

OSM Lab 2017: Math Pset 3

Wei Han Chia

Due: 10 July 2017

Problems from the Book

4.2

Let $V = \text{span}(\{1, x, x^2\})$ be the subspace of the inner product space $L^2([0, 1; \mathbb{R}])$. Consider the derivative operator D . We showed in a previous problem set that

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Now we are interested in finding all the eigenvalues and eigenspaces of D . We note that the characteristic polynomial of D is given by $\det(zI - D) = z^3$. It follows that all the eigenvalues of D are 0.

Now we note that the algebraic multiplicity of 0 is 3, but the span of the eigenspace of 0 is 1, given by $\text{span}(1)$.

4.4

Note that $\det[A^H - zI] = \overline{\det[A - \bar{z}I]}$. This implies that the eigenvalues of the Hermitian of a matrix are the conjugates of the eigenvalues of the original matrix.

(i) Let us consider A an Hermitian 2x2 matrix. From 4.3, note that $\det[A - \lambda I] = \lambda^2 - \text{tr}(A)\lambda + \det(A) = \lambda^2 - \text{tr}(A^H)\lambda + \det(A^H)$. Combining this and above, we know that $\lambda = \bar{\lambda}$, implying that all eigenvalues are real.

(ii) Let us consider A a skew-Hermitian 2x2 matrix. From 4.3, note that $\det[A^H - \lambda I] = \lambda^2 - \text{tr}(A)\lambda - \det(A)$. Comparing the solutions to the quadratic formula, and using the equality above, we see that $\bar{\lambda} = -\lambda$, i.e. that all eigenvalues are imaginary.

4.6

We know that the determinant of an upper triangular matrix is its diagonal entries. We can prove this by the definition of the expansion theorem of determinants and induction.

Now it follows that for any $n \times n$ upper triangular matrix (A) , $p_A(z) = \det(zI - A) = \pi_{i=1}^n (z - a_{ii})$. So solving the characteristic polynomial yields the eigenvalues of A as $\{a_{ii}\}_{i=1}^n$.

4.8

Let V be the span of the set $S = \{\sin(x), \cos(x), \sin(2x), \cos(2x)\}$.

(i) In the previous problem set, we proved that each element of S is pairwise orthogonal. This implies that they are a linearly independent set. To see this, consider constants such that $a_1 \sin(x) + a_2 \cos(x) + a_3 \sin(2x) + a_4 \cos(2x) = 0$

$$\begin{aligned} &\text{Taking inner product with respect to any } s \in S \\ &\implies a_1 = a_2 = a_3 = a_4 = 0 \end{aligned}$$

Now since V is the span of this linearly independent S , it follows that S is a basis of V .

(ii) Now let D be the derivative operator. We can construct D simply by considering the derivative of each of the members of S .

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

(iii) We want to find some D -invariant subspace of S , i.e. some $W \subset (S)$ s.t. $DW \subset w$. We note that the derivative of $\sin(x)$ and $\cos(x)$ will yield a function of $\sin(x)$ and $\cos(x)$, and so our two complementary D -invariant subspaces are $\text{span}\{\sin(x), \cos(x)\}$ and $\text{span}\{\sin(2x), \cos(2x)\}$.

4.13

Let

$$A = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}$$

We want to find the transition matrix P such that $P^{-1}AP$ is diagonal. We know that A has two distinct eigenvectors, $x_1 = [1, -1]$ and $x_2 = [2, 1]$, corresponding to eigenvalues of 0.4 and 1 respectively. From our diagonalization theorem, know that we can obtain the transition matrix P by letting $P = [x_1, x_2]$, i.e.

$$P = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$$

4.15

We will prove that if $(\lambda_i)_{i=1}^n$ are the eigenvalues of a semi-simple matrix A , and $f(x) = a_0 + a_1x + \dots + a_nx$, then $f(\lambda_i)_{i=1}^n$ are the eigenvalues of $f(A)$.

Proof. Since A is semi-simple, it is diagonalizable. Lets write $A = PDP^{-1}$. Now $A^k = PD^kP^{-1}$.

$$f(A) = P(a_0I + a_1D + \dots + a_nD^n)P^{-1}$$

Now since $f(A)$ is diagonalizable, we know that it is semi-simple, and that its eigenvectors are given by the columns of P , and its eigenvalues given by the diagonal entries of D . So, the eigenvalues of $f(A)$ are $a_0 + a_1\lambda + a_2\lambda^2 + \dots + a_n\lambda^n$ for all λ eigenvalues of A . So it follows that $(f(\lambda_i))_{i=1}^n$ are the eigenvalues of $f(A)$. \square

4.16

Let A be the same matrix as in 4.13.

(i) We want to compute $\lim_{n \rightarrow \infty} A^n$. With respect to the matrix 1-norm (max absolute column sum). Note that $A^n = PD^nP^{-1}$. Now $\|A^n\|_1 = \|P\| \|D^n\| \|P^{-1}\|$.

$$\|D^n\| = \begin{bmatrix} 0.4^n & 0 \\ 0 & 1 \end{bmatrix}$$

Now we can simply let $B = PCP^{-1}$, where $C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Now note that

$$\|A^k - B\| = X0.4^k$$

Where $X = \|P\| \|P^{-1}\|$. It is easy to see then that for any $\epsilon > 0$, there exists an N such that $X0.4^k < \epsilon$ for all $k > N$.

(ii) We note that the ∞ -norm and the Frobenius norm holds in the exact same way, sine we are only interested in the norm of the difference between B and A^k , which is a matrix with only 1 entry.

(iii) We can use Theorem 4.3.12 to note that the eigenvalues of $3I + 5A + A^3$ are $3 + 5 + 1 = 9$, and $3 + 2 + 0.064 = 5.064$.

4.18

If λ is an eigenvalue of $A \in M_n(\mathbb{F})$, then it is also an eigenvalue of A^T (this follows because the determinant is invariant to transpose). Now consider $\mathbf{x} \neq 0$ the eigenvector of A^T corresponding to λ .

$$A^T \mathbf{x} = \lambda \mathbf{x}$$

$$\text{Taking transpose } \mathbf{x}^T A = \lambda \mathbf{x}^T$$

4.20

Let A be Hermitian and orthonormally similar to B . We will show that B is also Hermitian, i.e. $B^H = B$.

$$\begin{aligned} B &= U^H A U \\ B^H &= U^H A^H U \\ &= U^H A U = B \end{aligned}$$

4.24

Given $A \in M_n(\mathbb{C})$, we define the Rayleigh quotient as

$$\rho(\mathbf{x}) = \frac{\langle \mathbf{x}, A\mathbf{x} \rangle}{\|\mathbf{x}\|^2}$$

We will show that the Rayleigh quotient is bounded by the eigenvalues of A , and so it will follow from 4.4 that it can only take on real values for Hermitian matrices and imaginary values for skew-Hermitian matrices.

First, note that both Hermitian and skew-Hermitian matrices are diagonalizable. This follows from Shur's lemma, and noting that an upper triangular matrix that is symmetric or skew-symmetric is diagonalizable. Now we know that these matrices are semi-simple, i.e. that the diagonal matrix D they are similar to is the matrix of its eigenvalues.

$$\begin{aligned} \rho(\mathbf{x}) &= \frac{\mathbf{x}^H A \mathbf{x}}{\mathbf{x}^H \mathbf{x}} = \frac{\mathbf{x}^H V^H D V \mathbf{x}}{\mathbf{x}^H V^H V \mathbf{x}} \\ &= \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \quad \text{Where } V \mathbf{x} = \sum_{i=1}^n y_i \mathbf{e}_i \end{aligned}$$

Clearly then $\rho(\mathbf{x})$ is a weighted average of eigenvalues of A , which allows us to conclude that in the case of Hermitian matrices, the Rayleigh quotient is always real, and in the case of skew-Hermitian matrices, it is always imaginary.

4.25

Let $A \in M_n(\mathbb{C})$ be a normal matrix with eigenvalues $(\lambda_1, \dots, \lambda_n)$ and corresponding orthonormal eigenvectors $[\mathbf{x}_1, \dots, \mathbf{x}_n]$.

(i) We will show that the identity matrix can be written $I = \mathbf{x}_1 \mathbf{x}_1^H + \dots + \mathbf{x}_n \mathbf{x}_n^H$. Consider any $\mathbf{v} \in \mathbb{C}^n$. We can represent \mathbf{v} as a $\sum_{i=1}^n \alpha_i \mathbf{x}_i$. Now

$$\begin{aligned} I \mathbf{v} &= \mathbf{v} \\ (\mathbf{x}_1 \mathbf{x}_1^H + \dots + \mathbf{x}_n \mathbf{x}_n^H) \mathbf{v} &= (\mathbf{x}_1 \mathbf{x}_1^H + \dots + \mathbf{x}_n \mathbf{x}_n^H) \sum_{i=1}^n \alpha_i \mathbf{x}_i \\ &= \sum_{i=1}^n \alpha_i \mathbf{x}_i = \mathbf{v} \end{aligned}$$

Where the last result follows using the usual inner product and the fact that each \mathbf{x} are orthonormal.

(ii) Now similarly, consider A .

$$\begin{aligned} A \mathbf{v} &= \sum_{i=1}^n \alpha_i A \mathbf{x}_i = \sum_{i=1}^n \alpha_i \lambda_i \mathbf{x}_i \\ (\lambda_1 \mathbf{x}_1 \mathbf{x}_1^H + \dots + \lambda_n \mathbf{x}_n \mathbf{x}_n^H) \mathbf{v} &= (\lambda_1 \mathbf{x}_1 \mathbf{x}_1^H + \dots + \lambda_n \mathbf{x}_n \mathbf{x}_n^H) \sum_{i=1}^n \alpha_i \mathbf{x}_i \\ &= \sum_{i=1}^n \alpha_i \lambda_i \mathbf{x}_i \end{aligned}$$

Therefore it similarly follows that A can be written as $\lambda_1 \mathbf{x}_1 \mathbf{x}_1^H + \dots + \lambda_n \mathbf{x}_n \mathbf{x}_n^H$.

4.27

Let $A \in M_n(\mathbb{F})$. From remark 4.5.2, we know that $x^T A x$ is always real valued and positive.

Now simply consider $e_i = [0, \dots, 0, 1, 0, \dots, 0]$ the standard basis.

$$e_i^T A e_i = a_{ii} > 0$$

Applying this for all i will show that every diagonal entry of A is real valued and positive.

4.28

Assume $A, B \in M_n(\mathbb{F})$ are positive semi-definite. Since B is positive semidefinite, we can diagonalize $B = U\Sigma U^T$, where Σ is diagonal. Now note that we can define $B^{1/2} = U\Sigma^{1/2}U^T$ such that $B^{1/2}B^{1/2} = B$. Now

$$\begin{aligned} \text{tr}(AB) &= \text{tr}(AB^{1/2}B^{1/2}) = \text{tr}(B^{1/2}AB^{1/2}) \\ B^{1/2}AB^{1/2} &\text{ is positive semi definite ince } A \text{ is psd so by 4.27 it has real positive diagonal entries} \\ \implies \text{tr}(AB) &\geq 0 \end{aligned}$$

Now for any diagonal matrix D and A with positive real entries on the diagonal, $\text{tr}(AD) \leq \text{tr}(A)\text{tr}(D)$. This follows by noting that $\text{tr}(AD) = \sum_{i=1}^n a_{ii}d_{ii} < (\sum_{i=1}^n a_{ii})(\sum_{j=1}^n d_{jj})$ So,

$$\begin{aligned} \text{tr}(AB) &= \text{tr}(AU\Sigma U^T) = \text{tr}(U^T AU\Sigma U^T) \\ &\leq \text{tr}(U^T AU)\text{tr}(\Sigma) \\ &\leq \text{tr}(A)\text{tr}(B) \end{aligned}$$

The only relevant condition we really need to check to prove that $\|\cdot\|_F$ is a matrix norm is the triangle inequality. The others follow from the linearity and scalar multiplicativity of the trace.

$$\begin{aligned} \|A + B\|_F^2 &= \text{tr}((A + B)^T(A + B)) = \text{tr}(A^T A) + \text{tr}(B^T B) + 2\text{tr}(A^T B) \\ &\leq \text{tr}(A^T A) + \text{tr}(B^T B) \quad \text{Since } \text{tr}(A^T B) > 0 \text{ from above} \\ &= \|A\|_F^2 + \|B\|_F^2 \end{aligned}$$

Now taking squareroots on both sides and using our usual inequality on real numbers yields the desired triangle inequality.

4.31

Assume $A \in M_{m \times n}(\mathbb{F})$, and A is not identically zero.

(i) Prove that $\|A\|_2 = \sigma_1$, where σ_1 is the largest singular value of A .

Proof.

$$\begin{aligned} \|A\|_2 &= \sup_{\|\mathbf{x}\|_2} \|A\mathbf{x}\|_2 \\ &= \sup \sqrt{\langle A\mathbf{x}, A\mathbf{x} \rangle} \\ &= \sup \sqrt{\langle \mathbf{x}, A^H A \mathbf{x} \rangle} \end{aligned}$$

Note that the singular values of A are precisely the eigenvalues of $A^H A$

Since $A^H A$ is hermitian, we can diagonalize it to get

$$\begin{aligned} &= \max_{1 \leq i \leq n} \sigma_i \|\mathbf{x}\|_2 \\ &= \sigma_1 \end{aligned}$$

□

(ii) Prove that if A is invertible, then $\|A^{-1}\|_2 = \sigma_n^{-1}$.

Proof. We proceed as above substituting A for A^{-1} , and noting that the largest singular value of A^{-1} is precisely σ_n^{-1} . We can see this by noting that the eigenvalues of an inverse matrix is precisely the inverse of its eigenvalues, and that the singular values of A are given by the eigenvalues of $A^H A$. □

(iii) We will first prove that $\|A^H\|_2^2 = \|A^T\|_2^2 = \|A\|_2^2$. Recall that by SVD, we can represent $A = U\Sigma V^H$ where U and V are orthonormal.

$$\begin{aligned}
\|A\|_2^2 &= \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|_2^2 \\
&= \sup_{\|\mathbf{x}\|=1} \|U\Sigma V^H \mathbf{x}\|_2^2 \\
&= \sup_{\|\mathbf{x}\|=1} \|\Sigma \mathbf{x}\|_2^2 \quad \text{See part iv} \\
&= \sup_{\|\mathbf{x}\|=1} \|\Sigma^T \mathbf{x}\|_2^2 = \|A^T\|_2^2 \\
&= \sup_{\|\mathbf{x}\|=1} \|\Sigma^H \mathbf{x}\|_2^2 = \|A^H\|_2^2
\end{aligned}$$

Next, we note the following:

$$\begin{aligned}
\|A^H A\|_2 &= \|U\Sigma V^H V \Sigma U^H\|_2 \\
&= \|U\Sigma^2 U^H\|_2 \\
&= \sup_{\|\mathbf{x}\|=1} \|\Sigma^2 \mathbf{x}\|_2 \\
&= \|A\|_2^2
\end{aligned}$$

(iv)

$$\begin{aligned}
\|UAV\|_2 &= \sqrt{\sup \langle UAV\mathbf{x}, UAV\mathbf{x} \rangle} \\
&= \sqrt{\sup \langle AV\mathbf{x}, AV\mathbf{x} \rangle} \quad \text{Using the adjoint and the fact that } U \text{ is orthonormal} \\
&= \sqrt{\sup \langle A\mathbf{x}, A\mathbf{x} \rangle} \quad \text{Using the fact that } V \text{ is orthonormal} \\
&= \|A\|_2
\end{aligned}$$

4.32

Assume $A \in M_{m \times n}(\mathbb{F})$ is of rank r .

(i) Prove that if $U \in M_m(\mathbb{F})$ and $V \in M_n(\mathbb{F})$ are orthonormal, then $\|UAV\|_F = \|A\|_F$.

Proof.

$$\begin{aligned}
\|UAV\|_F &= \sqrt{\text{tr}(V^H A^H U^H U A V)} \\
&= \sqrt{\text{tr}(A^H A)} = \|A\|_F
\end{aligned}$$

□

(ii) Prove that $\|A\|_F = (\sigma_1^2 + \dots + \sigma_r^2)^{1/2}$, where σ_i are the singular values of A .

Proof. Using SVD, we can write $A = U\Sigma V^H$ where U and V are orthonormal. From (i), we know that $\|U\Sigma V^H\|_F = \|\Sigma\|_F$.

$$\|\Sigma\|_F = \sqrt{\text{tr}(\Sigma^H \Sigma)} = \sqrt{\sigma_1^2 + \dots + \sigma_r^2}$$

□

4.33

From 4.31(i), we know that $\|A\|_2 = \sigma_1$. Now all we need to prove is that $\sup_{\|\mathbf{x}\|_2=1, \|\mathbf{y}\|_2=1} |\mathbf{y}^H A \mathbf{x}|$.

To see that σ_1 is an upper bound for $\sup_{\|\mathbf{x}\|_2=1, \|\mathbf{y}\|_2=1} |\mathbf{y}^H A \mathbf{x}|$

$$\begin{aligned}
\sup_{\|\mathbf{y}\|_2=\|\mathbf{x}\|_2=1} \mathbf{y}^H A \mathbf{x} &= \mathbf{y}^H U^H \Sigma V \mathbf{x} \\
&= \sup_{\|\mathbf{a}\|_2=\|\mathbf{b}\|_2=1} \mathbf{a}^H \Sigma \mathbf{b} \quad \text{Because } U \text{ and } V \text{ are orthogonal} \\
&= \sup \frac{\sum_{i=1}^n \sigma_i a_i b_i}{\sum_{i=1}^n a_i b_i}
\end{aligned}$$

It is easy to see therefore that the RHS is a weighted average of singular values, i.e. that σ_1 is an upper bound, and is indeed the least upper bound for the RHS.

4.36

Consider the matrix

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

A has eigenvalues $1, -1$, but has singular values $0.5 * (1 + \sqrt{5}), |0.5 * (1 - \sqrt{5})|$.

4.38

Proving proposition 4.6.2 is simply substituting in our definition of the Moore-Penrose pseudoinverse and the compact SVD of A .

(i)

$$AA^\dagger A = U_1 \Sigma_1 V_1^H V_1 \Sigma_1^{-1} U_1^H U_1 \Sigma_1 V_1^H = A$$

(ii)

$$A^\dagger AA^\dagger = V_1 \Sigma_1^{-1} U_1^H U_1 \Sigma_1 V_1^H V_1 \Sigma_1^{-1} U_1^H = A^\dagger$$

(iii)

$$(AA^\dagger)^H = I = AA^\dagger$$

(iv)

$$(A^\dagger A)^H = I = A^\dagger A$$

(v) Note that AA^\dagger is a map onto $R(A)$. Now from (i), $AA^\dagger AA^\dagger = AA^\dagger$, so AA^\dagger is a projection. Finally, from (iii), we can conclude that this projection is orthogonal.

(vi) Similarly, note that $A^\dagger A$ is a map onto $R(A^H)$. From (ii), $A^\dagger AA^\dagger A = A^\dagger A$, so $A^\dagger A$ is a projection. Finally, from (iv), we can conclude that this projection is orthogonal.