Matrix Analysis

Yunwei Ren

Contents

1 Eigenvalues, eigenvectors, and similarity

1.0 Introduction

1.

Proof. Let $S = \{x \in \mathbb{R}^n : x^T x = 1\}$, which is clearly a compact subset of $\mathbb{R}^n$. Consider the function $f : x \mapsto x^T Ax$. Since,

$$\|f(x+\delta) - f(x)\| = \|(x^T A)\delta + \delta^T (Ax) + \delta^T A\delta\| \le K\|\delta\|$$

for every $x \in \mathbb{R}$ and some fixed K, f is continuous. Hence, by Weierstrass's theorem, f attains its maximum value at some point $x \in S$. Namely, (1.0.3) has a solution x. Therefore, there exists some $\lambda \in \mathbb{R}$ such that $2(Ax - \lambda x) = 0$, implying that every real symmetric matrix has at least one real eigenvalue. □

2.

Proof. Let $S = \{x \in \mathbb{R}^n : x^T x = 1\}$ and m be the maximum value of $x \mapsto x^T Ax$ in S. Suppose λ is an eigenvalue of A and $u \neq 0$ is its associated eigenvector, then

$$Au = \lambda u \quad \Rightarrow \quad u^T Au = \lambda\|u\|^2 \quad \Rightarrow \quad (u/\|u\|)^T A(u/\|u\|) = \lambda \quad \Rightarrow \quad m \ge \lambda.$$

Meanwhile, by the previous discussion, m itself is a eigenvalue of A. Hence, it is the largest real eigenvalue of A. □

1.1 The eigenvalue-eigenvector equation

1.

Proof. It follows from

$$(A^{-1} - \lambda^{-1} I)x = (A^{-1} - \lambda^{-1} A^{-1} A)x = \lambda^{-1} A^{-1}(\lambda I - A)x = 0.$$

□

3.

Proof. Since $A \in M_n(\mathbb{R})$, $u, v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,

$$Ax = \lambda x \quad \Rightarrow \quad Au + iAv = \lambda u + i\lambda v$$

implies $Au = \lambda u$ and $Av = \lambda v$. As $x \neq 0$, at least one of u and v is nonzero and therefore A has a real eigenvector associated with λ. It can happen that only one of u and v is an eigenvector of A, because if $x \in \mathbb{R}^n$, which may happen as we discussed above, the imaginary part of x is 0. Finally, if x is a real eigenvector of A, then the eigenvalue λ it associated with must be real. Otherwise, at least one entry of λx is not real as $x \neq 0$, contradicting with the fact that Ax is real. □

5.

Proof. Let $p(t) = t^2 - t$. Since A is idempotent, $p(A) = A^2 - A = 0$. Hence, 0 is the only eigenvalue of $p(A)$. By Theorem 1.1.6, the only values the eigenvalues of A can be are the zeros of p, namely, 0 and 1.

Suppose A is nonsingular, then multiplying A^{-1} on the both sides of $A^2 = A$ yields $A = I$. □

7.

Proof. Suppose $\lambda \in \sigma(A)$ and x is its associated eigenvector, then

$$\begin{aligned} 0 &= (A - \lambda I)x = x^*(A^* - \bar{\lambda} I) = x^*(A - \bar{\lambda} I) \\ \Rightarrow \quad 0 &= x^*(A - \bar{\lambda} I)x = x^*Ax - \bar{\lambda}x^*x = (\lambda - \bar{\lambda})\|x\|^2. \end{aligned}$$

Hence, $\lambda = \bar{\lambda}$, implying all eigenvalues of A are real. □

9.

Solution. Solve the equation $\det(A - \lambda I) = 0$ and we get $\lambda = \pm i$. □

11.

Proof. If $\operatorname{rank}(A - \lambda I) < n - 1$, then $\operatorname{adj}(A - \lambda I) = 0$ by (0.8.2) and therefore we can always choose y to be the 0 and the other parts of the proposition clearly hold. Hence, in the following discussion, we assume that $\operatorname{rank}(A - \lambda I) = n - 1$.

Apply the full-rank factorization and we get $\operatorname{adj}(A - \lambda I) = \alpha x y^*$ for some nonzero $\alpha \in \mathbb{C}$ and $x, y \in \mathbb{C}^n$. Replacing x with αx and α with 1 proves the first part.

Suppose $\operatorname{adj}(A - \lambda I) = [\beta_1, \ldots, \beta_n]$, then

$$(A - \lambda I)\operatorname{adj}(A - \lambda I) \quad \Rightarrow \quad (A - \lambda I)\beta_k = 0 \quad (k = 1, 2, \ldots, n),$$

implying that β_k is an eigenvector of A associated with λ as long as it is nonzero. □

13.

Proof. If $\operatorname{rank} A < n - 1$, then x is always an eigenvector of $\operatorname{adj} A$ associated with 0 as $\operatorname{adj} A = 0$. Hence, we may assume that rank $= n - 1$. Then $\operatorname{adj} A = (\det A)A^{-1}$. By Exercise 1, x is an eigenvector of A^{-1} and therefore an eigenvector of $\operatorname{adj} A$. □

1.2 The characteristic polynomial and algebraic multiplicity

2.

Proof. Suppose $A = [a_{ij}]_{m,n} = [\alpha_1, \ldots, \alpha_n]^T$ and $B = [b_{ij}]_{n,m} = [\beta_1, \ldots, \beta_n]$, then

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} \alpha_i \beta_i = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji} = \sum_{j=1}^{m} \sum_{i=1}^{n} b_{ji} a_{ij} = \operatorname{tr}(BA).$$

Hence, for nonsingular $S \in M_n$, $\operatorname{tr}(S^{-1}AS) = \operatorname{tr}(S(S^{-1}A)) = \operatorname{tr}(A)$.

For $A \in M_n$, $\det(S^{-1}AS) = \det(S)\det(S^{-1})\det(A) = \det(A)$, which means the determinant function on M_n is similarity invariant. □

4.

Proof. It follows immediately from the fact that $\sigma(A) \subset \{0, 1\}$ and $S_k(A)$ is the sum of some $\prod \lambda_{i_j}$. □

6.

Proof. $\operatorname{rank}(A - \lambda I) = n - 1$ implies the matrix $A - \lambda I$ is singular, and therefore λ is an eigenvalue of A. However, it may not have multiplicity 1. For example[1], suppose

[1] Thanks to Zhihan Jin, one of my classmates.

$A = \left[\begin{smallmatrix} 1 & -1 \\ 1 & -1 \end{smallmatrix}\right]$. $\operatorname{rank} A = 1$ but 0, the only eigenvalue of A is of multiplicity 2. $\square$

8.

Proof. $p_{A+\lambda I}(t) = \det(tI - (A + \lambda I)) = \det((t-\lambda)I - A) = p_A(t - \lambda)$ and hence the eigenvalues of $A + \lambda I$, the zeros of $p_{A+\lambda}(t)$, are $\lambda_1 + \lambda, \dots, \lambda_n + \lambda$. $\square$

10.

Proof. Since $p_A(t)$ has n roots and non-real roots of a polynomial come in paris, at least one of the roots is real. Hence, A has at least one real eigenvalue. $\square$

12. TODO

14.

Proof. Suppose $C = \left[\begin{smallmatrix} \mu & 0 \\ * & B \end{smallmatrix}\right]$. By the exercise on p52,

$$p_A(t) = (t - \lambda)p_C(t) = (t - \lambda)p_{C^T}(t) = (t - \lambda)(t - \mu)p_B(t).$$

$\square$

16.

Proof. $f(t) = \det(A + (tx)y^T = \det A + y^T(\operatorname{adj} A)tx = \det A + t\beta$ where $\beta = y^T(\operatorname{adj} A)x$, a constant independent of t. Hence, for $t_1 \neq t_2$

$$\frac{t_2 f(t_1) - t_1 f(t_2)}{t_2 - t_1} = \frac{t_2(\det A + t_1\beta) - t_1(\det A + t_2\beta)}{t_2 - t_1} = \det A.$$

For the second part, we can get from calculation that

$$f(-b) = \det(A - b[1, \dots, 1]^T[1, \dots, 1]) = (d_1 - b)\cdots(d_n - b) = q(b)$$

and $f(-c) = q(-c)$. Hence, if $b \neq c$,

$$\det A = \frac{(-c)f(-b) - (-b)f(-c)}{(-c) - (-b)} = \frac{bq(c) - cq(b)}{b - c}.$$

Now suppose $b = c$. Note that $f(t)$ is a linear function of t, which is differentiable, implying that

$$\det A = \lim_{t_2 \to t_1} \frac{t_2 f(t_1) - t_1 f(t_2)}{t_2 - t_1} = f'(t_1)t_1 - f(t_1).$$

Meanwhile, since $q(t)$ is continuous, $q(t) \to f(-b)$ as $t \to b$. Thus,

$$\det A = \lim_{c \to b} \frac{(-c)f(-b) - (-b)f(-c)}{(-c) - (-b)} = q(b) - bq'(b).$$

Let

$$A_* = \lambda I - A = \begin{bmatrix} \lambda & -b & \cdots & -b \\ -c & \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & -b \\ -c & \cdots & -c & \lambda \end{bmatrix}.$$

and $q_*(t) = (\lambda - t)^n$, then by the previous result,

$$p_A(\lambda) = \frac{-bq_*(-c) - (-c)q_*(-b)}{-c-(-b)} = \frac{b(\lambda+c)^n - c(\lambda+b)^n}{b-c}, \qquad \text{if } b \neq c,$$
$$p_A(\lambda) = q_*(-b) - (-b)q_*'(-b) = (\lambda+b)^{n-1}(\lambda-(n-1)b), \qquad \text{if } b = c.$$

□

18.

Proof. The identity can be derived immediately from Observation 1.2.4 and the identity $a_1 = (-1)^{n-1} \operatorname{tr} \operatorname{adj}(A)$, the proof of which can be found on p53. □

20.

Proof. By (1.2.13),

$$\det(I+A) = (-1)^n p_A(-1) = (-1)^n \left((-1)^n + \sum_{k=1}^{n} (-1)^{n-k} E_k(A)(-1)^k \right) = 1 + \sum_{k=1}^{n} E_k(A).$$

□

22.

Proof. Suppose

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

then

$$\begin{aligned} p_{C_n(\varepsilon)}(t) &= \det(A + [0,\dots,0,1]^T[-\varepsilon,0,\dots,0]) \\ &= \det A - \varepsilon[1,0,\dots,0](\operatorname{adj} A)[0,\dots,0,1]^T \\ &= \det A - \varepsilon((\operatorname{adj} A)[1,n]) \\ &= \det A - \varepsilon \det A[\{n\}^c, \{1\}^c] \\ &= t^n - \varepsilon. \end{aligned}$$

And its spectrum, namely the set of roots of $p_{C_n(\varepsilon)}$, is $\{\varepsilon^{1/n} e^{2\pi i k/n} : k = 0, 1, \dots, n-1\}$. Hence,

$$\rho(I + C_n(\varepsilon)) = 1 + \rho(C_n(\varepsilon)) = 1 + \varepsilon^{1/n}.$$

□

1.3 Similarity

1.

Proof. (a) Since A and B are diagonalizable and commute, by Theorem 1.3.21, they are simultaneously diagonalizable. Hence, there exists some nonsingular $S \in M_n$ such that

$$\begin{aligned} A + B &= S^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)S + S^{-1}\operatorname{diag}(\mu_{i_1}, \ldots, \mu_{i+n})S \\ &= S^{-1}\operatorname{diag}(\lambda_1 + \mu_{i_1}, \ldots, \lambda_n + \mu_{i_n})S. \end{aligned}$$

Therefore, $\sigma(A + B) = \{\lambda_1 + \mu_{i_1}, \ldots, \lambda_n + \mu_{i_n}\}$.
(b) By Exercise 1.1.6, $\sigma(B) = \{0\}$, completing the proof.
(c) $\sigma(AB) = \{\lambda_1\mu_{i_1}, \ldots, \lambda_n\mu_{i_n}\}$, because

$$S^{-1}(AB)S = (S^{-1}AS)(S^{-1}BS) = \operatorname{diag}(\lambda_1\mu_{i_1}, \ldots, \lambda_n\mu_{i_n}).$$

□

2.

Proof. Suppose that $p(z) = \sum_{i=0}^{n} a_i z^i$ and $q(z) = \sum_{j=0}^{m} b_j z^j$, then

$$p(A)q(B) = \left(\sum_{i=0}^{n} a_i A^i\right)\left(\sum_{j=0}^{n} b_j B^j\right) = \sum_{i,j} a_i b_j A^i B^j = \sum_{i,j} a_i b_j B^j A^j = q(B)p(A).$$

□

3.

Proof.

$$\sum_{k=0}^{n} a_k A^k = \sum)k = 0^n S^{-1}(a_k \Lambda^k)S = S^{-1}p(\Lambda)S.$$

□

4.

Proof. First we assume that $A = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$. Since $AB = BA$, by (0.7.7), B is also diagonal matrix. Suppose that $B = \operatorname{diag}(\beta_1, \ldots, \beta_n)$ where b_i may coincide. Then $p(A) = B$ is equivalent to $p(\alpha_i) = \beta_i$ for $i = 1, \ldots, n$. Since α_i are distinct, we can construct such a polynomial by interpolation.

For the general case, note that in the proof of Theorem 1.3.12, $n_i = 1$ as long as α_i are distinct. Hence, A and B are simultaneously diagonalizable. Suppose that $A = S\operatorname{diag}(\alpha_1, \ldots, \alpha_n)S^{-1}$ and $B = S\operatorname{diag}(\beta_1, \ldots, \beta_n)S^{-1}$. Let p be the same polynomial as in the last paragraph. Then by P3,

$$p(A) = Sp(\operatorname{diag}(\alpha_1, \ldots, \alpha_n))S^{-1} = S\operatorname{diag}(\beta_1, \ldots, \beta_n)S^{-1} = B.$$

□

5.

Proof. Let $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = I$. Clear that A and B commute but are not simultaneously diagonalizable since A can not be diagonalized. This does not violate 1.3.12 because in 1.3.12, A and B are required to be diagonalizable. $\square$

6.

Proof. (a) It follows immediately from the fact that the multiplication of block diagonal matrices are block-wise.
(b) Suppose that $A = S\Lambda S^{-1}$, then

$$p_A(t) = \det(tI - A) = \det(tSS^{-1} - S\Lambda S^{-1}) = \det(tI - \Lambda) = p_\Lambda(t).$$

By P3, $p_\Lambda(A) = Sp_\Lambda(\Lambda)S^{-1}$. Therefore, $p_A(A) = p_\Lambda(A) = Sp_\Lambda(\Lambda)S^{-1} = 0$. $\square$

7.

Proof. Suppose $B = \operatorname{diag}(b_1, \ldots, b_n)$, then clear that $A = \operatorname{diag}(\sqrt{b_1}, \ldots, \sqrt{b_n})$ is a square root of B.

Assume that such A exists, then by Theorem 1.1.6, $\sigma(A) = \{0\}$, implying that

$$A = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}.$$

Therefore,

$$A^2 = \begin{bmatrix} ab & 0 \\ 0 & ab \end{bmatrix} \neq B.$$

Contradiction. $\square$

8.

Proof. If A and B are simultaneously diagonalizable, then clear that A and B commute. Now we suppose that A and B commute and $\lambda_1, \ldots, \lambda_n$ are the distinct eigenvalues of A. Since for each i, the eigenspace $E(A, \lambda_i)$ is an one-dimensional invariant subspace of B, then every vector in $E(A, \lambda_i)$ is also an eigenvector of B. Hence, there exists a basis of $\mathbb{C}^n$ consisting of the common eigenvectors of A and B, implying that they are simultaneously diagonalizable. $\square$

9.

Proof. By Theorem 1.3.22, AB and BA have the same eigenvalues. And as similar matrices are of the same rank, AB and BA are not similar. $\square$

10.

Proof. We argue by contradiction. Assume that there exists some vector in the list which belongs to the span of the previous vectors and suppose $x_q^{(p)}$ is the first such vector. Then it equals to some linear combination of the previous vectors. Compute $Ax_q^{(p)}$ using the linearity first and then using the fact that $x_q^{(p)}$ is an eigenvalues. Compare the two formula and we will obtain a contradiction. $\square$

11.

Proof. Suppose that $A, B \in M_n$ commute and $Ax = \lambda x$ where $x \neq 0$ and k the is the smallest integer such that $B^k x \in \operatorname{span}\{x, Bx, \ldots, B^{k-1}x\} = \mathcal{S}$. For every $u = \sum_{i=0}^{k-1} x_i B^i x \in \mathcal{S}$, Bu is a linear combination of $\sum_{i=0}^{k-2} x_i B^{i+1} x$ and $x_{k-1} B^k x \in \mathcal{S}$. Hence $Bu \in \mathcal{S}$ and therefore $\mathcal{S}$ is B-invariant. By Observation 1.3.18, there exists some $0 \neq y \in \mathcal{S}$ which is an eigenvector of B. Meanwhile, since

$$A \sum_{i=0}^{k-1} x_i B^i x = \sum_{i=0}^{k-1} x_i (AB^i) x = \sum_{i=0}^{k-1} x_i B^i \lambda x = \lambda \sum_{i=0}^{k-1} x_i B^i x,$$

every nonzero vector in $\mathcal{S}$ is a eigenvector of A and so does y. Hence, A and B have a common eigenvector y.

Now we argue by induction on m, the size of the finite commuting family $\mathcal{F} = \{A_1, \ldots, A_m\}$. Suppose that $y \neq 0$ is a common eigenvector of $A_1, \ldots, A_{m-1}$ and let k be the smallest integer such that $A_m^k y \in \operatorname{span}\{y, A_m y, \ldots, A_m^{k-1} y\} = \mathcal{S}$. Then by some argument similar to the previous one, $\mathcal{S}$ is A_m-invariant and hence contains a eigenvector z of A_m. Meanwhile, since A_i and A_m commute, every nonzero vector in $\mathcal{S}$ is an eigenvector of A_i for $i = 1, \ldots, m-1$ and so does z, concluding that matrices in a finite commuting families share a common eigenvector.

M_n is linear space of dimension n^2 and $\mathcal{F}$ is a subspace of M_n since for any $A, B, C \in \mathcal{F}$ and $a, b \in \mathbb{C}$,

$$(aA + bB)C = a(AC) + b(BC) = a(CA) + b(CB) = C(aA + bB).$$

Let $\mathcal{B} = \{B_1, \ldots, B_k\}$ be a basis of $\mathcal{F}$. Since $\mathcal{B}$ is finite and commuting, (b) shows that the matrices in $\mathcal{B}$ have a common eigenvector x. Hence, supposing $B_i x = \lambda_i x$,

$$\left(\sum_{i=1}^{k} b_i B_i \right) x = \sum_{i=1}^{k} b_i (B_i x) = \left(\sum_{i=1}^{k} b_i \lambda_i \right) x$$

where b_i are some scalars. Thus, x is a eigenvector of every $A \in \mathcal{F}$. $\square$

12.

Proof. Suppose that A is nonsingular. Then $BA = A^{-1}(AB)A$, i.e., $BA \sim AB$. Therefore BA is diagonalizable as long as AB is. We can produce the same result with a similar argument if B is nonsingular.

Suppose $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Then $AB = 0$, which is a diagonal matrix and $BA = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, which is not diagonalizable since all eigenvectors of BA are of form $[k, 0]^T$. $\square$

13.

Proof. Since similar diagonalizable matrices have the same eigenvalues and multiplicities, their characteristic polynomials are therefore the same and vice versa.

However, this is not true for two matrices which are not both diagonalizable. For example, $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $0_{2\times 2}$ have the same characteristic polynomial but they are not similar. $\square$

14. This exercise provides several ways to prove a matrix to be not diagonalizable.

Proof. (a) It follows immediately form the fact that the rank of a matrix is similarity invariance.
(b) It holds for diagonal matrices. And since rank is invariant under similarity, it holds for all diagonalizable matrices.
(c) A matrix is nilpotent iff it is similar to a matrix whose nonzero entries are all above the main diagonal. Hence a diagonalizable matrix is nilpotent iff it is similar to 0 iff it equals to 0.
(d) $\operatorname{tr} A = 0$ implies the nonzero eigenvalues of A comes in $\pm$ paris. Hence, $\operatorname{rank} A$ is even.
(e) $\operatorname{rank} B = 1$ but it has no nonzero eigenvalue. Hence, it is not diagonalizable by (a). Since $\operatorname{rank} B^2 = 0 \neq \operatorname{rank} B$, B is not diagonalizable by (b). B is a nonzero nilpotent matrix and therefore is not diagonalizable by (c). $\operatorname{tr} B = 0$ but $\operatorname{rank} B = 1$. Hence B is not diagonalizable by (d). □

15.

Proof. Suppose that $A = S\Lambda S^{-1}$ where Λ is a diagonal matrix, then

$$p(A) = \sum_{k=0}^{n} a_k (S\Lambda S^{-1})^k = \sum_{k=0}^{n} a_k S\Lambda^k S^{-1} = Sp(\Lambda)S^{-1}.$$

Clear that $p(\Lambda)$ is again a diagonal matrix. Hence, $p(A)$ is also diagonalizable.

However the converse is not true. For example, $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ is not diagonalizable but $A^2 = 0$ itself is a diagonal matrix. □

17.

Proof. Suppose that $A = TBT^{-1}$ where $T \in M_n(\mathbb{R})$ is nonsingular, then $\bar{A} = \overline{TBT^{-1}} = \bar{T}\bar{B}\bar{T}^{-1} = T\bar{B}T^{-1}$ since T is real. And the converse is obviously true. □

19.

Proof. Clear that $Q = Q^T$ and $Q^2 = I$, implying $Q = Q^{-1}$.
(a) Suppose that $A = \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right]$.

$$0 = K_{2n}A - AK_{2n} = \begin{bmatrix} A_{21} - A_{12} & A_{22} - A_{11} \\ A_{11} - A_{22} & A_{12} - A_{21} \end{bmatrix}.$$

Hence, A is 2-by-2 block centrosymmetric. And the proof of the converse is trivial. If A is nonsingular, then we have $A^{-1}K_{2n} = K_{2n}A^{-1}$, which implies A^{-1} is 2-by-2 block centrosymmetric. Meanwhile, since $K_{2n}^{-1} = K_{2n}$, $K_{2n}AK_{2n} = A$. Suppose B is a 2-by-2 block centrosymmetric matrix, then

$$K_{2n}AB = K_{2n}A(K_{2n}BK_{2n}) = (K_{2n}AK_{2n})B_{2n}K_{2n} = ABK_{2n}.$$

Therefore, AB is a 2-by-2 block centrosymmetric matrix as well.
(b)

$$\begin{aligned} Q^{-1}AQ &= \frac{1}{2}\begin{bmatrix} I_n & I_n \\ I_n & -I_n \end{bmatrix}\begin{bmatrix} B & C \\ C & B \end{bmatrix}\begin{bmatrix} I_n & I_n \\ I_n & -I_n \end{bmatrix} \\ &= \begin{bmatrix} B + C & 0 \\ 0 & B + C \end{bmatrix} = (B + C) \oplus (B - C). \end{aligned}$$

(c)

$$\det A = \det(Q^{-1}AQ) = \det \begin{bmatrix} B+C & 0 \\ 0 & B+C \end{bmatrix} = \det(B^2 + CB - BC - C^2)$$

and $\operatorname{rank} A = \operatorname{rank}(B+C) + \operatorname{rank}(B-C)$ follows immediately from $Q^{-1}AQ = (B+C) \oplus (B-C)$.
(d) $Q^{-1}\left[\begin{smallmatrix} 0 & C \\ C & 0 \end{smallmatrix}\right] = C \oplus (-C)$. Since $p_{C\oplus(-C)}(t) = p_C(t)p_{-C}(t)$, the eigenvalues occur in $\pm$ pairs. □

20.

Proof. (b) As A is nonsingular, $A^{-1}A = I_n$ and therefore $R_1(A^{-1})R_1(A) = R_1(A^{-1}A) = I_{2n}$ by (a). Hence, $R_1(A)$ is nonsingular and $R_1(A)^{-1} = R_1(A^{-1})$, which also implies that $R_1(A)^{-1}$ has the same block structure as $R_1(A)$.
(g) By (f), $R_1(A)$ is similar to $A \oplus \bar{A}$. Therefore, $\sigma(R_1(A)) = \{\lambda_1, \ldots, \lambda_n, \bar{\lambda}_1, \ldots, \bar{\lambda}_n\}$.
(h) By (f), $\det R_1(A) = \det(A \oplus \bar{A}) = |\det A|^2 \geq 0$. Since $\operatorname{rank}(A \oplus \bar{A}) = 2 \operatorname{rank} A$ and rank is invariant under similarity, $\operatorname{rank} R_1(A) = 2 \operatorname{rank} A$.
(i) It follows immediately from (h). □

23.

Proof. Suppose that there exists some $X \in M_{n,m}$ such that $C = BX$ and let $S = \left[\begin{smallmatrix} I_n & X \\ 0 & I_m \end{smallmatrix}\right]$. Clear that S is nonsingular and $S^{-1} = \left[\begin{smallmatrix} I_n & -X \\ 0 & I_m \end{smallmatrix}\right]$. Since

$$S \begin{bmatrix} B & BX \\ 0_n & 0_m \end{bmatrix} S^{-1} = \begin{bmatrix} B & 0 \\ 0 & 0_m \end{bmatrix},$$

A is similar to $B \oplus 0_m$.

Now we suppose that A is similar to $B \oplus 0_m$. Since similar matrices have the same rank, $\operatorname{rank}[B\, C] = \operatorname{rank} B$. □

28.

Proof.

$$\begin{aligned} \det(I_m + AB) &= \det \begin{bmatrix} I_m + AB & A \\ 0 & I_n \end{bmatrix} \\ &= \det \begin{bmatrix} I_m + AB & A \\ B + BAB & I_n + BA \end{bmatrix} \\ &= \det \begin{bmatrix} I_m & A \\ 0 & I_n + BA \end{bmatrix} \\ &= \det(I_n + BA). \end{aligned}$$

□

29.

Proof. Since $\det A = \sum_\sigma (\operatorname{sgn}\sigma \prod_{i=1}^n a_{i\sigma(i)})$, where σ is any permutation of $\{1, 2, \ldots, n\}$, the determinant of a matrix whose entries are integers is an integer.

Suppose that the a_{ij} is changed from -1 to 1 and denote the new matrix by $\tilde{A}$. Let $x, y \in \mathbb{C}^n$ be two vectors such that $x_i = 1$ and $y_j = 2$, then $\tilde{A} = A + xy^T$. By Cauchy's identity,

$$\det \tilde{A} = \det A + y^T(\operatorname{adj} A)x = \det A + 2 \det A[\{i\}^c, \{j\}^c].$$

Hence, the parity of $\det A$ is unchanged.

[2] Since changing a -1 entry to 1 does not change the parity of the determinant, we can change all the entries to 1. Hence, the parity of $\det A$ is the same as the parity of $\det(J_n - I)$. Induction on n yields that it is opposite to the parity of n. Thus, if n is even, then $\det A$ is odd and therefore nonzero, implying A is nonsingular. □

30.

Proof. By Theorem 1.3.27, there exists some nonsingular $R = \operatorname{diag}(r_1, \ldots, r_n)$ such that $T = SR$, Hence,

$$Tf(\Lambda)T^{-1} = SRf(\Lambda)R^{-1}S^{-1} = Sf(\Lambda)S^{-1}$$

as $f(\Lambda)$, a diagonal matrix, commute with every matrix. Therefore, $\cos^2 A + \sin^2 A = I$ as $\cos^2 x + \sin^2 x = 1$ for every $x \in \mathbb{R}$. □

31.

Proof. The characteristic polynomial of the matrix is $p(t) = (t-a)^2 + b^2 = (t-a-ib)(t-a+ib)$. Hence its eigenvalues are $a \pm ib$. □

33.

Proof.
(a) Since A is real, $A\bar{x} = \overline{\bar{A}x} = \overline{Ax} = \overline{\lambda x} = \bar{\lambda}\bar{x}$.
(b) Since λ is not real, x and $\bar{x}$ are associated with different eigenvalues. Hence, x and $\bar{x}$ are linear independent. Suppose that $mu + nv = 0$. Then

$$0 = m(x + \bar{x}) - in(x - \bar{x}) = (m - in)x + (m + in)\bar{x}.$$

Since $m - in = 0$ and $m + in = 0$, $m = n = 0$. Thus, u and v are also linear independent.
(c)

$$Au = \frac{1}{2}A(x + \bar{x}) = \frac{1}{2}(\lambda x + \bar{\lambda}\bar{x}) = \frac{1}{2}[(a + ib)(u + iv) + (a - bi)(u - iv)] = au - bv.$$

Similarly, $Av = bu + av$. Hence,

$$A[u, v] = [Au, Av] = [au - bv, bu + av] = [u, v]B.$$

(d) Since $S\left[\begin{smallmatrix} I_2 \\ 0 \end{smallmatrix}\right] = [u\, v]$, $S^{-1}[u\, v] = \left[\begin{smallmatrix} I_2 \\ 0 \end{smallmatrix}\right]$ and the proof of the next result is trivial.
(e) Since $p_A(t) = p_B(t)p_{A_1}(t)$, the result amounts to the fact that λ and $\bar{\lambda}$ are two roots of $p_B(t)$. □

[2] I don't know what J_n actually is here and assume it to be the matrix whose entries are all 1.

1.4 Left and right eigenvectors and geometric multiplicity

2.

Proof. Since

$$p_A(t) = p_{A^T}(t) = p_{-A}(t) = \det(tI + A) = (-1)^n \det((-t)I - A) = (-1)^n p_A(-t),$$

$-\lambda$ is an eigenvalue of A with multiplicity k as long as λ is. Thus, if n is odd, $0 \in \sigma(A)$ as the nonzero eigenvalues come in pairs. Hence, A is singular. Since the principal submatrices of a skew symmetric matrix are still skew symmetric, the ones with odd size are singular. Finally, since skew symmetric matrices are rank principal, having a nonsingular principal submatrix of size $r \times r$ if its rank is r, $\operatorname{rank} A$ is even. □

4.

Proof. Note that $S^{-1} = S$ and multiplying S on the left and right are respectively equivalent to changing the sign of the odd rows and columns. Hence, $S^{-1}AS = SAS = -A$. Since $p_{-A}(t) = (-1)^n p_A(-t)$ and similar matrices have the same eigenvalues with the same multiplicities, $-\lambda$ is an eigenvalue of A with multiplicity k as long as λ is. Since the eigenvalues of A come in $\pm$ paris, $0 \in \sigma(A)$, hence A is singular, if n is odd. □

6.

Proof. (a) Since x and y are entrywise positive, $y^*x > 0$. Hence, by Theorem 1.4.7, the subspace $\operatorname{span}(x)$ and the orthogonal complement W of y^* are both A-invariant. Let u be an entrywise nonnegative right eigenvector of A associated with eigenvalue μ. It belongs to some A-invariant subspace. Since $y^*u > 0$, it does not belong to W and therefore $u \in \operatorname{span}(x)$. Thus, $\mu = \lambda$. If u is a left eigenvector, the argument is similar.
(b) Since the algebraic multiplicity is no less than the geometric multiplicity, λ has geometric multiplicity 1. □

10.

Proof. Suppose $T = [\beta_1, \dots, \beta_n]$. Then

$$T^*A = \begin{bmatrix} \beta_1^* \\ \vdots \\ \beta_n^* \end{bmatrix} A = \begin{bmatrix} \beta_1^* A \\ \vdots \\ \beta_n^* A \end{bmatrix} = \begin{bmatrix} \lambda_1 \beta_1^* \\ \vdots \\ \lambda_n \beta_n^* \end{bmatrix} = \Lambda T^*.$$

where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Therefore, $AT^{-*} = T^{-*}\Lambda$, implying that T^{-*} are the right eigenvectors of A. □

12.

Proof. (a) First suppose that every list of $n-1$ columns of $A - \lambda I = [\beta_1, \dots, \beta_n]$ is linearly independent and let $x \neq 0$ be an eigenvector of A associated with λ. We argue by contradiction, assuming that x_i, the i-th entry of x, is zero. Then

$$0 = (A - \lambda)x = x_1\beta_1 + \cdots + x_{i-1}\beta_{i-1} + x_{i+1}\beta_{i+1} + \cdots + x_n\beta_n.$$

As $\beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_n$ are linearly independent, this implies $x_i = 0$ for each $i = 1, \dots, n$, contradicting with the assumption $x \neq 0$.
To prove the converse part, we continue to argue by contradiction and assume that $\beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_n$ are linearly dependent. Consider the equation $[\beta_1, \dots, \beta_n]x = 0$. Even if we restrict the i-th entry of x to be 0, the equation is still solvable, leading to a contradiction and completing the proof.
(b) The previous result shows that every list of $n-1$ columns of $A - \lambda I$ is linearly independent. Since $A - \lambda I$ is singular, this implies $\operatorname{rank}(A - \lambda I) = n - 1$. By the rank-nullity theorem, the geometric multiplicity of λ is 1. □

14.

Proof. (a) It follows immediately from

$$(A - \lambda I)\operatorname{adj}(A - \lambda I) = p_A(\lambda)I = 0 \quad \Rightarrow \quad A \operatorname{adj}(A - \lambda I) = \lambda \operatorname{adj}(A - \lambda I).$$

(b) The proof is similar to the one of (a).
(c) Given $\lambda \in \sigma(A)$, $\operatorname{adj}(A - \lambda) \neq 0$ iff $\operatorname{rank}(A - \lambda I) = n - 1$ iff λ has geometric multiplicity 1.
(d) It follows from

$$\operatorname{adj}(A - \lambda I) = \begin{bmatrix} d - \lambda & -b \\ -c & a - \lambda \end{bmatrix}$$

and the results of (a) and (b). □

2 Unitary Similarity and Unitary Equivalence

2.1 Unitary matrices and the QR factorization

8.

Proof.
(a) A complex orthogonal matrix A which is real is clearly unitary. Suppose A is unitary. Then $A^*A = A^TA$, implying that $A^* = A^T$ and hence A is real.
(b) Note that $S^2 = 1$. Hence,

$$A^TA = ((\cosh t)I - (i\sinh t)S)((\cosh t)I + (i\sinh t)S) = (\cosh^2 t)I + (\sinh^2 t)S^2 = I.$$

Namely, $A(t)$ is complex orthogonal. By (a), $A(t)$ is unitary only if it is real. Hence $A(t)$ being unitary implies $t = 0$.
(c) Let $A_n = \mathrm{diag}(\sqrt{n+1} + i\sqrt{n}, 1, \ldots, 1)$ which is complex orthogonal for each n. However, as $n \to \infty$, A_n is not bounded.
(d) First, every complex orthogonal matrix is invertible and the inverse of which is also complex orthogonal. Meanwhile, I is complex orthogonal. Finally, given complex orthogonal A and B,

$$(AB)^T(AB) = B^T(A^TA)B = I.$$

Namely, AB is also complex orthogonal. Hence, the set of complex orthogonal matrices of a given size forms a group.
(e) Since $1 = \det I = \det(A^TA) = (\det A)^2$, $|\det A| = 1$. Meanwhile, as e and $1/e$ are the eigenvalues of $A(t)$, A can have eigenvalues whose norm is not 1. □

10.

Proof. Since $(Ux)^*(Uy) = x^*U^*Uy = x^*y$, Ux and Uy are orthogonal iff x and y are orthogonal. □

11.

Proof. $A^{-1} = -A^T$ iff $-AA^T = I$ iff $(iA)(iA)^T = (-iA)(-iA)^T = I$ iff $\pm iA$ is orthogonal. Furthermore, $A^{-1} = A^{i\theta}A^T$ iff $Ae^{i\theta}A^T = I$ iff $(e^{i\theta/2}A)(e^{i\theta}A)^T = I$ iff $e^{i\theta/2}A$ is orthogonal. When $\theta = 0$, it is simply the orthogonal matrices and the skew orthogonal matrices when $\theta = \pi$. □

12.

Proof. Suppose $A = T^{-1}UT$ where U is unitary. Then $A^{-1} = T^{-1}U^{-1}T = T^{-1}U^*T$ and $A^* = T^*U^*T^{-*}$. Hence,

$$U^* = TA^{-1}T^{-1} = T^{-*}A^*T^* \quad \Rightarrow \quad A^{-1} = T^{-1}T^{-*}A^*T^*T = (T^*T)^{-1}A^*(T^*T).$$

Namely, A^{-1} and A^* are similar. □

23.

Proof. Since Q is unitary, $\det Q = \pm 1$. Hence $|\det A| = |\det Q| \det R = r_{11} \cdots r_{nn}$ as R is an upper triangular matrix with nonnegative main diagonal entries.

Meanwhile, for each i, $a_i = Qr_i$. Since unitary matrices are isometries, $\|a_i\|_2 = \|r_i\|_2$. And clear that $\|r_i\|_2 = \sqrt{\sum |r_{ij}|^2} \geq r_{ii}$. The equality holds iff $r_{ij} = 0$ for $j \neq i$ iff $a_i = r_{ii} q_i$.

Therefore,

$$|\det A| = r_{11} \cdots r_{nn} \leq \|a_1\|_2 \cdots \|a_n\|_2.$$

If $\|a_i\|_2 \neq 0$ for every i, then the equality holds iff R is diagonal, which implies $A = QR$ has orthogonal columns. □

2.3 Unitary and real orthogonal triangularizations

9.

Proof. Let $U = [x\, U_1]$ be a unitary matrix whose first column is $x/\|x\|_2$. One can construct such a matrix explicitly via 2.3.P1. Note that $U^* x = \|x\|_2 e_1^{(n)}$, the first vector in the standard base of $\mathbb{C}^n$ and $U^* A U = \left[\begin{smallmatrix} \lambda & \bigstar \\ 0 & U_1^* A U_1 \end{smallmatrix}\right]$. And the eigenvalues of $U_1^* A U_1$ are $\lambda_2, \ldots, \lambda_n$.

Let $V = [1] \oplus U$. Then

$$V^* \mathcal{A} V = \begin{bmatrix} \alpha & y^* U \\ \|x\|_2 e_1^{(n)} & U^* A U \end{bmatrix} = \begin{bmatrix} B & \bigstar \\ 0 & U_1^* A U_1 \end{bmatrix}.$$

Let $D = \operatorname{diag}(1, \|x\|_2^{-1})$, then $DBD^{-1} = \left[\begin{smallmatrix} \alpha & y^* x \\ 1 & \lambda \end{smallmatrix}\right] = E$. Hence, they have the same eigenvalues. Therefore, the eigenvalues of $\mathcal{A}$ are the eigenvalues of E together with $\lambda_2, \ldots, \lambda_n$. If $y \perp x$, then $y^* x = 0$ and therefore the eigenvalues of E are just α and λ. □

5 Norms for Vectors and Matrices

5.1 Definitions of norms and inner products

1.

Proof. By the homogeneity and the triangular inequality of seminorm,

$$\|x\| = \left\| \sum_{i=1}^{n} x_i e_i \right\| \leq |x_i| \|e_i\|.$$

□

2.

Proof.

(a) Since V_0 and S are subspaces of V, 5.5.1.(1, 2, 3) holds. We only need to show the positivity. Clear that $\|0\| = 0$. Let $x \in S$ be any vector with norm zero. Then $x \in V_0$. Since $V_0 \cap S = \{0\}$, $x = 0$. Thus, $\|\cdot\|$ is a norm on S.

(b) Clear that $x \sim x$ and $x \sim y$ iff $y \sim x$. For the transitivity, suppose $\|x - y\| = \|y - z\| = 0$. Then by the triangular inequality

$$0 \leq \|x - z\| \leq \|x - y\| + \|y - z\| = 0.$$

Hence, $\sim$ is an equivalence relation.

Let $\hat{x}$ be the equivalence class containing x. For every $x, y \in \hat{x}$, since $\|x - y\| = 0$, $x - y \in V_0$.

Let $\hat{V}$ be the set consisting of all equivalence classes. Let $c \in \mathbb{F}$ and $\hat{x}, \hat{y} \in \hat{V}$. Define $c\hat{x} = \widehat{cx}$ and $\hat{x} + \hat{y} = \widehat{x + y}$. Now we shows that it is well-defined and therefore $\hat{V}$ is a vector space. Suppose $x_1, x_2 \in \hat{x}$. Since $c(x_1 - x_2) \in V_0$, $\widehat{cx_1} = \widehat{cx_2}$. Hence, the scalar multiplication is well-defined. Further suppose $y_1, y_2 \in \hat{y}$. Then $x_1 + y_1 - x_2 - y_2 = (x_1 - y_1) + (x_2 - y_2) \in V_0$. Hence, $\widehat{x_1 + y_1} = \widehat{x_2 + y_2}$. Namely, the addition is well-defined.

Clear that $\|\hat{x}\|$ is well-defined. By Lemma 5.1.2, $\|\hat{x}\|$ is actually a single value set so we may think it as a real-valued function. Meanwhile, it is a seminorm as $\|x\|$ is. And the argument in (a), *mutatis, mutandis*, gives the positivity. Hence, it is a norm on $\hat{V}$.

(c) For every vector seminorm, the norm described in (b) is a natural norm associated with it.

(d) Yes.

(e) Clear that it is a seminorm. And for every vector x orthogonal to z, not necessarily zero, $\|x\| = 0$. Hence, it is not a norm. The equivalence classes of hyperplanes orthogonal to z. □

4.

Proof.

(a) Some computation yields the identity. Consider x and y as two adjacent edges of a parallelogram. Then $x - y$ and $x + y$ are the two diagonals of the parallelogram. This gives a geometric explanation of the identity. □

6.

Proof.

$$\frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) = \frac{1}{2}(\langle x,y\rangle + \overline{\langle y,x\rangle}) = \mathrm{Re}\langle x,y\rangle.$$

By P4, $\|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 - \|x+y\|^2$ and therefore

$$\mathrm{Re}\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - 2\|x\|^2 - 2\|y\|^2 + \|x+y\|^2) = \frac{1}{2}(\|x+y\|^2 - \|x\|^2 - \|y\|^2).$$

□

10.

Proof. The homogeneity implies $\|0\| = 0$. For every x,

$$0 = \|x - x\| \le \|x\| + \|-x\| = 2\|x\|,$$

where the inequality comes from the triangular inequality. □

12.

Proof.

(a) Some computation gives $\langle x,x\rangle = \|x\|^2$ and the nonnegativity and positivity follows. Since $\langle x,y\rangle \in \mathbb{R}$, $\overline{\langle y,x\rangle} = \langle y,x\rangle$. And clear that it is symmetric. Thus, it satisfies axioms (1), (1a) and (4).

(b) It comes from some straightforward computation.

(c) For every nonnegative integer n and m, by the additivity,

$$\langle nx,y\rangle = \left\langle \sum_{i=1}^{n} x, y \right\rangle = n\langle x,y\rangle.$$

As $x = mx/m$, $m\langle m^{-1}nx,y\rangle = n\langle x,y\rangle$ and hence

$$\left\langle \frac{n}{m}x, y\right\rangle = \frac{n}{m}\langle x,y\rangle.$$

Meanwhile,

$$\begin{aligned}\langle -x,y\rangle &= \frac{1}{2}(\|x-y\|^2 - \|x\|^2 - \|y\|^2)\\ &= \frac{1}{2}(2\|x\|^2 + 2\|y\|^2 - \|x+y\|^2 - \|x\|^2 - \|y\|^2)\\ &= -\langle x,y\rangle.\end{aligned}$$

Hence, $\langle ax,y\rangle = a\langle x,y\rangle$ for every $a \in \mathbb{Q}$.

(d) For fixed x and y, let $p(t) = t^2\|x\|^2 + 2t\langle x,y\rangle + \|y\|^2$, a polynomial in t. For rational t, by (c), $p(t) = \|tx+y\|^2 \ge 0$. By the continuity of p, $p(t) \ge 0$ for all $t \in \mathbb{R}$. Hence, the discriminant of p is non-positive. Namely, $|\langle x,y\rangle|^2 \le \|x\|^2\|y\|^2$.

(e) Clear that the inequality holds for any $a \in \mathbb{R}$ and $b \in \mathbb{Q}$. Hence $|a-b|$ can be arbitrarily small, giving the homogeneity. Together with the previous parts, $\langle \cdot,\cdot\rangle$ is an inner product on V.

(f) We have show that $\mathrm{Re}\langle x, y\rangle$ is an inner product when the vector space is over $\mathbb{R}$. The nonnegativity and positivity comes from some computation and the additivity from the previous discussion. Finally,

$$\begin{aligned}\langle y, x\rangle &= \mathrm{Re}\langle x, y\rangle + \frac{i}{2}(\|ix + y\|^2 - \|ix\|^2 - \|y\|^2) \\ &= \mathrm{Re}\langle x, y\rangle + \frac{i}{2}(\|x - y\|^2 - \|x\|^2 - \|y\|^2) \\ &= \mathrm{Re}\langle x, y\rangle - \frac{i}{2}(\|x + iy\|^2 - \|x\|^2 - \|y\|^2) \\ &= \overline{\langle x, y\rangle},\end{aligned}$$

where the last equality comes from (5.1.9). Hence, $\langle\cdot,\cdot\rangle$ is an inner product on V. $\square$