CS 513 Assignment 2

Ruochen Lin

February 27, 2018

1

1.1

The characteristic equation of matrix A is

$$(1-\lambda)(2-\lambda)+2=0,$$

solving which will give us the two eigenvalues of A:

$$\lambda = \frac{3 \pm \sqrt{7}i}{2}.$$

They both have the same 'length' in the complex plane:

$$|\lambda| = \sqrt{(\frac{3}{2})^2 + (\frac{\sqrt{7}}{2})^2} = \sqrt{\frac{16}{4}} = 2,$$

which is the spectral radius of A.

1.2

The l_1- and l_∞-norms of A are easy to find:

$$\|A\|_1 = 2+1 = 3,\ \|A\|_\infty = 2+2 = 4.$$

Finding the l_2-norm, on the other hand, goes hand-in-hand with the SVD of A; so we give the result below, but postpone the corresponding calculation to the next subsection:

$$\|A\|_2 = 2\sqrt{2}.$$

1.3

$$A^TA = \begin{bmatrix} 2 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix},$$
$$AA^T = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}.$$

Schur decomposing A^TA and AA^T gives us the left and right singular matrices of A, U and V:

$$A^TA = V\Lambda V^T,$$
$$AA^T = U\Lambda U^T.$$

Solving the characteristic equation of A^TA,

$$(5-\lambda)^2 - 9 = 0,$$

gives us the eigenvalues of A^TA,

$$\lambda_1 = 2,\ \lambda_2 = 8,$$

and the corresponding normalised eigenvectors:

$$v_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix},$$
$$v_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}.$$

Thus we have:

$$\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix},$$
$$V = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

The square roots of the eigenvalues gives us the sigular values of A:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{2},\ \sigma_2 = \sqrt{\lambda_2} = 2\sqrt{2}.$$

The largest of the two ($\sigma_2 = 2\sqrt{2}$) is also the l_2-norm of A.

In order to find U, we can decompose AA^T in the same fashion; however, since we have figured out Λ and V, an easier way of finding U would be recognizing that:

$$\begin{aligned} Av_1 &= \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix} \\ &= \sqrt{2}\, u_1, \\ Av_2 &= \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{-\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} -2\sqrt{2} \\ 0 \end{bmatrix} \\ &= 2\sqrt{2}\, u_2. \end{aligned}$$

Hence

$$U = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

and

$$A = U\Sigma V^T = U\Lambda^{\frac{1}{2}}V^T,$$

where Σ is defined as a diagonal matrix with diagonal entries $\Sigma_{ii} = \sigma_i$.

Indeed, we find that

$$\begin{aligned} AV &= \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 & -\sqrt{2} \\ \sqrt{2} & 0 \end{bmatrix} \\ &= U\Sigma = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix} \end{aligned}$$

and that

$$\begin{aligned} A^T U &= \begin{bmatrix} 2 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 1 & 2 \end{bmatrix} \\ &= V\Sigma = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}. \end{aligned}$$

2

2.1

After playing around in `MATLAB`, (code can be found in Appendix B,) my finding is that

Theorem Given a symmetric matrix A, (λ, v) or $(-\lambda, v)$ is an eigenpair of $A^T A = A^2$ if and only if (λ^2, v) is an eigenpair of $A^T = A$.

We might be tempted to state a stronger form of the theorem, *i.e.* (λ, v) is an eigenpair of A if and only if (λ^2, v) is one for A^2. However, if we set $\lambda = 2$ and

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix},$$

then apparently $(\lambda^2 = 4, \begin{bmatrix} 0 & 1 \end{bmatrix}^T)$ is an eigenpair of

$$A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix};$$

however, $\lambda = 2$ is not an eigevalue of A, and thus the 'if' part of the stronger form of the theorem is false. We hence have to live with the more 'verbose' version of the theorem.

2.2

Proof:

If

$$Av = \lambda v \text{ or } Av = -\lambda v,$$

then

$$A^2 v = A(Av) = \pm\lambda Av = \lambda^2 v;$$

in other words, $Av = \lambda v$ or $Av = -\lambda v$ only if $A^2 v = \lambda^2 v$.

To prove the 'if' part, since A is symmetric, it can be decomposed as the following:

$$A = Q\Lambda Q^T$$

with Λ being a digonal matrix constructed from the eigenvalues of A and Q being orthogonal. We can also express A^2 in terms of Λ and Q:

$$A^2 = AA = (Q\Lambda Q^T)(Q\Lambda Q^T) = Q\Lambda^2 Q^T.$$

We recognise that the equation above happens to be the Schur decomposition of A^2, because Λ^2 is also diagonal. If λ^2 is one of the eigenvalues of A^2, then it must be one of the diagonal terms of Λ^2, say $\lambda^2 = (\Lambda^2)_{ii} = (\Lambda_{ii})^2$,

$$\Rightarrow \Lambda_{ii} = \pm\lambda.$$

This is to say that, if λ^2 is the eigenvalue of A^2 that corresponds to the eigenvector of q_i, then either λ or $-\lambda$ must also be the eigenvalue of A that corresponds to eigenvector q_i. Hence we finished the 'if' part of proof.

2.3

Since

$$\|A\|_2 = \sqrt{\max \sigma(A^T A)}$$

and

$$\lambda_A = \pm\sqrt{\lambda_{A^T A}},$$

the l_2-norm of a real symmetric matrix A is equal to the element in its spectrum with the largest absolute vaue, *i.e.*

$$\|A\|_2 = \max_{\lambda} |\lambda|\,,\ \ \lambda \in \sigma(A).$$

Apply this result to

$$A = \begin{bmatrix} -8 & 144 \\ 144 & -92 \end{bmatrix},$$

whose eigenvalues are

$$\lambda_1 = -200,\ \ \lambda_2 = 100,$$

we predict the l_2-norm of A to be

$$\|A\|_2 = \max\{|-200|\,, |100|\} = 200,$$

which is indeed the l_2-norm of A. (Can be verified with `MATLAB`.)

2.4

The theorem from preceeding parts does not hold for more general matrices; for example, given

$$A = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix},$$

$$\sigma(A) = \{-\sqrt{2},\ \sqrt{2}\};$$

following the previous theorem we would predict that

$$\|A\|_2 = \sqrt{2}.$$

However, if we calculate the l_2-norm in the canonical way, we would find that

$$\begin{aligned}
\sigma(A^T A) &= \{1,\, 4\}, \\
\|A\|_2 &= \sqrt{4} = 2 \\
&\neq \sqrt{2},
\end{aligned}$$

which contradicts the theorem.

2.5

Following the discussion above, we can say that the sigular values of a symmetric matrix are equal to the absolute values of its eigenvalues.

3

3.1

If A is a non-singular matrix and (λ, v) is one of its eigenpairs, *i.e.*

$$Av = \lambda v,$$

then

$$\begin{aligned}
A^{-1}Av &= (A^{-1}A)v = v \\
&= A^{-1}(Av) = \lambda A^{-1} v
\end{aligned}$$

$$\Rightarrow A^{-1}v = \frac{1}{\lambda} v.$$

In addition, A is also the inverse of A^{-1}, so $A^{-1}v = \frac{1}{\lambda}v$ will lead to $Av = \lambda v$ as well. Hence we proved that $Av = \lambda v$ if and only if $A^{-1}v = \frac{1}{\lambda}v$.

3.2

Given that A is non-singular,

$$\left\|A^{-1}\right\|_2 = \frac{1}{\min\{\sigma\}},\ \sigma \geqslant 0,\ \sigma^2 \in \sigma(A^T A).$$

Proof:

Given an SVD of A:

$$A = U\Sigma V^T,$$

then

$$A^{-1} = V\Sigma^{-1}U^T,$$

where $(\Sigma^{-1})_{ij} = 0$ if $i \neq j$ and $(\Sigma^{-1})_{ii} = \frac{1}{\Sigma_{ii}}$.

It can easily be verified that Σ^{-1} and A^{-1} expressed as such are exactly the inverses of Σ and A respectively, and thus $\{\sigma_i = \Sigma_{ii}\}$ and $\{\sigma_i^{-1} = \Sigma_{ii}^{-1}\}$ are the sets of sigular values of A and A^{-1}. The l_2-norm of A^{-1} can be expressed as the following:

$$\left\|A^{-1}\right\|_2 = \max_{\sigma_i^{-1}}\left|\sigma_i^{-1}\right| = \max_{\sigma_i}\frac{1}{\sigma_i} = \frac{1}{\min\{\sigma_i\}}.$$

Consider both $\{\sigma\}$ and $\{\sigma_i\}$ represent the set of singular values of A, we've proved the theorem.

4

$$Z = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 7 \\ 4 & 2 & 3 \\ 4 & 2 & 2 \end{bmatrix}$$

Since Z is a 5×3 rectangular matrix, it can be QR-decomposed into the product of a 5×5 orthogonal matrix Q and a 5×3 upper triangular matrix R:

$$Z_{5\times 3} = Q_{5\times 5}R_{5\times 3}.$$

The l_2-norm (abbreviated as simply norm below) of the first column of Z, denoted as x_1, is

$$\|x_1\|_2 = \sqrt{1^2 + 4^2 + 7^2 + 4^2 + 4^2} = \sqrt{98} = 7\sqrt{2};$$

thus

$$y_1 = \begin{bmatrix} 7\sqrt{2} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\|x_1 - y_1\| = \sqrt{196 - 14\sqrt{2}} \approx 13.2741,$$

and

$$w_1 = \frac{x_1 - y_1}{\|x_1 - y_1\|} = \begin{bmatrix} \frac{1-7\sqrt{2}}{\sqrt{196-14\sqrt{2}}} \\ \frac{4}{\sqrt{196-14\sqrt{2}}} \\ \frac{7}{\sqrt{196-14\sqrt{2}}} \\ \frac{4}{\sqrt{196-14\sqrt{2}}} \\ \frac{4}{\sqrt{196-14\sqrt{2}}} \end{bmatrix} \approx \begin{bmatrix} -0.6704 \\ 0.3013 \\ 0.5273 \\ 0.3013 \\ 0.3013 \end{bmatrix}.$$

We can construct the first Householder matrix as the following:

$$\begin{aligned} H_1 &= I - 2ww^T \\ &\approx \begin{bmatrix} 0.1010 & 0.4041 & 0.7071 & 0.4041 & 0.4041 \\ 0.4041 & 0.8184 & -0.3178 & -0.1816 & -0.1816 \\ 0.7071 & -0.3178 & 0.4438 & -0.3178 & -0.3178 \\ 0.4041 & -0.1816 & -0.3178 & 0.8184 & -0.1816 \\ 0.4041 & -0.1816 & -0.3178 & -0.1816 & 0,8184 \end{bmatrix} \end{aligned}$$

Multiply Z by H_1 on the left, we have:

$$H_1 Z \approx \begin{bmatrix} 9.8995 & 9.4954 & 9.6975 \\ 0 & 1.6311 & 2.9897 \\ 0 & 2.1044 & 1.7320 \\ 0 & -1.3689 & -0.0103 \\ 0 & -1.3689 & -1.0103 \end{bmatrix}$$

Following this strategy we have:

$$x_2 \approx \begin{bmatrix} 9.4954 \\ 1.6311 \\ 2.1044 \\ -1.3689 \\ -1.3689 \end{bmatrix}, \ y_2 \approx \begin{bmatrix} 9.4954 \\ 3.2919 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ w_2 \approx \begin{bmatrix} 0 \\ -0.5023 \\ 0.6364 \\ -0.4140 \\ -0.4140 \end{bmatrix},$$

$$\begin{aligned} H_2 &= I - 2w_2w_2^T \\ &\approx \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.4955 & 0.6393 & -0.4158 & -0.4158 \\ 0 & 0.6393 & 0.1900 & 0.5269 & 0.5269 \\ 0 & -0.4158 & 0.5269 & 0.6572 & -0.3428 \\ 0 & -0.4158 & 0.5269 & -0.3428 & 0.6572 \end{bmatrix}, \\ H_2H_1Z &\approx \begin{bmatrix} 9.8995 & 9.4952 & 9.6975 \\ 0 & 3.2919 & 3.0129 \\ 0 & 0 & 1.7026 \\ 0 & 0 & 0.0089 \\ 0 & 0 & -0.9911 \end{bmatrix} \end{aligned}$$

and that

$$x_3 \approx \begin{bmatrix} 9.6975 \\ 3.0129 \\ 1.7026 \\ 0.0089 \\ -0.9911 \end{bmatrix}, \ y_3 \approx \begin{bmatrix} 9.6975 \\ 3.0129 \\ 1.9701 \\ 0 \\ 0 \end{bmatrix}, \ w_3 \approx \begin{bmatrix} 0 \\ 0 \\ -0.2606 \\ 0.0086 \\ -0.9654 \end{bmatrix},$$

$$H_3 = I - 2w_3w_3^T \approx \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.8642 & 0.0045 & -0.5031 \\ 0 & 0 & 0.0045 & 0.9999 & 0.0167 \\ 0 & 0 & -0.5031 & 0.0167 & -0.8641 \end{bmatrix},$$

$$R = H_3H_2H_1Z \approx \begin{bmatrix} 9.8995 & 9.4954 & 9.6975 \\ 0 & 3.2919 & 3.0129 \\ 0 & 0 & 1.9701 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\begin{aligned} Q &= (H_3H_2H_1)^T = H_1^TH_2^TH_3^T \\ &\approx \begin{bmatrix} 0.1010 & 0.3162 & 0.5420 & 0.3408 & -0.6928 \\ 0.4041 & 0.3534 & 0.5162 & -0.5730 & 0.3422 \\ 0.7071 & 0.3906 & -0.5248 & 0.2684 & 0.0028 \\ 0.0041 & -0.5580 & 0.3871 & 0.5006 & 0.3534 \\ 0.4041 & -0.5580 & -0.1204 & -0.4825 & -0.5273 \end{bmatrix}. \end{aligned}$$

And we find that $\|Z - QR\|_2 \sim 10^{-15}$, implying that $Z \approx QR$, just as expected.

5

5.1

Suppose $Y = HX = (I - 2ww^T)X = X - 2ww^TX$ is our target matrix to compute, and x_i and y_i are the ith column of matrices X and Y respectively, then

$$y_i = x_i - 2(w^Tx_i)w.$$

If we compute y_is in this fashion, then the computation of the inner product between w and x_i requires m multiplications and $m-1$ additions, incurring a cost of $O(m)$; multiplying w by $-2w^T x_i$ needs another m multiplications, and adding the resulting vector to x_i requires m additions. In general, calculating y_i needs $O(m) + O(m) + O(m) = O(m)$ elemental operations, and there are m of them to calculate; thus getting the whole matrix of Y following this strategy would be of $O(m^2)$ complexity.

5.2

The strategy we propose is that, given a linear system

$$Ax = b,$$

and assuming that A can be QR-factorized into

$$A = QR,$$

$$\begin{aligned} \Rightarrow Rx &= Q^T b = H_m H_{m-1} \dots H_1 b \\ &= (I - 2w_m w_m^T)(I - 2w_{m-1} w_{m-1}^T) \dots (I - 2w_1 w_1^T) b := \tilde{b}. \end{aligned}$$

Thus while we are QR-factorizing A with matrices $(I - 2w_i w_i^T)$, we multiply the vector on the right hand side by the transformation matrix as well, (this is also of $O(m^2)$ complexity, thus will not change the asymptotic behaviour of the algorithm;) in the kth step we only need to multiply $(m - k + 2)$ columns ($m - k + 1$ on the left hand side, 1 on the right hand side) by the transformation matrix; however, the comlexity of QR-factorization is still of $O(m^3)$ overall. At the end of the factorization, we should end up with

$$Rx = \tilde{b},$$

the solution of which is also the solution of the original problem. Since R is upper triangular, we can solve for each component of x bottom up, which add a complexity of $O(m^2)$. In general, we expect the algorithm to solve the linear system within $O(m^3)$ time.

5.3

We generated two 5×5 and two 5×1 matrices with random entries as our test As and bs; by calculating the l_2-norm of $Ax - b$, we verified our algorithm to be correct. (Please refer to the attached diary.)

5.4

As stated in 5.2, we expect our algorithm to have the time complexity of $O(m^3)$; to verify this, we tested our script with random matrices of $2\times$ all the way up to 21×21; by fitting the corresponding number of operations used against the dimensionality, we find that the leading term of complexity is $2m^3$, just as expected. If we use quartic model (`'poly4'` in `MATLAB` terminology), then the coefficient of the m^4 term will be negligible.

A MATLAB code that verifies results in Problem 1

The script that verifies our result from Problem 1 is the following:

```
A = [2, -2; 1, 1];
rho = max(abs(eig(A)))
norm_1 = norm(A, 1)
norm_2 = norm(A, 2)
norm_inf = norm(A, Inf)
[U, Sigma, V] = svd(A)
```

And the output is:

```
rho =

2

norm_1 =

3

norm_2 =

2.8284

norm_inf =

4


U =

-1.0000   -0.0000
-0.0000    1.0000


Sigma =

2.8284         0
0    1.4142
```

```
V =

-0.7071    0.7071
0.7071    0.7071
```

Note that the numerical values of ρ and the norms corroborate the results we got in Problem 1. The result from `MATLAB svd` function minorly differs from our result in Problem 1; however, since both the singular values and the corresponding left/right singular vectors are indentical, our result in Problem 1 must be correct as well.