

AM 213A HW3

Joseph Moore

Winter 2022

25
30

Part 1

1. The first run of this algorithm gave me

$$\begin{bmatrix} 5.000000 & -4.231182 & -0.274848 & -0.146834 \\ -4.231182 & 5.787425 & 1.298516 & 0.693716 \\ -0.274848 & 1.298516 & 4.499262 & 1.335199 \\ -0.146834 & 0.693716 & 1.335199 & 2.713313 \end{bmatrix}$$

This is not an iterative method, there is no "convergence"

5/10

I then applied it in series four times to get

$$\begin{bmatrix} 5.000000 & 4.242640 & 0.003049 & 0.000373 \\ 4.242640 & 5.997951 & 1.404444 & 0.171993 \\ 0.003049 & 1.404444 & 4.957691 & 0.362209 \\ 0.000373 & 0.171993 & 0.362209 & 2.044357 \end{bmatrix}$$

It should solve a tri-diag directly.

Clearly converging to a tridiagonal matrix.

this isn't a valid criterion for convergence, but ok.

- i) With out shift I would assume that the convergence would happen much slower than with convergence. For measuring the error in the matrix I took the absolute value of the difference between the two norm of the first column and the first element in that column. This essentially measures how diagonal the matrix is. I looped this algorithm until the error was less than 10^{-8} in 14 iterations. I got

$$\begin{bmatrix} 3.732051 & 0.000204 & 0.000000 \\ 0.000204 & 2.000000 & 0.000000 \\ 0.000000 & 0.000000 & 0.267949 \end{bmatrix}$$

not quite converged yet.

10/10

- ii) For this one I used the same error method. I was, however, unable to iterate the loop past a certain point before I ran into errors. If I require the error to be less than 10^{-8} the matrix elements become 'nans'. The algorithm achieved this accuracy in 12 iterations. I got

$$\begin{bmatrix} 3.732051 & 0.000195 & 0.000000 \\ 0.000195 & 2.000000 & 0.000000 \\ 0.000000 & 0.000000 & 0.267949 \end{bmatrix}$$

This is one motivation for deflation.

3. For this problem I solved $(A - \lambda I)y = x$ for y instead of finding $B = (A - \lambda I)^{-1}$ and applying it to x . They are analytically equivalent while the former being much easier to compute. I ran the program for each eigenvalue.

$\frac{10}{10}$

$$\lambda_1 = -8.0286$$

$$\begin{bmatrix} -0.263462 \\ -0.659041 \\ 0.199634 \\ 0.675573 \end{bmatrix}$$

✓

$$\lambda_2 = 7.932900$$

$$\begin{bmatrix} -0.560145 \\ -0.211633 \\ -0.776708 \\ -0.195382 \end{bmatrix}$$

✓

$\lambda_3 = 5.668900$ The vector for this problem kept alternating across the origin so that $y - x$ was exactly 2 even though the eigenvector was correct. I had to manually stop the loop by inspection. It took less than 10 iterations to get a satisfactory answer.

$$\begin{bmatrix} -0.378703 \\ -0.362419 \\ 0.537935 \\ -0.660199 \end{bmatrix}$$

✓

$$\lambda_3 = -1.573200$$

$$\begin{bmatrix} -0.688048 \\ 0.624123 \\ 0.259801 \\ 0.263750 \end{bmatrix}$$

✓

→ should use

$$\min(\|x - y\|, \|x + y\|)$$

For this reason.

Part 2

1. First we decompose A into QTQ^* where Q is unitary and thus preserves the eigenvalues of a matrix. From this we can deduce that $\rho(U) = \rho(A)$. We can also use the fact that $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|_2^{1/k}$ and thus $\|A^n\|_2 = \|U^n\|_2$ as $n \rightarrow \infty$. Using the fact that all norms are equivalent in a finite vector space we need only show that $\lim_{n \rightarrow \infty} \|U^n\|_2 = 0 \Leftrightarrow \rho(U) < 1$. We have also shown that $\rho(U^n) = \rho(U)^n$. Now,

$$\rho(U)^n = \rho(U^n) \leq \|U^n\|.$$

If $\lim_{n \rightarrow \infty} \|U^n\| = 0$ then $\lim_{n \rightarrow \infty} \rho(U)^n = 0$ which means that $\rho(U) < 1$.

We then split up U into $D + N$. where D is only the diagonal elements and N is the rest. Now using the binomial expansion we have

$$(D + N)^n = \sum_{k=0}^n \binom{n}{k} D^{n-k} N^k.$$

The eigenvalues of a upper triangular matrix are the diagonal entries thus $\rho(D) = \rho(T)$. We have also shown before that the 2-norm of a diagonal matrix is $\max_i |\lambda_i|$. Thus by taking the norm of both sides we can now say

$$\|U^n\| \leq \sum_{k=0}^n \binom{n}{k} \|D\|^{n-k} \|N\|^k = \sum_{k=0}^n \binom{n}{k} \rho(U)^{n-k} \|N\|^k.$$

Now if $\rho(U) < 1$ then $\lim_{n \rightarrow \infty} \rho(U)^{n-k} = 0$ and thus $\lim_{n \rightarrow \infty} \|U^n\| = 0$.

2. If we can create the same characteristic polynomial then we will have the same eigenvalues. Thus it is sufficient to show that

$$\det \begin{pmatrix} AB - \lambda I & 0 \\ B & -\lambda I \end{pmatrix} = \det \begin{pmatrix} -\lambda I & 0 \\ B & BA - \lambda I \end{pmatrix}.$$

The first matrix can become

$$\det \begin{pmatrix} AB - \lambda I & 0 \\ B & -\lambda I \end{pmatrix} = \det \begin{pmatrix} A & -\lambda I \\ I & 0 \end{pmatrix} \begin{pmatrix} B & -\lambda I \\ I & -A \end{pmatrix}$$

Now because $\det(AB) = \det(BA)$ we have

$$\begin{aligned} \det \begin{pmatrix} B & -\lambda I \\ I & -A \end{pmatrix} \begin{pmatrix} A & -\lambda I \\ I & 0 \end{pmatrix} &= \det \begin{pmatrix} BA - \lambda I & \lambda B \\ 0 & -\lambda I \end{pmatrix} \\ &= \det(BA - \lambda I) \det(-\lambda I) = \det \begin{pmatrix} -\lambda I & 0 \\ B & BA - \lambda I \end{pmatrix}. \end{aligned}$$

3. First we establish that the eigenvalues of A^T are the same as A . If we perform cofactor expansion on both matrices and chose the first row and column of the matrices respectively, we will get the same coefficients multiplied by A_{ij} and A_{ij}^T respectively. If we continue this process recursively, we will arrive at the same base case for both matrices, $|a_{ij}|$. Thus the eigenvalues of A^T are the same as A . It is then obvious that the row sum of A minus the diagonal element is the same as the respective column sum of A^T minus the diagonal element.

4. First we look at rowwise disks

$$r_1 = 0.3 + 0.1 + 0.4 = 0.8 - (disk : 1.0 \pm 0.8)$$

$$r_2 = 0.0 + 0.0 + 0.1 = 0.1 - (disk : 2.0 \pm 0.1)$$

$$r_3 = 0.0 + 0.4 + 0.0 = 0.4 - (disk : 3.0 \pm 0.4)$$

$$r_3 = 0.1 + 0.0 + 0.0 = 0.1 - (disk : 4.0 \pm 0.1)$$

no columnwise disks

$$c_1 = 0.0 + 0.0 + 0.1 = 0.1 - (disk : 1.0 \pm 0.1)$$

$$c_2 = 0.3 + 0.4 + 0.0 = 0.7 - (disk : 2.0 \pm 0.7)$$

$$c_3 = 0.1 + 0.0 + 0.0 = 0.1 - (disk : 3.0 \pm 0.1)$$

$$r_3 = 0.4 + 0.1 + 0.0 = 0.5 - (disk : 4.0 \pm 0.5)$$

There are four disjointed disks which means that each one should contain at least one eigenvalue. The matrix can have at most four eigenvalues and thus each disk must have exactly one eigenvalue. We can tell by inspection that each row disk corresponds with a column disk and must describe the same eigenvalue. We can take the smaller of the two disks as a more accurate range. Thus each eigenvalue can be said to lie in one of the four circles

$$|z - k| \leq 0.1, k = 1, 2, 3, 4.$$

5. First we need to establish what scaling a matrix does to the eigenvalues. We note that characteristic polynomial for a matrix A scaled by α is $\det(\alpha A - \lambda I)$. Then we have

$$\det(\alpha(A - \frac{\lambda}{\alpha}I)) = \alpha^m \det(A - \frac{\lambda}{\alpha}I) = 0 \Rightarrow \det(A - \frac{\lambda}{\alpha}I) = 0.$$

from this we can see that any eigenvalue λ_i will be given by

$$\left(\frac{\lambda}{\alpha} - \lambda_i\right) = 0 \Rightarrow \lambda = \alpha \lambda_i.$$

Thus every eigenvalue will be scaled by the same value. From here we note that the eigenvectors of a real valued SPD matrix will have an orthogonal set of eigenvectors and thus span \mathbf{R}^m . We can then write any vector y as a linear combination of eigenvectors $\sum_{i=1}^m a_i v_i$. From here we see that

$$\begin{aligned} \frac{y^T A^{k+1} y}{y^T A^k y} &= \frac{\lambda_1^{k+1} y^T (\frac{1}{\lambda_1} A)^{k+1} y}{\lambda_1^k y^T (\frac{1}{\lambda_1} A)^k y} = \frac{\lambda_1 y^T (\frac{1}{\lambda_1} A)^{k+1} \sum_{i=1}^m a_i v_i}{y^T (\frac{1}{\lambda_1} A)^k \sum_{i=1}^m a_i v_i} \\ &\Rightarrow \frac{\lambda_1 y^T \sum_{i=1}^m a_i (\frac{1}{\lambda_1} A)^{k+1} v_i}{y^T \sum_{i=1}^m a_i (\frac{1}{\lambda_1} A)^k v_i} = \frac{\lambda_1 y^T \sum_{i=1}^m a_i (\frac{\lambda_i}{\lambda_1})^{k+1} v_i}{y^T \sum_{i=1}^m a_i (\frac{\lambda_i}{\lambda_1})^k v_i} \end{aligned}$$

Because all the eigenvalues are divided by the maximum eigenvalue, the maximum eigenvalue will be one while all others will be less than one. This will cause all components of the sum, except $i = 1$, to vanish as $k \rightarrow \infty$. We are then left with

$$\frac{\lambda_1 y^T a_1 (\frac{\lambda_1}{\lambda_1})^{k+1} v_i}{y^T a_1 (\frac{\lambda_1}{\lambda_1})^k v_i} = \frac{\lambda_1 y^T a_1 v_i}{y^T a_1 v_i} = \lambda_1.$$

6. The maximum value any eigenvalue can take is a diagonal element plus the corresponding rows radius of convergence determined by $r_i = \sum_{j=1}^m a_{ij} - a_{ii}$. Suppose that the maximum of this sum is given by row i . We then have

$$\lambda_{max} \leq a_{ii} + r_i = a_{ii} + \left(\sum_{j=1}^m a_{ij} - a_{ii} \right) = \sum_{j=1}^m a_{ij} = 1.$$

A similar argument can be made for the minimum eigenvalue with i chosen to find the minimum value in any disk. We have

$$\lambda_{min} \geq a_{ii} - r_i = a_{ii} - \left(\sum_{j=1}^m a_{ij} - a_{ii} \right) = a_{ii} - \sum_{j=1}^m a_{ij} + a_{ii} = 2a_{ii} - 1 \geq -1.$$

The last inequality being true because all elements are positive.

7.

a) First we know that a normal matrix can be diagonalized and decomposed into SVD. It follows then that

$$\begin{aligned} A^T A &= (U \Sigma V^*)^* (U \Sigma V^*) = V \Sigma U^* U \Sigma V^* = V \Sigma^* \Sigma V^* \\ &= V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_m^2 \end{bmatrix} V^* = V D_\lambda V^*. \end{aligned}$$

Thus the eigenvalues of $A^T A$ are σ_i^2 . Since A^* is diagonalizable we can do

$$A^T = (V D_\lambda V^T)^T = V D_\lambda^T V^T = V D_\lambda V^T.$$

Thus the eigenvalues are preserved by the transpose. We can then write

$$A^* A v_i = A^* \lambda_i v_i = \lambda_i A^* v_i = |\lambda_i|^2 v_i.$$

From this we can conclude that

$$|\lambda_i|^2 = \sigma_i^2 \Rightarrow |\lambda_i| = \sigma_i$$

where $1 \leq i \leq m$.

b) We start with a vector decomposed into the sum of orthonormal eigenvectors $x = \sum_{i=1}^m a_i v_i$. We will also assume that $\|x\| = 1$. Then

$$\|x\|^2 = \sum_{i=1}^m a_i^2 v_i^T v_i = \sum_{i=1}^m a_i^2 \Rightarrow \|x\| = \sqrt{\sum_{i=1}^m a_i^2}.$$

If $\|x\| = 1$ then

$$1 = \sqrt{\sum_{i=1}^m a_i^2} \Rightarrow 1 = \sum_{i=1}^m a_i^2.$$

and

$$A^T Ax = A^T A \sum_{i=1}^m a_i v_i = \sum_{i=1}^m a_i A^T A v_i = \sum_{i=1}^m a_i \lambda_i^2 v_i$$

and

$$\|Ax\|^2 = (Ax)^T Ax = x^T A^T Ax = x^T \sum_{i=1}^m a_i \lambda_i^2 v_i = \sum_{i=1}^m a_i^2 \lambda_i^2 v_i^T v_i = \sum_{i=1}^m a_i^2 \lambda_i^2 \leq \lambda_1^2 \sum_{i=1}^m a_i^2$$

$$\Rightarrow \|Ax\| \leq \lambda_1 \sqrt{\sum_{i=1}^m a_i^2} = \lambda_1$$

gets us

$$\|A\| = \max\{\|Ax\|, \|x\| = 1\} = \max_{\|x\|=1} \sqrt{\|Ax\|^2} \leq \max_{\|x\|=1} \sqrt{\lambda_1^2 \sum_{i=1}^m a_i^2} = \lambda_1.$$

If we let $x = v_1$ then we notice that

$$\|A\|^2 \geq \|Ax\| = x^T A^T Ax = \lambda_1^2 \|x\| = \lambda_1^2.$$

$$\Rightarrow \|A\| \geq \lambda_1.$$

Therefore $\|A\| = \lambda_1$. It follows immediately that $\|A\| = \sigma_1 = \rho(A)$.