HW4

xh2433

03/26/2024

1 Question 1

Let $Q = \begin{bmatrix} Q' & Q'' \end{bmatrix}$ where $Q' \in \mathbb{R}^{m\times n}$. We know that $A = Q\begin{bmatrix} R \\ 0 \end{bmatrix}$ where $R \in \mathbb{R}^{n\times n}$, so we can find that $A = Q'R + Q''0 = Q'R$. $A^TA = R^TQ'^TQ'R$. Since Q is orthogonal, every column vector of Q is orthonormal, so every column vector of Q' is also orthonormal. Therefore, $Q'^TQ' = I$, and $A^TA = R^TR$. Since A is full rank, x such that $Ax = 0$ implies x is the trivial solution. We can see that A^TA is non-singular by proving that the only solution to $A^TAy = 0$ is trivial. Assume there exists v such that $A^TAv = 0$ and v is not trivial. Notice that $Av \in col(A)$ and $Av \in null(A^T)$. Since for any matrix, $col(A) \perp null(A^T)$, so $Av = 0$. Since A is full rank, the only solution to $Av = 0$ is trivial, which contradicts the assumption. Therefore the only solution to $A^TAy = 0$ is trivial, and A^TA is non-singular. We can get that $det(A^TA) \neq 0$, so $det(R^TR) = det(R^T)det(R) \neq 0$, and $det(R) \neq 0$. By definition of QR-factorization, R is a upper triangular matrix. Since R is non-singular, its diagonal must be nonzero because the product of the elements that lines on the main diagonal of a triangular matrix is the determinant of the matrix, thus ends the proof.

2 Question 2

If μ is a scalar that solves the least squares problem $\min \|Av - \mu v\|_2$, we should know that the inner production between μv and $(Av - \mu v)$ should be 0, so $(\mu v)^T(Av - \mu v) = 0$. Take $\mu = \frac{v^TAv}{v^Tv}$ and substitute in the right side of the statement we will get $\frac{(v^TAv)^2}{v^Tv} - \frac{(v^TAv)^2}{(v^Tv)^2}v^Tv = 0$. Therefore, $\mu = \frac{v^TAv}{v^Tv}$ solves the least squares problem.

3 Question 3

python code:

```
import numpy as np
```

```python
def powerMethod(A, threshold):
    vk1 = np.random.random(size=A.shape[0])
    vk2 = np.random.random(size=A.shape[0])
    it = 0
    while 1:

        vk1 = vk2 / np.linalg.norm(vk2)
        vk2 = A @ vk1
        vk2 = vk2 / np.linalg.norm(vk2)

        Avk2 = A @ vk2
        Avk1 = A @ vk1
        lambda2 = np.dot(vk2, Avk2)
        lambda1 = np.dot(vk1, Avk1)

        print(f"iteration: {it}, diff: {abs(lambda1 - lambda2)}")
        if abs(lambda1 - lambda2) < threshold:
            return vk2, lambda2, it
        it += 1

itPower = []
for num in range (100):
    size = 100
    A = np.random.uniform(low=-10, high=100, size=(size, size))
    # A = A.T @ A
    # this step is to ensure that we don't get complex eigenvalues
    eigenvalues = np.linalg.eigvals(A)
    v, lamda, it = powerMethod(A, 0.00001)
    itPower.append(it)
    max_eigenvalue = np.max(eigenvalues)
itPowernp = np.array(itPower)
avgPower = np.mean(itPower)


def inversePowerMethod(A, threshold):
    vk1 = np.random.random(size=A.shape[0])
    vk2 = np.random.random(size=A.shape[0])
    vk1 = vk1 / np.linalg.norm(vk1)
    vk2 = vk2 / np.linalg.norm(vk2)
    it = 0
    alpha = np.linalg.norm(A)
    B = A - alpha * np.identity(A.shape[0])
    while 1:
        vk1 = vk2
        vk2 = np.linalg.solve(B, vk1)
        vk2 = vk2 / np.linalg.norm(vk2)
```

```python
        lambda2 = np.dot(vk2, A @ vk2)
        lambda1 = np.dot(vk1, A @ vk1)
        B = A - lambda2 * np.identity(A.shape[0])
        it += 1

        print(f"iteration: {it}, diff: {abs(lambda1 - lambda2)}")
        if abs(lambda1 - lambda2) < threshold:
            return vk2, lambda2, it

itInverse = 0
for num in range(100):
    size = 100
    A = np.random.uniform(low=-10, high=100, size=(size, size))
    v, lamda, it= inversePowerMethod(A, 0.00001)
    itInverse += it
    eigenvalues = np.linalg.eigvals(A)
    max_eigenvalue = np.max(eigenvalues)
avgInverse = itInverse / 100

print(f"avg Power: {avgPower}, avg Inverse: {avgInverse}")
```

3.1 iii

from the code I get the following results: **avg Power: 5.76, avg Inverse: 4.0** we see that for inverse power method, the average iteration is 1.76 lower than power methods, so inverse power method performs better.

4 Question 4

4.1 i

Since $A(\alpha) = \alpha P + (1-\alpha)E$, by definition of E, for any j where $j = 1, ..., n$, we will have the j_{th} column sum of $A(\alpha)$ to be $\sum_{i=1}^{n} \alpha P_{ij} + \frac{1-\alpha}{n} = \alpha(\sum_{i=1}^{n} P_{ij}) + n\frac{1-\alpha}{n} = \alpha \cdot 1 + 1 - \alpha = 1$. Since all the elements of P and all the elements of E are non-negative, all the elements of $A(\alpha)$ should also be non-negative. Since for $A(\alpha)$ every column add to 1 and all the elements are non-negative, we can see that $A(\alpha)$ is column-stochastic matrix.

Now observe that $A(\alpha)^T = \alpha P^T + (1-\alpha)E^T = \alpha P^T + (1-\alpha)E$. Since we know that the rows of P also add up to 1, with the same method of proving, we can see that $A(\alpha)^T$ also has column sums of 1. Therefore, both the column sum and row sum of A add up to 1.

4.2 ii

From the conclusion from part (i), we know that A is a column-stochastic matrix with non-negative entries, and A has column sums of 1 and row sums of 1. We can see that $l = (1, \ldots, 1)^T, u \in \mathbb{R}^n$ is an eigenvector of A, because $Al = (\sum_{i_1=1}^n A_{1i_1} \cdot 1, \ldots, \sum_{i_n=1}^n A_{ni_n} \cdot 1)^T = (1, \ldots, 1)^T$, where A_{ij} represent the $(i, j)th$ element of A. Therefore, l is an eigenvector of A and 1 is the eigenvalue corresponding to l.

Suppose λ, u is a random eigenvalue-eigenvector pair of A. Observe that for some $j = 1, ..., n$, $\lambda u_j = \sum_{i=1}^n A_{ji} u_i$. Suppose $u_k = \max_j u_j$. Then $|\lambda| \cdot |u_k| = |\sum_{i=1}^n A_{ki} u_i| \leq \sum_{i=1}^n A_{ki} |u_i| \leq \sum_{i=1}^n A_{ki} |u_k| = 1 \cdot |u_k|$. Therefore, any random eigenvector of A is at most 1, so 1 is the largest eigenvalue of A, and l is the eigenvector corresponding to the largest eigenvalue.

4.3 iii

As proved above, by definition of stochastic-column matrices, all elements of A, P are non-negative. Compose $A' = A - v_1 v_1^T$ where $v_1 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})^T$. Observe that for every $v_i, i \neq 1$, where v_i is an eigenvector of A, there exists $u_i := (v_i - \frac{1}{\lambda_i} v_1 v_1^T v_i)$ where $Av_i = \lambda_i v_i$ such that u_i is the eigenvector of A' with $A' u_i = \lambda_i u_i, i \neq 1$, because:

$$\begin{aligned}
A' u_i &= A v_i - \frac{1}{\lambda_i} v_1 v_1^T v_i - v_1 v_1^T v_i + \frac{1}{\lambda_i} v_1 v_1^T v_1 v_1^T v_i \\
&= A v_i - v_1 v_1^T v_i \\
&= \lambda_i v_i - v_1 v_1^T v_i \\
&= \lambda_i (v_i - \frac{1}{\lambda_i} v_1 v_1^T v_i)
\end{aligned}$$

By definition of E, every entry of E is $\frac{1}{n}$, so notice that $v_1 v_1^T = E$. By definition of A, $A = \alpha P + (1-\alpha)E$, so $A' = \alpha P + (1-\alpha)E - v_1 v_1^T = \alpha P + (1-\alpha)E - (1-\alpha)E - \alpha v_1 v_1^T = \alpha(P - v_1 v_1^T)$. Therefore, $\lambda_2(A) = \lambda_1(A') = \alpha \lambda_2(P)$. Since P is an column-stochastic matrix, $|\lambda_2(P)| \in [0, 1]$, so $\lambda_2(A)$ is bounded by α.

5 Question 5

5.1 i

We know that $(D + \epsilon A)(e + \epsilon u) = (\lambda + \epsilon \mu)(e + \epsilon u)$. Since we assume that $\epsilon^2 = 0$, expand both sides and we get $De + \epsilon Du + \epsilon Ae = \lambda e + \epsilon \lambda u + \epsilon \mu e$. Rearrange the equation we get $De - \lambda e = \epsilon(\lambda u + \mu e - Du - Ae)$. Multiply both sides by ϵ and we get $\epsilon(De - \lambda e) = 0$, and $De - \lambda e = 0$. Therefore, λ and e is an eigenvalue-eigenvector pair of D. Since D is a diagonal matrix, we know that $\lambda = d_{jj}$ and e is the jth column vector of identity matrix $I \in \mathbb{R}^{n \times n}$. Subtract both sides of the original equation by λe and we get $Du + Ae = \lambda u + \mu e$ so $(D - \lambda I)u = (\mu I - A)e$. Observe that the left-hand side of the equation can be represented as $((d_{11}-\lambda)u_1,\ldots,(d_{nn}-\lambda)u_n)^T$, and the right-hand side is essentially the jth column of matrix $\mu I - A$, which can be represented as $(-a_{1j},\ldots,\mu-a_{jj},\ldots,-a_{nj})^T$. Therefore, we see that $(d_{ii}-\lambda)u_i = -a_{ij}$ when $i \neq j$, so we get $u_i = -\frac{a_{ij}}{d_{ii}-\lambda}$. When $i = j$, $(d_{jj}-\lambda)u_j = 0 = \mu - a_{jj}$ so $\mu = a_{jj}$. Since we want eigenvectors to be normalized, $\|e+\epsilon u\|_2 = 1$, and $(e+\epsilon u)^T(e+\epsilon u) = \|e+\epsilon u\|_2^2 = 1$. Expand the expression to get $e^Te+2\epsilon e^Tu = 1$. e is proved above to be the jth column of the identity matrix, so $e^Te = 1$, and $2\epsilon e^Tu = 0$, $e^Tu = 0$. e^Tu gives us the jth element in u, so $u_j = 0$.

5.2 ii

Write the matrix multiplication in linear equations and we get

$$d_{11}e + \epsilon d_{11}u + \epsilon^2 d_{11}x + \epsilon A_1 e + \epsilon^2 A_1 u + \epsilon A_2 f + \epsilon^2 A_2 v = \lambda e + \lambda\epsilon u + \lambda\epsilon^2 x + \epsilon\mu e + \epsilon^2\mu u + \epsilon^2 ve$$
$$\epsilon A_2^T e + \epsilon^2 A_2^T u + D_{n-k}f + \epsilon D_{n-k}v + \epsilon^2 D_{n-k}y + \epsilon A_3 f + \epsilon^2 A_3 v = \lambda f + \epsilon\lambda u + \epsilon^2\lambda y + \epsilon\mu f + \epsilon^2\mu u + \epsilon^2 uf$$

equating both sides with the same power of ϵ we get: $d_{1,1}e = \lambda e$, so $\lambda = d_{1,1}$. Similarly, $D_{n-k}f = \lambda f$. This implies that f is an eigenvector of D_{n-k}, which implies $\lambda = d_{n-k,n-k}, d_{n-k+1,n-k+1}, ..., d_{n,n}$, but $\lambda = d_{1,1}$ and $d_{1,1}$ is distinct to $d_{n-k,n-k}, d_{n-k+1,n-k+1}, ..., d_{n,n}$, so f can only be 0.

Simplify the equations to get:

$$d_{11}e + \epsilon d_{11}u + \epsilon^2 d_{11}x + \epsilon A_1 e + \epsilon^2 A_1 u + \epsilon^2 A_2 v = \lambda e + \epsilon\lambda u + \epsilon^2\lambda x + \epsilon\mu e + \epsilon^2\mu u + \epsilon^2 ve$$
$$\epsilon A_2^T e + \epsilon^2 A_2^T u + \epsilon D_{n-k}v + \epsilon^2 D_{n-k}y + \epsilon^2 A_3 v = \epsilon\lambda u + \epsilon^2\lambda y + \epsilon^2\mu u$$

Similarly equating both sides to get

$$\begin{aligned} d_{11}u + A_1 e &= \lambda u + \mu e \\ d_{11}x + A_1 u + A_2 v &= \lambda x + \mu u + ve \\ A_2^T e + D_{n-k}v &= \lambda u \\ A_2^T u + D_{n-k}y + A_3 v &= \lambda y + \mu u \end{aligned}$$

substitute λ by d_{11} and simplify to get:

$$A_1 e = \mu e \tag{1}$$
$$A_1 u + A_2 v = \mu u + ve \tag{2}$$
$$A_2^T e + D_{n-k}v = \lambda u \tag{3}$$
$$A_2^T u + D_{n-k}y + A_3 v = \lambda y + \mu u \tag{4}$$

Therefore, e is an eigenvector of A_1, v can be solved from the equation (3) $A_2^T e + D_{n-k}v = \lambda u$, which is $(D_{n-k}-\lambda I)v = -A_2^T e$. From equation (2) we see $(A_1 - \mu I)u = ve - A_2 v$.