HW3

xh2433

03/08/2024

1 Question 1

1.1 (i)

Let $\delta \in \mathbb{R}^+$, $A \in \mathbb{R}^{n\times n}$, A has eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m$. $a_1, a_2, ..., a_m$ represents the number of times each eigenvalue repeatedly appeared in the roots of the characteristic equation of A. It is obvious that $\sum_{i=1}^m a_i = n$. Let $\rho(A)$ represents the absolute value of the eigenvalue with the greatest absolute value among all eigenvalues. We first apply Jordan Normal Form on A such that $A = MJM^{-1}$ where $M \in \mathbb{R}^{n\times n}$.
By definition of Jordan Normal Form, J is in the form of

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & J_{n_2}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{n_m}(\lambda_m) \end{bmatrix}$$

where $J_{n_i}(\lambda_i) \in \mathbb{R}^{n_i \times n_i}$ is in the form of

$$J_{n_i}(\lambda_i) = \begin{bmatrix} \lambda_{n_i} & 1 & \dots & 0 \\ 0 & \lambda_{n_i} & 1\dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{n_i} \end{bmatrix}$$

$J_{n_i}(\lambda_i)$ can be further decomposed into $J_{n_i}(\lambda_i) = \lambda_i I + Bd_{n_i}$ where Bd_{n_i} represents the bi-diagonal identity matrix.

Then we define diagonal matrix $D(\delta) \in \mathbb{R}^{n\times n}$ such that

$$D(\delta) = \begin{bmatrix} D_{n_1}(\delta) & 0 & \dots & 0 \\ 0 & D_{n_2}(\delta) & 0\dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D_{n_m}(\delta) \end{bmatrix}$$

where $D_{n_i}(\delta) \in \mathbb{R}^{n_i \times n_i}$ is the following:

$$D_{n_i}(\delta) = \begin{bmatrix} \delta & 0 & \dots & 0 \\ 0 & \delta^2 & 0\dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta^{n_i} \end{bmatrix}$$

Since $D_{n_i}(\delta)$ is a diagonal matrix, for all $\delta \in \mathbb{Z}^+$ it is obvious that $(D_{n_i}(\delta))^{-1} = D_{n_i}(\delta^{-1})$, so $(D(\delta))^{-1} = D(\delta^{-1})$.
Then we apply $D(\delta)^{-1}$ and $D(\delta)$ on both sides of J to get a bi-diagonal matrix $B(\delta) = D(\delta)^{-1}JD(\delta)$. For all m blocks on the diagonal of B, we can get $B_{n_i}(\delta) \in \mathbb{R}^{n_i \times n_i}$:

$$B_{n_i}(\delta) = (D_{n_i}(\delta)^{-1})(J_{n_i}(\lambda_i))(D_{n_i}(\delta)) =$$
$$(D_{n_i}(\delta)^{-1})(\lambda_i I)(D_{n_i}(\delta)) + (D_{n_i}(\delta)^{-1})(Bd_{n_i})(D_{n_i}(\delta)) =$$
$$(\lambda_i I) + \begin{bmatrix} \delta^{-1} & 0 & \dots & 0 \\ 0 & \delta^{-2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta^{-n_i} \end{bmatrix} \begin{bmatrix} 0 & \delta^2 & 0 & \dots & 0 \\ 0 & 0 & \delta^3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \delta^{n_i} \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} =$$
$$\begin{bmatrix} \lambda_{n_i} & \delta & 0 & \dots & 0 \\ 0 & \lambda_{n_i} & \delta & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_{n_i} & \delta \\ 0 & \dots & 0 & 0 & \lambda_{n_i} \end{bmatrix}$$

Let $S = D(\delta)^{-1}M$, we can see that $B(\delta) = S^{-1}AS$. Since B and A are similar, we can see that every eigenvalue of B is also an eigenvalue of A, so $\rho(B) = \rho(A)$. We can see that $\inf_{\delta>0} \|B\|_2 = \rho(B) = \rho(A)$.

Then, by the definition of infimum and the definition of Matrix 2-norm, there exist $\epsilon, \delta > 0$ such that

$$\rho(A) \leq \|B(\delta)\|_2 = \|SAS^{-1}\|_2 \leq \rho(A) + \epsilon$$

so $\|SAS^{-1}\|_2 - \epsilon \leq \rho(A)$.

Suppose $Av = \rho(A)v, \|v\|_2 = 1$. By definition of norm, $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\| \geq \|Av\| = \rho(A)$.

Put the conclusions together, we get $\|SAS^{-1}\|_2 - \epsilon \leq \rho(A) \leq \|A\|_2$, which concludes the proof.

1.2 (ii)

Suppose $A \in \mathbb{R}^{n\times n}$. Since A can be diagonalized by one orthogonal matrix, then A can be written as $A = PDP^T$. We can see that under this condition, A is symmetric, so A is diagonalizable. This provides n linear independent eigenvector and n different eigenvalue. Since $\rho(A)$ represents the eigenvalue with the largest absolute value and $\|A\|_2 = \lambda^{1/2}$ where λ represents the eigenvalue with the largest absolute value of $A^T A$. $A^T A = PD^2P^T$. The largest value in D^2 equals $\rho(A)^2$ as D is a diagonal matrix. So we can see that $\|A\|_2 = \lambda^{1/2} = \rho(A)$.

2 Question 2

python code:

```
import numpy as np

def find_U_b(A, b):
    dim = A.shape[0]
    for i in range(dim - 1):
        maxInd = np.argmax(A[i:, i])
        A[[i, i + maxInd], :] = A[[i + maxInd, i], :]
        b[i], b[i + maxInd] = b[i + maxInd], b[i]
        for j in range(1, dim - i):
            L = A[i + j, i] / A[i, i]
            A[i + j, i:] -= L * A[i, i:]
            b[i + j] -= L * b[i]
    return A, b

def find_x(A, b):
    dim = A.shape[0]
    x = np.zeros(dim)
    for i in range(dim - 1, -1, -1):
        x[i] = (b[i] - np.dot(A[i, i+1:], x[i+1:]))/ A[i,i]
    return x

A = np.random.uniform(low= -10, high=10, size=(100, 10))
b = np.random.uniform(low = -10, high = 10, size=(100))
ATA = A.transpose() @ A
ATA_new , b_new = find_U_b(ATA, b)
x = find_x(ATA_new, b_new)
print(x)
print(A @ x)
print(b)
```

3 Question 3

Suppose $b \in \mathbb{R}^n$. Suppose $\alpha, \beta \in \mathbb{R}^+$.
Statement 1: $\|b\|_1 \leq \alpha \|b\|_2$
Proof: By Cauchy–Schwarz inequality, we know that $\sum_{i=1}^{n} u_i v_i \leq \|u\|_2 \|v\|_2$. Therefore, $\sum_{i=1}^{n} b_i (signb)_i \leq \|b\|_2 \|(signb)\|_2$. By definition of 1-norm, we know

that $\|b\|_1 = \sum_{i=1}^{n} b_i(signb)_i$ and $\|(signb)\|_2 = \sqrt{n}$. Choose α to be $\sqrt{n}$, so we have $\|b\|_1 \leq \alpha\|b\|_2$.
Statement 2: $\|b\|_2 \leq \beta\|b\|_1$
Proof: $\|b\|_2 = \sqrt{\sum_{i=1}^{n}(b_i)^2} \leq \sqrt{n(\max_i bi)^2} = \sqrt{n}\,|(\max_i bi)| \leq \sqrt{n}\sum_{i=1}^{n}|bi| = \sqrt{n}\|b\|_1$. Select $\beta = \sqrt{n}$ completes the proof.

4 Question 4

Let $A \in \mathbb{R}^{m\times n}$ where $m \geq n$. Let $x, \delta x \in \mathbb{R}^n$, and $b, \delta b \in \mathbb{R}^m$. Since $(A^TA)^{-1}$ exists, we can see that A should be full rank. Construct pseudo inverse of A to be $A^\dagger = (A^TA)^{-1}A^T$. It is easy to prove that $\|AA^\dagger b\|_2 = \|b\|_2$. So we have $\|A\|_2\|A^\dagger\ b|_2 \geq \|b\|_2$ and $\frac{1}{\|A^\dagger b\|_2} \leq \frac{\|A\|_2}{\|b\|_2}$. We know that x solves $\min\|Ax - b\|_2$ and $x + \delta x$ solves $\min\|A(x + \delta x) - (b - \delta b)\|_2$. So we have normal equations $A^TAx = A^Tb$ and $A^TA(x + \delta x) = A^T(b + \delta b)$. With basic algebra we get $A^TA\delta x = A^T\delta b$, $x = A^\dagger b$, and $\delta x = A^\dagger \delta b$. Therefore, we have:

$$\frac{\|\delta x\|_2}{\|x\|_2} = \frac{\|A^\dagger \delta b\|_2}{\|A^\dagger b\|_2} \geq \frac{\|A^\dagger\|_2\|\delta b\|_2}{\|A^\dagger b\|_2} \geq \|A\|_2\|A^\dagger\|_2\frac{\|\delta b\|_2}{\|b\|_2}$$

Substituting $A^\dagger$ with $(A^TA)^{-1}A^T$ completes the proof.

5 Question 5

python code:

```
import numpy as np

size = (200, 100)
A = np.random.uniform(low = -10, high=10, size = size)
Q = np.zeros(shape=size)
R = np.zeros(shape = (size[1], size[1]))

for col in range(size[1]):
    Qk = Q[:, :col]
    Qtk = Qk.transpose()
    Acol = A[:, col].reshape(-1, 1)
    res = Qtk @ Acol
    diff = Qtk * res
    sum = np.sum(diff, axis = 0).reshape(-1, 1)
    v = Acol - sum
    norm = np.linalg.norm(v)
    q = (v / norm).flatten()
    Q[:, col] = q
    result = np.zeros(size[1])
    result[:len(res)] = res.flatten()
    result[col] = norm
```

```
    R[:, col] = result

print(Q @ R)
print(A)

print(Q @ Q.transpose())
```

6 Question 6

Let $A \in \mathbb{R}^{n \times n}$ be a full rank matrix.
Statement: Prove that the QR factorization of A is unique.
Proof:
We will prove by contradiction. Suppose $A = Q_1 R_1 = Q_2 R_2, Q_1 \neq Q_2$ and $R_1 \neq R_2$, where Q_1, Q_2 are orthogonal matrices and R_1, R_2 are upper triangular matrices. It is easy to prove that the diagonal of R_1, R_2 are positive. Therefore, we can get $Q_2^T Q_1 = R_2 R_1^{-1}$, so $R_2 R_1^{-1}$ is orthogonal and upper triangular. Suppose $U = R_2 R_1^{-1} = (u_1, u_2, ..., u_n)$. We can see that $u_i^T u_j = 0$ when $i \neq j$, $u_i^T u_j = 1$ when $i \neq j$ and the elements starting at $i + 1$ for every u_i is zero. I will prove by induction that that for $i = 1, 2, ..., n$, $u_i = e_i$. When $i = 1$, we know that elements from the second element to the nth element of u_1 are 0, and $u_1^T u_1 = 1$, so the first element of $u_1 = 1$. Suppose for $i \leq k$, $u_i = e_i$, and $u_{k+1} = (v_1, v_2, ..., v_{k+1}, 0, \ldots, 0)^T$ where $j = 1, 2, ..., k$. Since we know that $u_1 = e_1$ and u_1 is orthogonal to u_{k+1}, $u_{k+1}^T u_1 = 0 = v_1$. Similarly, calculating the inner production of u_{k+1} with u_i for $i = 2, 3, ..., k$ will show that $v_2, v_3, ..., v_k = 0$. Since $u_{k+1}^T u_{k+1} = 1$, we can see that $v_{k+1} = 1$, which concludes the proof that $U = I$. Since $U = R_2 R_1^{-1} = I$, $R_2 = R_1$, which contradicts the assumption. Therefore, A can only have one set of QR factorization.