ECS 230 - Homework 05

- Student: Qirun Dai (qrdai@ucdavis.edu)

1. Question 1 (Proof)

12/8. Hw 05.

1. Let's prove the statement by showing that: "the characteristic polynomial of matrix A is the same as that of matrix A^T."

For A, $P(\lambda) \triangleq \det(\lambda I - A)$.

For A^T, $P_T(\lambda) \triangleq \det(\lambda I - A^T)$

Lemma 1: For any matrix $X \in \mathbb{R}^{n\times n}$, $\det(X) = \det(X^T)$.

simple proof: $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$, $X^T = \begin{bmatrix} x_{11} & x_{21} & & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & x_{nn} \end{bmatrix}$.

Let's prove by induction:

① Base condition: $X \in \mathbb{R}^{1\times 1}$. obvious.

② Inductive Hypothesis:

For any $X \in \mathbb{R}^{(n-1)\times(n-1)}$, $\det(X) = \det(X^T)$ always holds.

③ Inductive Step: now prove for any $X \in \mathbb{R}^{n\times n}$, $\det(X) = \det(X^T)$ still holds.

Using Laplace Expansion on the first row of X:

$$\det(X) = x_{11} \cdot \det\left(\begin{bmatrix} x_{22} & \cdots & x_{2n} \\ & \vdots & \\ x_{n2} & \cdots & x_{nn} \end{bmatrix}\right) + \cdots + (-1)^{1+n} x_{1n} \cdot \det\left(\begin{bmatrix} x_{21} & \cdots & x_{2,n-1} \\ & & \\ x_{n1} & \cdots & x_{n,n-1} \end{bmatrix}\right)$$

Using Laplace Expansion on the first column of X^T:

$$\det(X^T) = x_{11} \cdot \det\left(\begin{bmatrix} x_{22} & \cdots & x_{n2} \\ & \vdots & \\ x_{2n} & \cdots & x_{nn} \end{bmatrix}\right) + \cdots + (-1)^{1+n} x_{1n} \cdot \det\left(\begin{bmatrix} x_{21} & & x_{n1} \\ \vdots & \cdots & \vdots \\ x_{2,n-1} & & x_{n,n-1} \end{bmatrix}\right)$$

From Inductive Hypothesis, we know all the $(n-1)\times(n-1)$ determinant in the 2 expressions above are correspondingly equal. Then $\det(X) = \det(X^T)$. $\square$

Now with Lemma 1 proved, we can proceed to prove the original statement easily:

$$P_T(\lambda) = \det(\lambda I - A^T) = \det((\lambda I - A)^T) \underset{\text{Lemma 1}}{=} \det(\lambda I - A) = P(\lambda).$$

Therefore $P_T(\lambda) = P(\lambda)$; Then for any eigenvalue λ of A, λ must also be an eigenvalue of A^T.

2. Question 2-4 (Basic Power Method)

2.1 Question 2 and 3

1. Please refer to source file `power.c` for the implementation of Power Method with $l1-normalization$ (without shifting for now). The results from the paper can be reproduced and verified.

2.2 Question 4

1. Using the strategy described in Exercise 1, the score of node 3 can be improved. The resulting importance scores of the 5 nodes are presented below:

$\begin{pmatrix} 0.244898 \\ 0.081633 \\ 0.367347 \\ 0.122449 \\ 0.183673 \end{pmatrix}$, while the original importance scores without adding node 5 are $\begin{pmatrix} 0.387097 \\ 0.129032 \\ 0.290323 \\ 0.193548 \end{pmatrix}$

   - As can be seen, the score of node 5 increases from 0.290 to 0.367.
   - The effect of this strategy can be intuitively interpreted. After adding node 5 that only links to node 3 and also adding an additional link from node 3 to node 5, two things happen. First, node 5 gives a strong vote to node 3, since the only forward link from node 5 goes to node 3. Also, the additional link from node 3 to 5 splits the vote from node 3 into two halves, making node 1 receive fewer importance scores from node 3. As a result, the importance score of node 3 increases and the score of node 1 decreases, finally making node 3 the node with the highest importance.

3. Question 5-7 (Shifted Power Method)

3.1 Question 5

1. The link matrix of the linear chain with `n=4` is presented below. Note that it's a tridiagonal matrix with all-zero diagonal components.
   - $A = \begin{pmatrix} 0.000000 & 0.500000 & 0.000000 & 0.000000 \\ 1.000000 & 0.000000 & 0.500000 & 0.000000 \\ 0.000000 & 0.500000 & 0.000000 & 1.000000 \\ 0.000000 & 0.000000 & 0.500000 & 0.000000 \end{pmatrix}$
2. Using the power method without shifting, we start from two different initial vectors to compute the eigenvector associated with eigenvalue $\lambda = 1$.
   a. Using initial vector `[1, 1, 1, 1]`, we can normally compute the resulting eigenvector: $\begin{pmatrix} 0.166667 \\ 0.333333 \\ 0.333333 \\ 0.166667 \end{pmatrix}$, or in precise fraction form $\begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}$, which is indeed an eigenvector of $\lambda = 1$.

b. However, using initial vector `[1, 2, 3, 4]`, the power method without shifting fails to converge. The computed result oscillates between $\begin{pmatrix} 0.133333 \\ 0.400000 \\ 0.266667 \\ 0.200000 \end{pmatrix}$ and $\begin{pmatrix} 0.200000 \\ 0.266667 \\ 0.400000 \\ 0.133333 \end{pmatrix}$, both of which have the same residue value 0.133333.

3. Now let's analyze why the results from different initial vectors are different. Let's first compute the eigenvalues of A, which yields four distinct eigenvalues: $1, \frac{1}{2}, -\frac{1}{2}, -1$. We can see there are two distinct eigenvalues with the same absolute value: 1 and -1. This actually violates the convergence condition of the power method, which requires there is only one eigenvalue with the largest absolute value. Therefore, under such conditions, the vector sequence produced by power method does not necessarily converge to a unique eigenvector of λ_1, and its convergence depends on the choice of the initial vector. Actually, now the vector sequence produced by power method has multiple subsequences which converge to different resulting vectors.

3.2 Question 6

1. Please refer to source file **`shifted_power.c`** for the implementation of the shifted power method.
2. We have already shown that the four distinct eigenvalues of matrix A are $1, \frac{1}{2}, -\frac{1}{2}, -1$. In order to solve the problem in question 5, we can apply shifting to the power method so that there aren't two distinct eigenvalues with the same absolute value. Specifically, there are two choices of the shift, depending on the eigenvector of which eigenvalue we want to compute. Firstly, we denote σ as the shift value, and the shifted matrix as $A - \sigma I$.
   a. If we still want to compute the eigenvector of $\lambda = 1$ and also want to ensure that it can always be computed given whatever initial vector, we can take $\sigma = -\frac{1}{4}$. Now the eigenvalues of matrix $A - \sigma I$ are $\frac{5}{4}, \frac{3}{4}, -\frac{1}{4}, -\frac{3}{4}$, with $\frac{5}{4}$ and $\frac{3}{4}$ being the largest and second largest eigenvalues, and the convergence rate is $O\left(\left(\frac{3}{5}\right)^k\right)$ accordingly. For any initial vectors (except for those orthogonal to the eigenvector of $\lambda = 1$), we can compute the eigenvector of $\lambda = 1$, which is the same as above: $\begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}$.

b. If we want to compute the eigenvalue of $\lambda = -1$, we can similarly take $\sigma = \frac{1}{4}$. Now the eigenvalues of matrix $A - \sigma I$ are $\frac{3}{4}, \frac{1}{4}, -\frac{3}{4}, -\frac{5}{4}$, with $-\frac{5}{4}$ and $\frac{3}{4}$ being the largest and second largest eigenvalues, and the convergence rate is also $O\left(\left(\frac{3}{5}\right)^k\right)$. For any initial vectors (except for those orthogonal to the eigenvector of $\lambda = -1$), we can compute the eigenvector of $\lambda = -1$, which is the similar in format to the result above: $\begin{pmatrix} \frac{1}{6} \\ -\frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{6} \end{pmatrix}$. Specifically, since `[1, 1, 1, 1]` is orthogonal to the resulting eigenvector, we cannot use it as the initial vector, otherwise the power method cannot converge even with shifting.

3.3 Question 7

1. The relative merits of shifted power method:
   a. It still computes the eigenvectors and eigenvalues of the original matrix A.
2. The relative merits of the modified power method using the matrix S in the paper:
   a. From the paper we know that for any column-stochastic matrix A, the matrix we compute, $M = (1-m)A + mS$, must be strictly positive and also column-stochastic. This ensures that there is only one unique eigenvector q of eigenvalue $\lambda = 1$ and it can be computed and converge to starting from any initial vector x_0. Therefore, we do not need to worry about the actual eigenvalue distribution of the target matrix, and can always directly compute a correct answer using normal power method. In contrast, for the shifted power method, we must first know the exact eigenvalue distribution of the target matrix, before we can decide the shift value. Sometimes, even if we know all the eigenvalues of the target matrix, it may still be difficult to find an appropriate shift value that can accelerate convergence.
   b. Now let's compare the convergence rate of these two methods in the limit of large n. For shifted power method, as n goes to $+\infty$ for the link matrix of the linear chain, its eigenvalues will still be evenly distributed on the positive and negative side of 0, and will gradually concentrate near ± 1. Therefore, even the optimal convergence rate as n goes to $+\infty$ will also gradually converge to 1. For the modified power method with S, since $M = (1-m)A + mS$, and from the paper we know that it can be shown that the absolute value of λ_2 is always lower than or equal to $1-m$, the convergence rate of whatever large `n` is thus always $1-m$, where `m` is a hyperparameter that can be tuned by ourselves. Therefore, the modified power method with S also has a better and more flexible convergence rate in the limit of large `n` for the linear chain.