Problem Set #3, Spectral Theory

OSM Lab: Math
Rebekah Dix

Exercise 4.2

Observe that we can write any element of $L^2([0,1];\mathbb{R})$ as,$p(x) = a_0 + a_1 x + a_2 x^2$, and thus can represent $p(x)$ by the vector $[a_0, a_1, a_2]$. Then, the derivative operator can be written as,

$$D[p](x) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = a_1 + 2a_2 x \tag{1}$$

Then, the characteristic polynomial of $D[p](x)$ is given by,

$$det(\lambda I - D[p](x)) = \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -2 \\ 0 & 0 & \lambda \end{vmatrix} = \lambda^3 \tag{2}$$

Therefore, $D[p](x)$ has one eigenvalue $\lambda = 0$, which has an algebraic multiplicity of 3. To find the eigenspace, observe that an eigenvalue must have $a_1 = a_2 = 0$, but a_0 is free to be anything. Therefore, the eigenspace is simply the span of $[1, 0, 0]$. This has a geometric multiplicity of 1.

Exercise 4.4

Claim 1. An Hermitian 2x2 matrix has only real eigenvalues.

Proof. Let $A \in M_2(\mathbb{F})$ be an Hermitian 2x2 matrix. By definition, $A^H = A$. This leads to a restriction on the entries of A. By inspection, we find that A must take the form,

$$A = \begin{bmatrix} a & c + di \\ c - di & b \end{bmatrix} \tag{3}$$

where $a, b, c, d \in \mathbb{R}$. That is, the diagonal elements of A must be real, and the off-diagonal elements of A must be the complex conjugate of each other. Next, recall that the characteristic polynomial of any 2x2 matrix as the form $p(\lambda) = \lambda^2 - tr(A)\lambda + det(A)$. Thus, the characteristic polynomial of A is,

$$p(\lambda) = \lambda^2 - (a + b)\lambda + (ab - (c^2 + d^2)) \tag{4}$$

Equation (4) is a quadratic equation in λ, and therefore has real roots if and only if the discriminant is positive. That is, if $(a+b)^2 - 4(ab - (c^2 - d^2))$ is positive. However,

$$(a + b)^2 - 4(ab - (c^2 - d^2)) = (a - b)^2 + 4(c^2 + d^2) \tag{5}$$

Observe that the expression in (5) is a sum of squares and therefore always positive. Thus, the characteristic polynomial of A has real roots. Therefore, A has real eigenvalues. Our choice of A represented an arbitrary Hermitian 2x2 matrix, so that an Hermitian 2x2 matrix has only real eigenvalues. $\square$

Claim 2. A skew-Hermitian 2x2 matrix only has imaginary eigenvalues.

Proof. Let $A \in M_2(\mathbb{F})$ be a skew-Hermitian 2x2 matrix. By definition, $A^H = -A$. This leads to a restriction on the entries of A. By inspection, we find that A must take the form,

$$A = \begin{bmatrix} ai & c+di \\ -c+di & bi \end{bmatrix} \tag{6}$$

where $a, b, c, d \in \mathbb{R}$. The characteristic polynomial of A is,

$$p(\lambda) = \lambda^2 - i(a+b)\lambda - ab + c^2 + d^2 \tag{7}$$

This equation is quadratic in λ, and therefore has imaginary roots if and only if the discriminant is negative. The discriminant of this quadratic equation is,

$$(-i(a+b))^2 - 4(ab + c^2 + d^2) = -(a-b)^2 - 4c^2 - 4d^2 \tag{8}$$

Observe that the above equation is a difference of squares, and therefore is weakly negative. Then, the characteristic polynomial will have imaginary roots so that all eigenvalues of A are imaginary. Our choice of A represented an arbitrary skew-Hermitian 2x2 matrix, so that a skew-Hermitian 2x2 matrix has only imaginary eigenvalues. $\square$

Exercise 4.6

Claim 3. The diagonal entries of an upper-triangular (or lower-triangular matrix) are its eigenvalues.

Proof. Let $A \in M_n(\mathbb{F})$ be an upper-triangular matrix. Fix $\lambda \in \mathbb{C}$. Observe that $\lambda I - A$ is also an upper-triangular matrix. Recall that the determinant of an upper-triangular matrix is the product of the elements on the diagonal. Therefore, we find the characteristic polynomial of A is,

$$p(\lambda) = det(\lambda I - A) = \prod_{i=1}^{n} (\lambda - a_{ii}) \tag{9}$$

The roots of this equation are precisely the diagonal elements of A, so that the diagonal entries of an upper-triangular are its eigenvalues. The proof is analogous for a lower-diagonal matrix since the determinant of a lower-diagonal matrix is also the product of its diagonal elements. $\square$

Exercise 4.8

Part (i) Let V be the span of the set $S = \{\sin(x), \cos(x), \sin(2x), \cos(2x)\}$ in the vector space $C^\infty(\mathbb{R}; \mathbb{R})$.

Claim 4. S is a basis for V.

Proof. Clearly, by the definition of V, S spans V. Therefore, we must show that S is linearly independent. For the sake of contradiction, assume that S is linearly dependent. Then, there exist $a, b, c, d, \in \mathbb{R}$ not all zero such that $a\sin(x) + b\cos(x) + c\sin(2x) + d\cos(2x) = 0$ for all x. Now, let us consider special cases of this condition to pin down the values of the constants. First, suppose $x = 0$. Then, we find that $b + d = 0$. Next, suppose $x = \pi$. Then, we find that $-b + d = 0$. These two equations imply that $b = d = 0$. Now, consider $x = \frac{\pi}{2}$. This implies that $a = 0$. Next, suppose that $x = \frac{\pi}{4}$. This in turn implies that $d = 0$. Therefore, we have arrived at a contradiction, and it must bbe that S is indeed a linearly independent set. Therefore, S is a basis for V. $\square$

Part (ii) Let D be the derivative operator. Since S is a basis for V, we may write any element of V as a linearly combination of the elements of S. We order the space as follows: $v(x) = a\sin(x) + b\cos(x) + c\sin(2x) + d\cos(2x)$. Then,

$$D[v](x) = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = a\cos(x) - b\sin(x) + 2c\cos(2x) - 2d\sin(2x) \tag{10}$$

Part (iii) Consider the subspaces $U = \{\sin(x), \cos(x)\}$ and $V = \{\sin(2x), \cos(2x)\}$. We show that these are two complementary D-invariant subspaces in V. Consider $u(x) = a\sin(x) + b\cos(x)$. Then, $D[u](x) = a\cos(x) - b\sin(x) \in U$. Similarly, consider $v(x) = a\sin(2x) + b\cos(2x)$. Then, $D[v](x) = 2a\cos(2x) - 2b\sin(2x) \in V$. Therefore, U and V are D-invariant. Clearly, U and V are complementary.

Exercise 4.13

Let $A = \begin{bmatrix} .8 & .4 \\ .2 & .6 \end{bmatrix}$. The characteristic polynomial of A is $p(\lambda) = det(\lambda I - A) = (\lambda - 1)(\lambda - \frac{2}{5})$. Thus, the eigenvalues are $\lambda = 1, \frac{2}{5}$. By direct computation, the corresponding eigenvectors are $[2, 1]^T$ and $[-1, 1]^T$ respectively. Then, the matrix $P = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$ causes $P^{-1}AP$ to be diagonal. Indeed, it follows by direct calculation that,

$$D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{2}{5} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} .8 & .4 \\ .2 & .6 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = P^{-1}AP \tag{11}$$

Exercise 4.15

Claim 5. If $(\lambda_i)_{i=1}^n$ are the eigenvalues of a semisimple matrix $A \in M_n(\mathbb{F})$ and $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ is a polynomial, then $(f(\lambda_i))_{i=1}^n$ are the eigenvalues of $f(A) = a_0 I + a_1 A + \cdots + a_n A^n$.

Proof. Let A be a semisimple matrix, then A is diagonalizable. Thus, $A = PDP^{-1}$, where the columns of P are the eigenvectors corresponding to the eigenvalues of A

and D is a diagonal matrix whose diagonal elements are the eigenvalues of A. Then,

$$\begin{aligned} f(A) = f(PDP^{-1}) &= a_0 I + a_1 PDP^{-1} + \cdots + a_n (PDP^{-1})^n \\ &= a_0 I + a_1 PDP^{-1} + \cdots + a_n PD^n P^{-1} \\ &= P(a_0 I + a_1 D + \cdots + a_n D^n) P^{-1} \end{aligned}$$

Let $\hat{D} = a_0 I + a_1 D + \cdots + a_n D^n$. Thus, $f(A)$ is similar to $\hat{D}$, so they have the same eigenvalues. The eigenvalues of $\hat{D}$ are $(f(\lambda_i))_{i=1}^n$, the diagonal elements of $\hat{D}$. Thus, the eigenvalues of $f(A)$ are $(f(\lambda_i))_{i=1}^n$. $\square$

Exercise 4.16

Let $A = \begin{bmatrix} .8 & .4 \\ .2 & .6 \end{bmatrix}$. Let P and D be as in Exercise 4.13.
Part (i)

$$\begin{aligned} A^k = PD^k P^{-1} &= \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{2}{5}^k \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 + .4^k & 2 - 2 * .4^k \\ 1 - .4^k & 1 + 2 * .4^k \end{bmatrix} \end{aligned}$$

And,

$$\begin{aligned} \lim_{n\to\infty} A^n = \lim_{n\to\infty} PD^n P^{-1} &= P \left(\lim_{n\to\infty} D^n \right) P^{-1} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \left(\lim_{n\to\infty} \begin{bmatrix} 1 & 0 \\ 0 & \frac{2}{5} \end{bmatrix} \right) \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Then,

$$A^k - B = \begin{bmatrix} .4^k & -2 * .4^k \\ -.4^k & 2 * .4^k \end{bmatrix}$$

The 1-norm of a matrix is the largest column sum. Clearly, $A^k - B$ converges with respect to the 1-norm since the columns both sum to 0.
Part (ii) The ∞-norm of matrix is the largest row sum. However, clearly $-.4^k + 2 * .4^k \to 0$ as $k \to \infty$. Therefore, the same matrix B works. Next,

$$\begin{aligned} \left\| A^k - B \right\|_F &= \sqrt{tr \left(\begin{bmatrix} .4^k & -.4^k \\ -2 * .4^k & 2 * .4^k \end{bmatrix} \begin{bmatrix} .4^k & -2 * .4^k \\ -.4^k & 2 * .4^k \end{bmatrix} \right)} \\ &= \sqrt{tr \left(\begin{bmatrix} 2 * .4^{2k} & -4 * .4^{2k} \\ -4 * .4^{2k} & 8 * .4^{2k} \end{bmatrix} \right)} \\ &= \sqrt{10 * .4^{2k}} \end{aligned}$$

Clearly, $\sqrt{10 * .4^{2k}}$ goes to 0 as k goes to infinity. Therefore, $\left\|A^k - B\right\|_F \to 0$. Thus, convergence does not appear to depend on the choice of norm.
Part (iii) By Theorem 4.3.12, we know that $(f(\lambda_i))_{i=1}^2$ are the eigenvalues of $f(A) = 3I + 5A + A^3$, where $f(x) = 3 + 5x + x^3$. Therefore, the eigenvalues of $f(A)$ are $f(1) = 9$ and $f(.4) = 5.064$.

Exercise 4.18

Claim 6. If λ is an eigenvalue of $A \in M_n(\mathbb{F}^n)$, then there exists a nonzero row vector x^T such that $x^T A = \lambda x^T$.

Proof. Let $A \in M_n(\mathbb{F}^n)$. First note that A and A^T have the same eigenvalues. Indeed, A and A^T have the same characteristic polynomial:

$$\begin{aligned} p_A(z) &= det(zI - A) \\ &= det((zI - A)^T) \\ &= det(zI - A^T) \\ &= p_{A^T}(z) \end{aligned}$$

Therefore, if λ is an eigenvalue of A, then λ is an eigenvalue of A^T. Thus, suppose λ is an eigenvalue of A (and hence A^T). Then there exists some nonzero $y \in \mathbb{F}^n$ such that $A^T y = \lambda y$. Taking the transpose of both sides of this equation, observe that $y^T A = \lambda y^T$. Thus, there exists a nonzero row vector y^T such that $y^T A = \lambda y^T$. □

Exercise 4.20

Claim 7. If A is Hermitian and orthonormally similar to B, then B is also Hermitian.

Proof. Let $A, B \in M_n(\mathbb{F}^n)$, A be Hermitian, and A be orthonormally similar to B. Therefore, $A = A^H$ there exists an orthonormal matrix U such that $B = U^H A U$. Consider taking the transpose of each side of this equation. We see that $B^H = (U^H A U)^H = U^H A^H U = U^H A U = B$. Therefore, $B = B^H$, so that B is Hermitian. □

Exercise 4.24

Claim 8. Given $A \in M_n(\mathbb{C}^n)$, define the Rayleigh quotient as

$$\rho(x) = \frac{\langle x, Ax \rangle}{\|x\|^2} \tag{12}$$

The Rayleigh quotient can only take on real values for Hermitian matrices and only imaginary values for skew-Hermitian matrices.

Proof. First, suppose that $A \in M_n(\mathbb{C}^n)$ is Hermitian, so that $A^H = A$. Fix $x \in \mathbb{C}^n$. Then,

$$\langle x, Ax \rangle = x^H A x = x^H A^H x = (Ax)^H x = \langle Ax, x \rangle = \overline{\langle x, Ax \rangle} \tag{13}$$

Where the second inequality follows because A is Hermitian and the final inequality follows by the conjugate symmetry of the inner product. Therefore, we have that $\langle x, Ax\rangle = \overline{\langle x, Ax\rangle}$. Therefore, $\langle x, Ax\rangle$ must be a real number as it equals its complex conjugate. Also observe that $\|x\|^2$ is by definition a real number. Therefore, $\rho(x) = \frac{\langle x, Ax\rangle}{\|x\|^2}$ can only take on real values when A is Hermitian.

Next, suppose $A \in M_n(\mathbb{C}^n)$ is skew-Hermitian, so that $A^H = -A$. Fix $x \in \mathbb{C}^n$. Then,

$$\langle x, Ax\rangle = x^H Ax = (A^H x)^H x = (-Ax)^H x = -(Ax)^H x = -\langle Ax, x\rangle = -\overline{\langle x, Ax\rangle} \tag{14}$$

Therefore, we have that $\langle x, Ax\rangle = -\overline{\langle x, Ax\rangle}$. However, for this equality to hold, it must be that the real part of this number is 0, or that $\langle x, Ax\rangle$ is imaginary. To see this more clearly, suppose $\langle x, Ax\rangle = a + bi$ where $a, b \in \mathbb{R}$. Now, $\langle x, Ax\rangle = -\overline{\langle x, Ax\rangle}$ implies that $a + bi = -a + bi$. Matching up the real and imaginary parts implies that $a = -a$, or that $a = 0$. Thus, $\langle x, Ax\rangle$ is imaginary; and therefore, $\rho(x) = \frac{\langle x, Ax\rangle}{\|x\|^2}$ can only take on real values when A is skew-Hermitian. □

Exercise 4.25

Let $A \in M_n(\mathbb{C}^n)$ be a normal matrix with eigenvalues $(\lambda_1, \ldots, \lambda_n)$ and corresponding orthonormal eigenvectors $[x_1, \ldots, x_n]$.

Claim 9. The identity matrix can be written as $I = x_1 x_1^H + \cdots + x_n x_n^H$.

Proof. Let $x_j \in [x_1, \ldots, x_n]$. Observe that $(x_1 x_1^H + \cdots + x_n x_n^H)x_j = x_j x_j^H x_j = x_j = Ix_j$. The second and third equalities follow because $[x_1, \ldots, x_n]$ is an orthonormal set (so $x_i^H x_j = 0$ for all $i \neq j$ and $x_j^H x_j = 1$). Thus, it follows by the final equality that $x_1 x_1^H + \cdots + x_n x_n^H = I$. □

Claim 10. A can be written as $A = \lambda_1 x_1 x_1^H + \cdots + \lambda_n x_n x_n^H$.

Proof.

$$\begin{aligned}
A &= AI \\
&= A(x_1 x_1^H + \cdots + x_n x_n^H) && \text{(by the above claim)} \\
&= Ax_1 x_1^H + \cdots + Ax_n x_n^H \\
&= \lambda_1 x_1 x_1^H + \cdots + \lambda_n x_n x_n^H && \text{(because } Ax_i = \lambda_i x_i\text{)}
\end{aligned}$$

□

Exercise 4.27

Claim 11. If $A \in M_n(\mathbb{F}^n)$ is positive definite, then all its diagonal entries are real and positive.

Proof. Let $A \in M_n(\mathbb{F}^n)$ be positive definite. By definition, A is Hermitian. Note that all Hermitian matrices must have real elements on the diagonal. This follows because if A is Herimitian, then entry $a_{ii} = \overline{a_{ii}}$, which implies that a_{ii} is real. Thus, positive definite matrices have real diagonal entries.

Now, consider the canonical basis vector e_i (i.e. an $n \times 1$ vector of zeros with a 1 in position i). Let A be written in terms of its columns as $A = [a_1, a_2, \ldots, a_n]$. Since A is positive definite, we have that,

$$0 < \langle e_i, Ae_i \rangle = e_i^T A e_i = e_i^T a_i = a_{ii} \tag{15}$$

Therefore, $a_{ii} > 0$ so that the diagonal elements of A are both real and positive. □

Exercise 4.28

Claim 12. Let $A, B \in M_n(\mathbb{F})$ be positive semidefinite. Then,

$$0 \leq tr(AB) \leq tr(A)tr(B) \tag{16}$$

Then, $\|\cdot\|_F$ is a matrix norm.

Proof. Since A and B are positive semidefinite matrices, there exist matrices S and T such that $A = S^H S$ and $B = T^H T$. Thus,

$$tr(AB) = tr(S^H S T^H T) = tr(TS^H (TS^H)^H) = tr((TS^H)^H TS^H). \tag{17}$$

Then, $(TS^H)^H TS^H$ is a positive semidefinite. By Exercise 4.26, we know that positive semidefinite matrices have nonnegative diagonal elements, so the trace is weakly positive. Therefore, $tr(AB) = tr((TS^H)^H TS^H) \geq 0$. Next, we diagonalize A and B as $A = PDP^{-1}$ and $B = QEQ^{-1}$. Then,

$$\begin{aligned}
tr(AB) &= tr(PDP^{-1}QEQ^{-1}) \\
&= tr(PP^{-1}QDEQ^{-1}) \\
&= tr(QQ^{-1}DE) \\
&= tr(DE) \\
&= \sum_i \lambda_i \mu_i && \text{(where } \lambda_i \text{ and } \mu_i \text{ are the eigenvalues of } A \text{ and } B) \\
&\leq \sum_i \lambda_i \sum_i \mu_i \\
&= tr(A)tr(B)
\end{aligned}$$

We now verify the properties required of a matrix norm.

(Positivity) Note that $\|A\|_F = tr(A^H A)$. $A^H A$ is a positive semidefinite matrix, so that is diagonal elements are weakly positive, which implies its trace is weakly positive. Thus, $\|A\|_F \geq 0$. Conversely, suppose $\|A\|_F = 0$. Since the diagonal entries of $A^H A$ are weakly positive, it must be that the are all 0 for $\|A\|_F = 0$. This in turn implies that the singular values of A are all 0. Then, it must be that A is the zero matrix.

(Scale preservation) Fix $\alpha \in \mathbb{R}$. Observe that,

$$\|\alpha A\|_F = \sqrt{tr((\alpha A)^H(\alpha A)} = \sqrt{\alpha^2 tr(A^H A)} = \alpha\sqrt{tr(A^H A)} = \alpha \|A\|_F \tag{18}$$

(Triangle Inequality)

$$\|A+B\|_F^2 = tr((A+B)^H(A+B)) = tr(A^H A + A^H B + B^H A + B^H B) = \tag{19}$$

$$\begin{aligned}
\|A+B\|_F^2 &= tr((A+B)^H(A+B)) = tr(A^H A + A^H B + B^H A + B^H B) \\
&= tr(A^H A) + tr(A^H B) + tr(B^H A) + tr(B^H B) \\
&= tr(A^H A) + tr(B^H B) + 2tr(A^H B) \\
&\leq \|A\|_F^2 + \|B\|_F^2 + 2\|A\|\|B\| && \text{(by Cauchy-Schwarz)} \\
&= (\|A\|_F + \|B\|_F)^2
\end{aligned}$$

Therefore, $\|A+B\|_F \leq \|A\|_F + \|B\|_F$.

(Submultiplicativity) This is implied by the inequality $0 \leq tr(AB) \leq tr(A)tr(B)$. Therefore, $\|\cdot\|_F$ is a matrix norm. □

Exercise 4.31

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

Claim 13. $\|A\|_2 = \sigma_1$, where σ_1 is the largest singular value of A.

Proof. Let $A = U\Sigma V^H$ be the singular value decomposition of A.

$$\begin{aligned}
\|A\|_2 &= \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} \\
&= \sup_{x\neq 0} \frac{\left\|U\Sigma V^H x\right\|_2}{\|x\|_2} \\
&= \sup_{x\neq 0} \frac{\left\|\Sigma V^H x\right\|_2}{\|x\|_2} && \text{(because } U \text{ orthonormal)} \\
&= \sup_{y\neq 0} \frac{\|\Sigma y\|_2}{\|Vy\|_2} && \text{(change of variables)} \\
&= \sup_{y\neq 0} \frac{\|\Sigma y\|_2}{\|y\|_2} && \text{(because } V \text{ orthonormal)} \\
&= \|\Sigma\|_2
\end{aligned}$$

Observe that Σ is a diagonal matrix, and it is well known that $\|\Sigma\|_2 = \sigma_1$, the largest diagonal element. However, we prove this fact here for completeness.

Lemma 14. Let $B \in M_n(\mathbb{F})$ be a diagonal matrix with diagonal entries $b_1, b_2, \ldots, b_n$. Then $\|B\|_2$ is equal to the largest diagonal element.

Proof. We provide a lower and upper bound on $\|B\|_2$, and show that these are both equal to the largest diagonal element of B. Without loss of generality, suppose that b_k is the largest diagonal element.

Upper bound:

$$\begin{aligned}\|B\|_2^2 &= \sup_{x\neq 0} \frac{\|Bx\|_2^2}{\|x\|_2^2} \\ &= \sup_{x\neq 0} \frac{b_1^2x_1^2 + \cdots + b_n^2x_n^2}{x_1^2 + \cdots + x_n^2} \\ &\leq \sup_{x\neq 0} \frac{b_k^2(x_1^2 + \cdots + x_n^2)}{x_1^2 + \cdots + x_n^2} \\ &= b_k^2\end{aligned}$$

Lower bound: Let y be a vector of zeros, with a 1 in entry k. Then,

$$\begin{aligned}\|B\|_2^2 &= \sup_{x\neq 0} \frac{\|Bx\|_2^2}{\|x\|_2^2} \\ &\geq \frac{\|By\|_2^2}{\|y\|_2^2} \\ &= b_k^2\end{aligned}$$

Therefore, $\|B\|_2 = b_k$, the largest diagonal element. □

Therefore, by the above lemma, $\|A\|_2 = \sigma_1$. □

Claim 15. If A is invertible, then $\|A^{-1}\|_2 = \sigma_n^{-1}$.

Proof. Suppose A is in invertible, and let $A = U\Sigma V^H$ be the singular value decomposition of A. Then, $A^{-1} = (V^H)^{-1}\Sigma^{-1}U^{-1}$. Observe that this is also a singular value decomposition, because $(V^H)^{-1} = V$ and $U^{-1} = U^H$, because U and V are orthonormal matrices. Note that the largest singular value of Σ^{-1} is now $\frac{1}{\sigma_n}$. Therefore, by the first claim in this problem, we have that $\|A^{-1}\|_2 = \sigma_n^{-1}$. □

Claim 16. $\left\|A^H\right\|_2^2 = \left\|A^T\right\|_2^2 = \left\|A^HA\right\|_2 = \|A\|_2^2$

Proof. Let $A = U\Sigma V^H$ be the singular value decomposition of A. Then,

$$\begin{aligned}A^H &= V\Sigma^HU^H \\ A^T &= \overline{V}\Sigma U^T \\ A^HA &= (V\Sigma^HU^H)(U\Sigma V^H) = V(\Sigma^H\Sigma)V^H\end{aligned}$$

Observe that each of these is also a singular value decomposition. Now, consider the singular values of each of these decompositions. Since the singular values are real numbers, we have that $\Sigma^H = \Sigma$. Therefore, A^H, A^T and A all have the same singular values, and by the first claim in this problem, we have that $\left\|A^H\right\|_2^2 = \left\|A^T\right\|_2^2 = \|A\|_2^2 = \sigma_1^2$. Next, observe that the diagonal elements of $(\Sigma^H\Sigma)$ are simply the singular values squared. Therefore, $\left\|A^HA\right\|_2 = \sigma_1^2$. □

Claim 17. If $U \in M_m(\mathbb{F})$ and $V \in M_n(\mathbf{F})$ are orthonormal, then $\|UAV\|_2 = \|A\|_2$.

Proof.

$$\begin{aligned}
\|UAV\|_2 &= \sup_{x\neq 0} \frac{\|UAVx\|_2}{\|x\|_2} \\
&= \sup_{x\neq 0} \frac{\|AVx\|_2}{\|x\|_2} && \text{(because } U \text{ orthonormal)} \\
&= \sup_{y\neq 0} \frac{\|Ay\|_2}{\|Vy\|_2} && \text{(change of variables)} \\
&= \sup_{y\neq 0} \frac{\|Ay\|_2}{\|y\|_2} && \text{(because } V \text{ orthonormal)} \\
&= \|A\|_2
\end{aligned}$$

□

Exercise 4.32

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

Claim 18. If $U \in M_m(\mathbb{F})$ and $V \in M_n(\mathbb{F})$ are orthonormal, then $\|UAV\|_F = \|A\|_F$.

Proof. Let $U \in M_m(\mathbb{F})$ and $V \in M_n(\mathbb{F})$ be orthonormal. Then,

$$\begin{aligned}
\|UAV\|_F^2 &= tr((UAV)^H(UAV)) \\
&= tr(V^H A^H U^H U^H AV) \\
&= tr(V^H A^H AV) \\
&= tr(VV^H A^H A) \\
&= tr(A^H A) \\
&= \|A\|_F^2
\end{aligned}$$

Therefore, $\|UAV\|_F = \|A\|_F$. □

Claim 19. $\|A\|_F = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2)^{\frac{1}{2}}$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A.

Proof. Let $A = U\Sigma V^H$ be the singular value decomposition of A. Then,

$$\begin{aligned}
\|A\|_F &= \left\|U\Sigma V^H\right\|_F \\
&= \|\Sigma\|_F && \text{(by the above claim, as } U, V \text{ are orthonormal)} \\
&= (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2)^{\frac{1}{2}}
\end{aligned}$$

□

Exercise 4.33

Claim 20. Let $A \in M_n(\mathbb{F})$. Then,

$$\|A\|_2 = \sup_{\|x\|_2=1, \|y\|_2=1} |y^H A x| \tag{20}$$

Proof. Let $A \in M_n(\mathbb{F})$. We showed in Exercise 4.31 (i) that $\|A\|_2 = \|\Sigma\|_2$, where the singular value decomposition is $A = U\Sigma V^H$. Then,

$$\begin{aligned}
\sup_{\|x\|_2=1, \|y\|_2=1} |y^H \Sigma x| &= \sup_{\|x\|_2=1, \|y\|_2=1} |\langle y, \Sigma x\rangle| \\
&\leq \sup_{\|x\|_2=1, \|y\|_2=1} \|y\|_2 \|\Sigma x\|_2 && \text{(by Cauchy-Schwarz)} \\
&= \sup_{\|x\|_2=1} \|\Sigma x\|_2 \\
&= \|\Sigma\|_2 \quad \text{(by the definition of the 2-norm and lemma in 4.31)} \\
&= \|A\|_2
\end{aligned}$$

□

Exercise 4.36

Consider the matrix,

$$A = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix} \tag{21}$$

Then, the singular values of A are $\sigma_1 = 3$ and $\sigma_2 = 2$. However, the eigenvalues of A are -3 and -2.

Exercise 4.38

Claim 21. If $A \in M_{m\times n}(\mathbb{F})$, then the Moore-Penrose pseudoinverse of A satisfies the following:

1. $AA^\dagger A = A$.
2. $A^\dagger AA^\dagger = A^\dagger$.
3. $(AA^\dagger)^H = AA^\dagger$
4. $(A^\dagger A)^H = A^\dagger A$
5. $AA^\dagger = proj_{\mathscr{R}(A)}$ is the orthogonal projection onto $\mathscr{R}(A)$
6. $A^\dagger A = proj_{\mathscr{R}(A^H)}$ is the orthogonal projection onto $\mathscr{R}(A^H)$

Proof. Let $A \in M_{m\times n}(\mathbb{F})$, and let $A = U_1\Sigma_1 V_1^H$ be the compact form of the SVD of A. The Moore-Penrose pseudoinverse of A is $A^\dagger = V_1\Sigma_1^{-1}U_1^H$.

1. Observe that,

$$\begin{aligned}
AA^\dagger A &= (U_1\Sigma_1 V_1^H)(V_1\Sigma_1^{-1}U_1^H)(U_1\Sigma_1 V_1^H)\\
&= U_1\Sigma_1(V_1^H V_1)\Sigma_1^{-1}(U_1^H U_1)\Sigma_1 V_1^H\\
&= U_1\Sigma_1\Sigma_1^{-1}\Sigma_1 V_1^H\\
&= U_1\Sigma_1 V_1^H = A
\end{aligned}$$

where the second line follows because U_1 and V_1 are orthonormal. Therefore, $AA^\dagger A = A$.

2. Observe that,

$$\begin{aligned}
A^\dagger AA^\dagger &= (V_1\Sigma_1^{-1}U_1^H)(U_1\Sigma_1 V_1^H)(V_1\Sigma_1^{-1}U_1^H)\\
&= V_1\Sigma_1^{-1}(U_1^H U_1)\Sigma_1(V_1^H V_1)\Sigma_1^{-1}U_1^H\\
&= V_1\Sigma_1^{-1}\Sigma_1\Sigma_1^{-1}U_1^H\\
&= V_1\Sigma_1^{-1}U_1^H
\end{aligned}$$

where the second line follows because U_1 and V_1 are orthonormal. Therefore, $A^\dagger AA^\dagger = A^\dagger$.

3. Observe that,

$$\begin{aligned}
(AA^\dagger)^H &= ((U_1\Sigma_1 V_1^H)(V_1\Sigma_1^{-1}U_1^H))^H\\
&= U_1(\Sigma_1^{-1})^H V_1^H V_1\Sigma_1^H U_1^H\\
&= U_1U_1^H\\
&= U_1 I U_1^H\\
&= U_1\Sigma_1\Sigma_1^{-1}U_1^H\\
&= U_1\Sigma_1 V_1^H V_1\Sigma_1^{-1}U_1^H\\
&= AA^\dagger
\end{aligned}$$

4. Observe that,

$$\begin{aligned}
(A^\dagger A)^H &= ((V_1\Sigma_1^{-1}U_1^H)(U_1\Sigma_1 V_1^H))^H\\
&= V_1\Sigma_1^H U_1^H U_1(\Sigma_1^{-1})^H V_1^H\\
&= V_1V_1^H\\
&= V_1 I V_1^H\\
&= V_1\Sigma_1^{-1}\Sigma_1 V_1^H\\
&= V_1\Sigma_1^{-1}U_1^H U_1\Sigma_1 V_1^H\\
&= A^\dagger A
\end{aligned}$$

5. We need to show that $AA^\dagger$ is indeed a projection and also orthogonal. To show that $AA^\dagger$ is a projection, we must show that $(AA^\dagger)(AA^\dagger) = AA^\dagger$. This easily

follows from (1). Indeed, $(AA^\dagger)(AA^\dagger) = (AA^\dagger A)A^\dagger = AA^\dagger$. Next, write U_1 in terms of its columns as $U_1 = [u_1, \cdots, u_r]$. By the SVD, U_1 is an orthonormal basis for $\mathscr{R}(A)$. Then, using (3),

$$\begin{aligned}
AA^\dagger x &= U_1 U_1^H x \\
&= U_1[u_1^H x, \cdots, u_n^H x] \\
&= \sum_{i=1}^{r} u_i^H x u_i && \text{(r is the number of singular values of A)} \\
&= \sum_{i=1}^{r} \langle u_i, x \rangle u_i \\
&= proj_{\mathscr{R}(A)} x
\end{aligned}$$

Therefore, by definition, $AA^\dagger$ is an orthogonal projection onto $\mathscr{R}(A)$.

6. The proof is analogous to the proof of (5). Write $V_1 = [v_1, \cdots, v_r]$. By the SVD, V_1 is an orthonormal basis for $\mathscr{R}(A^H)$. Then, by (4),

$$\begin{aligned}
A^\dagger A x &= V_1 V_1^H x \\
&= V_1[v_1^H x, \cdots, v_n^H x] \\
&= \sum_{i=1}^{r} v_i^H x v_i && \text{(r is the number of singular values of A)} \\
&= \sum_{i=1}^{r} \langle v_i, x \rangle v_i \\
&= proj_{\mathscr{R}(A^H)} x
\end{aligned}$$

Therefore, by definition, $A^\dagger A$ is an orthogonal projection onto $\mathscr{R}(A^H)$.

$\square$