R}^n$, where $z = [w_1, w_1 + w_2, \dots, (w_1 + w_2 + \dots + w_n)]$. Each row i is equal to the sum of w_1, \dots, w_i from $i = \{1, n\}$. For some matrix X ,

$$X \in \mathbb{R}^{n \times n}, \quad SX = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{11} + X_{21} & X_{12} + X_{22} & \dots & X_{1n} + X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{11} + \dots + X_{n1} & X_{12} + \dots + X_{n2} & \dots & X_{1n} + \dots + X_{nn} \end{bmatrix}, \quad \text{so } X = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & -1 & 1 \end{bmatrix}.$$

This matrix is equal to S^{-1} , and conceptually represents, when multiplied into another matrix, Y , as $S^{-1}Y$, it will subtract each element by the element to its left, or if as YS^{-1} , each element by the element below it.

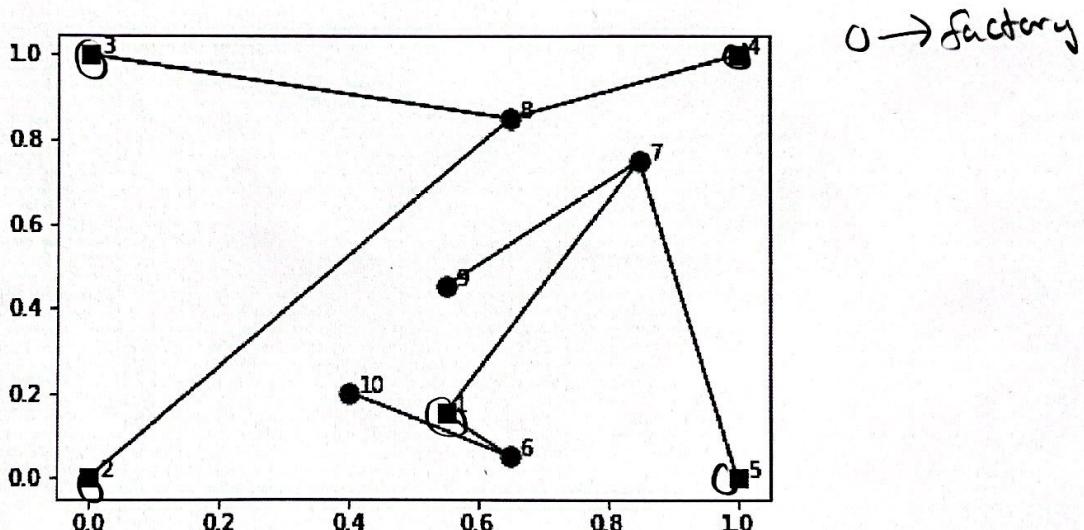
Problem 3 (Least Squares Placement) [18pts]. This problem has parts a–e. We have provided space below each problem for the solution.

The vectors p_1, \dots, p_N , each in \mathbb{R}^2 represent the locations of N objects. There are two types of objects: factories and warehouses. The first K objects are factories, whose locations are fixed and given. Our goal in the placement problem to choose the locations of the last $N - K$ objects i.e the warehouses.

Our choice of the locations is guided by an undirected graph; an edge between two objects means we would like them to be close to each other. In least squares placement, we choose the locations p_{K+1}, \dots, p_N of warehouses so as to minimize the sum of the squares of the distances between objects connected by an edge, where the L edges of the graph are given by the set E . For a specific location of factories p_1, \dots, p_K , we can frame our task as solving the following optimization:

$$g(p_1, \dots, p_K) = \min_{p_{K+1}, \dots, p_N} \sum_{(i,j) \in E} \|p_i - p_j\|^2$$

For illustration, see the figure below. The factories are denoted by red squares, and warehouses by blue circles; we want to move the blue circles so that the objective is minimized.



- a. In each of the simplified cases below, state what the optimal objective value would be. Further, what assignment to $\{p_{K+1}, \dots, p_N\}$ would achieve this optimal objective? If there are multiple optimal assignments, state any one.
- i. The factories are fixed at $p_{N-K+1} = \dots = p_{N-1} = p_N = \mathbf{0}$.
 - ii. There is $K = 1$ factory. Write your answer in terms of the fixed location p_1 .
 - iii. All edges are from one warehouse to another. In other words, for any $(i, j) \in E$, both (p_i, p_j) are warehouses. Write your answer in terms of E .

Solution part a.

- a) i. The minimization occurs if we place p_1, \dots, p_{K-k} all at 0, so $p_1 = p_2 = \dots = p_{N-k+1} = \dots = p_N = 0$, so $g(p_1, \dots, p_N) = 0$
- ii. minimized when p_1, \dots, p_{N-k}, p_N , so $g(p_1, \dots, p_N) = 0$, type in gain
- iii. since we have the freedom to place any warehouse anywhere, $g(p_1, \dots, p_N) = 0$, with every warehouse on E in the same spot, so $\sum_{i \in E} p_i = 0$

Exercise 2: the sum of the squares of the potential differences is expressed as

Exercise 2: the sum of the squared distances between the N factories can be expressed as

Exercise 2: if $a = ((p_1), \dots, (p_N))$ and $b = ((q_1), \dots, (q_N))$ are N -vectors consisting of

Exercise 2: squared coordinates of the factories, respectively.

Exercise 2: part b.

- b. Now, we move towards solving the problem in the more general case.

Let \mathcal{D} be the Dirichlet energy of the graph, which is defined as follows (more details are given in §7.3 of VMLS, and specifically page 135; you may want to take a look):

Consider a graph $\mathcal{G} = (\{1, \dots, N\}, E)$ composed of a set of edges E and N nodes $\{1, \dots, N\}$ such that nodes $i, j \in \{1, \dots, N\}$ are connected if and only if there exists an edge $(i, j) \in E$. Let B be the incidence matrix of the graph, and let $w \in \mathbb{R}^N$ be the potential for the graph—i.e., w_i is the value of some quantity at node $i \in \{1, \dots, N\}$. The Dirichlet energy is $\mathcal{D}(w) = \|B^T w\|^2$ and can be equivalently expressed as

$$\mathcal{D}(v) = \sum_{(i,j) \in E} (w_i - w_j)^2$$

Solution

1 if edge leaves
 -1 if edge enters
 0 else

which is the sum of the squares of the potential differences of w across all edges in the graph.

Show that the sum of the squared distances between the N factories can be expressed as $\mathcal{D}(u) + \mathcal{D}(v)$, where $u = ((p_1)_1, \dots, (p_N)_1)$ and $v = ((p_1)_2, \dots, (p_N)_2)$ are N -vectors containing the first and second coordinates of the factories, respectively.

Solution part b.

$$\begin{aligned}
 \text{For } p_1, \dots, p_N, \quad \mathcal{D}(w) = \|B^T w\|^2 = \sum_{i,j \in E} (w_i - w_j)^2 = \|w_i - w_j\|^2, \text{ translates to} \\
 \|p_{1i} - p_{1j}\|^2 + \|p_{2i} - p_{2j}\|^2 + \dots + \|p_{Ni} - p_{Nj}\|^2 = \left\| \begin{bmatrix} v_{1i} - v_{1j} \\ v_{2i} - v_{2j} \\ \vdots \\ v_{Ni} - v_{Nj} \end{bmatrix} \right\|^2 + \dots + \left\| \begin{bmatrix} v_{1i} - v_{1j} \\ v_{2i} - v_{2j} \\ \vdots \\ v_{Ni} - v_{Nj} \end{bmatrix} \right\|^2 = \\
 \left(\sqrt{(v_{1i} - v_{1j})^2 + (v_{2i} - v_{2j})^2 + \dots + (v_{Ni} - v_{Nj})^2} \right)^2 = \\
 (v_{1i} - v_{1j})^2 + (v_{2i} - v_{2j})^2 + \dots + (v_{Ni} - v_{Nj})^2 + (v_{Ni} - v_{Nj})^2 = \mathcal{D}(u) + \mathcal{D}(v)
 \end{aligned}$$

- c. Express the least squares placement problem as a least squares problem, with variable $x = (u_{1:(N-K)}, v_{1:(N-K)})$. In other words, express the objective above (the sum of squares of the distances across edges) as $\|Ax - b\|^2$, for an appropriate $m \times n$ matrix A and m -vector b . You will find that $m = 2L$.

Hint: You can use the fact that $D(y) = \|B^T y\|^2$, where B is the incidence matrix of the graph.

Solution part c.

Least squares = $D(u) + D(v) = \|B^T u\|^2 + \|B^T v\|^2$ for the incidence matrix B , which has dimension $N \times L$. Then, the first $N-K$ rows of B correlate to the factories, B_f , and the last $(N-L)$ rows of B correlate to warehouses, B_w . We can also split u, v (so that they also split into $[1 \dots K]$, $[N-K \dots N]$). Then u_f, u_w, v_f, v_w are created so we can rewrite the least squares as:

$$\|B^T u\|^2 + \|B^T v\|^2 = \|B_f^T u_f + B_w^T u_w\|^2 + \|B_f^T v_f + B_w^T v_w\|^2.$$

To get this into form $\|Ax - b\|^2$, we need to minimize 2 separate least squares equations, where the two equations are:

$$A = - \begin{bmatrix} B_w^T & 0 \\ 0 & B_w^T \end{bmatrix} \text{ in both, and } b = - \begin{bmatrix} B_f^T u_f \\ B_f^T v_f \end{bmatrix}, \quad b =$$

$$A = - \begin{bmatrix} B_w^T & 0 \\ 0 & B_w^T \end{bmatrix} \text{ and } b = - \begin{bmatrix} B_f^T v_f \\ B_f^T u_f \end{bmatrix}$$

- d. Solve the least squares placement problem for the specific problem with $N = 10, K = 5, L = 13$, fixed locations

$$p_1 = (0.55, 0.15), \quad p_2 = (0, 0), \quad p_3 = (0, 1), \quad p_4 = (1, 1), \quad p_5 = (1, 0),$$

The edges are:

$$(1, 6), \quad (2, 6), \quad (5, 6), \quad (1, 7), \quad (4, 7), \quad (2, 8), \\ (3, 8), \quad (3, 9), \quad (5, 9), \quad (5, 10), \quad (7, 9), \quad (6, 10), \quad (7, 10).$$

Using the provided Python notebook, plot the locations, showing the graph edges as lines connecting the locations. Specifically, solve **Python-P3d** in the provided notebook and print the pdf and attach it at the end of your exam. You will have to upload the final notebook to Canvas to get credit in addition to attaching a print out pdf to the exam you submit!

Problem 4 (Recursive Least Squares) [12pts]. Suppose we have data $(x^{(1)}, \dots, x^{(m)})$ and $(y^{(1)}, \dots, y^{(m)})$ and we believe that $y^{(i)} \approx (x^{(i)})^\top \theta$ —that is, $y^{(i)}$ is approximately a linear function of $x^{(i)}$. The $x^{(i)} \in \mathbb{R}^n$ are n -dimensional vectors and $y^{(i)} \in \mathbb{R}$ are scalars. The least squares method seeks to minimize

$$J(\theta) = \sum_{k=1}^m (y^{(k)} - (x^{(k)})^\top \theta)^2$$

We saw in lecture that

$$\hat{\theta}_m := \left(\sum_{k=1}^m x^{(k)} (x^{(k)})^\top \right)^{-1} \sum_{k=1}^m x^{(k)} y^{(k)}$$

minimizes $J(\theta)$. There are three parts to the problem and on the next page put your solutions to parts a and b. For part c, print your pdf and attach it at the end of your exam. You will also need to provide the ipynb of your solutions.

- a. Define $X_m^\top = [x^{(1)} \ x^{(2)} \ \dots \ x^{(m)}]$ and $Y_m = [y^{(1)} \ y^{(2)} \ \dots \ y^{(m)}]^\top$. Show that

$$\hat{\theta}_m := (X_m^\top X_m)^{-1} X_m^\top Y_m$$

minimizes $\|Y_m - X_m \theta\|^2$ by taking the derivative of $\|Y_m - X_m \theta\|^2$ and setting it to zero.

- b. Now suppose one more data pair $(x^{(m+1)}, y^{(m+1)})$ becomes available and define

$$X_{m+1}^\top = [X_m^\top \ x^{(m+1)}] \quad \text{and} \quad Y_m = [Y_m^\top \ y^{(m+1)}]^\top$$

Hence, the new least squares solution is

$$\hat{\theta}_{m+1} := (X_{m+1}^\top X_{m+1})^{-1} X_{m+1}^\top Y_{m+1} = R_{m+1}^{-1} \sum_{k=1}^{m+1} x^{(k)} y^{(k)} \quad (1)$$

where

$$R_{m+1} := \sum_{k=1}^{m+1} x^{(k)} (x^{(k)})^\top.$$

Show that the least squares solution can be obtained by the following recursive procedure:

$$\hat{\theta}_{m+1} = \hat{\theta}_m + R_{m+1}^{-1} x^{(m+1)} (y^{(m+1)} - (x^{(m+1)})^\top \hat{\theta}_m) \quad (2)$$

$$R_{m+1} = R_m + x^{(m+1)} (x^{(m+1)})^\top \quad (3)$$

That is, show that these equations hold given the definition of R_m and the expression for the least squares solution.

Hint: First, verify (3) holds given the definition of R_{m+1} . Then, to verify (2), take the expression in (1) for $\hat{\theta}_{m+1}$, and expand the sum $\sum_{k=1}^{m+1} x^{(k)} y^{(k)} = x^{(m+1)} y^{(m+1)} + \sum_{k=1}^m x^{(k)} y^{(k)}$. Now, use (3) and the expression you have for $\hat{\theta}_m$ to show (2) holds.

- c. Implement recursive least squares. In the provided Python notebook, solve **Python-P4**. You will have to upload the final notebook to Canvas to get credit in addition to attaching a print out pdf to the exam you submit!

Solution.

a. Provide your solution to part a. here.

$$\text{Derivative of } \|Y_m - X_m \theta\|^2 = \text{derivative of } Y_m^T Y_m - 2(Y_m^T X_m) \theta + \theta^T X_m^T X_m \theta$$

$$= -2X_m^T Y_m + (X_m^T X_m + (X_m^T X_m)^T) \theta. \text{ Let } \Theta = (X_m^T X_m)^{-1} X_m^T Y_m =$$

$$X_m^{-1} (X_m^{-1})^T X_m^T Y_m = X_m^{-1} Y_m, \text{ so the above equation becomes}$$

$$-2X_m^T Y_m + \underbrace{X_m^T X_m}_{I} \underbrace{X_m^{-1} Y_m}_{I} + X_m^T \underbrace{X_m X_m^{-1}}_I Y_m = -2X_m^T Y_m + X_m Y_m + X_m Y_m = 0,$$

and when $\nabla \|Y_m - X_m \theta\|^2 = 0$, it is minimized

b. Provide your solution to part b. here.

we know $R_{m+1} = \sum_{k=1}^{m+1} x^k (x^k)^T$, which is equal to $\sum_{k=1}^{m+1} x^k (x^k)^T + x^{m+1} (x^{m+1})^T$

$$= R_m + x^{m+1} (x^{m+1})^T.$$

we know $\hat{\theta}_{m+1} = (x_{m+1}^T x_{m+1})^{-1} x_{m+1}^T y_{m+1} = R_{m+1}^{-1} \sum_{k=1}^{m+1} x^k y^k$.

If $\hat{\theta}_{m+1} = \hat{\theta}_m + R_{m+1}^{-1} x^{m+1} (y^{m+1} - (x^{m+1})^T \hat{\theta}_m)$, then this equals

$$\hat{\theta}_m + R_{m+1}^{-1} (x^{m+1} y^{m+1} - x^{m+1} (x^{m+1})^T \hat{\theta}_m) = R_m^{-1} \sum_{k=1}^m x^k y^k + R_{m+1}^{-1} x^{m+1} y^{m+1}$$

$$- R_{m+1}^{-1} x^{m+1} (x^{m+1})^T \hat{\theta}_m = R_{m+1}^{-1} x^k y^k, \text{ since } y^{k+1} = (x^k)^T \hat{\theta}.$$

Therefore, the two are equivalent.

Problem 5 (Least Norm Solution) [6pts]. In class we saw how to obtain the least squares approximate solution to $Ax = b$ corresponding to an over-determined set of equations—that is, where $A \in \mathbb{R}^{m \times n}$ with $m > n$ (“tall matrix”). In this case, there is rarely an exact solution to $Ax = b$, and instead we find an approximate solution by minimizing $\|Ax - b\|^2$.

If on the other hand A is a “wide matrix”, meaning $n > m$, then the set of equations is under-determined. This means there are more unknowns than equations and hence, potentially many solutions. In this case we select, amongst the solutions, the one with the smallest norm—i.e., the least norm solution. That is, we seek \hat{x} such that $A\hat{x} = b$ and $\|\hat{x}\| \leq \|x\|$ for all other x such that $Ax = b$.

Consider an under-determined system of equations $Ax = b$ where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ are given with $n > m$. Suppose that A is full rank and $b \in \text{range}(A)$ —this ensures there is at least one solution. Show that $\hat{x} = A^\top(AA^\top)^{-1}b$ is the least norm solution to $Ax = b$. In particular, letting $S = \{x \in \mathbb{R}^n \mid Ax = b\}$ be the set of solutions to $Ax = b$, you need to show that

$$\|\hat{x}\| \leq \|x\| \quad \forall x \in S.$$

$$\text{Let } x = \hat{x} + u$$

Solution.

$$\begin{aligned} \|Ax - b\|^2 &= (Ax - b)^\top (Ax - b) = (A\hat{x} + Au - b)^\top (A\hat{x} + Au - b) = (Au + (A\hat{x} - b))^\top (Au + (A\hat{x} - b)) \\ (Au + (A\hat{x} - b))^\top (Au + (A\hat{x} - b)) &= (Au)^\top Au + (A\hat{x} - b)^\top (A\hat{x} - b) + 2(Au)^\top (A\hat{x} - b) = \\ \|Au\|^2 + \|A\hat{x} - b\|^2 &+ 2u^\top (A^\top A\hat{x} - A^\top b), \text{ which if we let } u=0, \\ \text{is minimized when } x = \hat{x}, \text{ let } \hat{x} &= A^\top(AA^\top)^{-1}b = A^\top A^{-1} A^\top b = \bar{A}^\top b, \\ \text{so } A\hat{x} &= A\bar{A}^\top b = b, \text{ thereby fulfilling the equation.} \end{aligned}$$