7

Spectral Properties of Matrices

7.1 Introduction

The existence of directions that are preserved by linear transformations (which are referred to as eigenvectors) has been discovered by L. Euler in his study of movements of rigid bodies. This work was continued by Lagrange, Cauchy, Fourier, and Hermite. The study of eigenvectors and eigenvalues acquired increasing significance through its applications in heat propagation and stability theory. Later, Hilbert initiated the study of eigenvalue in functional analysis (in the theory of integral operators). He introduced the terms of eigenvalue and eigenvector. The term *eigenvalue* is a German-English hybrid formed from the German word *eigen* which means “own” and the English word “value”. It is interesting that Cauchy referred to the same concept as *characteristic value* and the term *characteristic polynomial* of a matrix (which we introduce in Definition 7.1) was derived from this naming.

We present the notions of geometric and algebraic multiplicities of eigenvalues, examine properties of spectra of special matrices, discuss variational characterizations of spectra and the relationships between matrix norms and eigenvalues. We conclude this chapter with a section dedicated to singular values of matrices.

7.2 Eigenvalues and Eigenvectors

Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. An *eigenpair* of A is a pair $(\lambda, \mathbf{x}) \in \mathbb{C} \times (\mathbb{C}^n - \{\mathbf{0}\})$ such that $A\mathbf{x} = \lambda\mathbf{x}$. We refer to λ is an *eigenvalue* and to $\mathbf{x}$ is an *eigenvector*. The set of eigenvalues of A is the *spectrum* of A and will be denoted by $\mathsf{spec}(A)$.

If $(\lambda, \mathbf{x})$ is an eigenpair of A, the linear system $A\mathbf{x} = \lambda\mathbf{x}$ has a non-trivial solution in $\mathbf{x}$. An equivalent homogeneous system is $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$ and this system has a non-trivial solution only if $\det(\lambda I_n - A) = 0$.

Definition 7.1. *The* characteristic polynomial *of the matrix A is the polynomial p_A defined by $p_A(\lambda) = \det(\lambda I_n - A)$ for $\lambda \in \mathbb{C}$.*

Thus, the eigenvalues of A are the roots of the characteristic polynomial of A.

Lemma 7.2. *Let $A = (\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_n) \in \mathbb{C}^n$ and let B be the matrix obtained from A by replacing the column $\boldsymbol{a}_j$ by $\mathbf{e}_j$. Then, we have*

$$\det(B) = \det\left(A\begin{bmatrix} 1 & \cdots & j-1 & j+1 & \cdots & n \\ 1 & \cdots & j-1 & j+1 & \cdots & n \end{bmatrix}\right).$$

Proof. The result follows immediately by expanding B on the j^{th} column.

The result obtained in Lemma 7.2 can be easily extended as follows. If B is the matrix obtained from A by replacing the columns $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_k}$ by $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}$ and $\{i_1, \ldots, i_p\} = \{1, \ldots, n\} - \{j_1, \ldots, j_k\}$, then

$$\det(B) = \det\left(A\begin{bmatrix} i_1 & \cdots & i_p \\ i_1 & \cdots & i_p \end{bmatrix}\right). \tag{7.1}$$

In other words, $\det(B)$ equals a principal p-minor of A.

Theorem 7.3. *Let $A \in \mathbb{C}^{n \times n}$ be a matrix. Its characteristic polynomial p_A can be written as*

$$p_A(\lambda) = \sum_{k=0}^{n} (-1)^k a_k \lambda^{n-k},$$

where a_k is the sum of the principal minors of order k of A.

Proof. By Theorem 5.146 the determinant

$$p_A(\lambda) = \det(\lambda I_n - A) = (-1)^n \det(A - \lambda I_n)$$

can be written as a sum of 2^n determinants of matrices obtained by replacing each subset of the columns of A by the corresponding subset of columns of $-\lambda I_n$. If the subset of columns of $-\lambda I_n$ involved are $-\lambda\mathbf{e}_{j_1}, \ldots, -\lambda\mathbf{e}_{j_k}$ the result of the substitution is $(-1)^k \lambda^k \det\left(A\begin{bmatrix} i_1 \cdots i_p \\ i_1 \cdots i_p \end{bmatrix}\right)$, where $\{i_1, \ldots, i_p\} = \{1, \ldots, n\} - \{j_1, \ldots, j_k\}$. The total contribution of sets of k columns of $-\lambda I_n$ is $(-1)^k \lambda^k a_{n-k}$. Therefore,

$$p_A(\lambda) = (-1)^n \sum_{k=0}^{n} (-1)^k \lambda^k a_{n-k}.$$

Replacing k by $n-k$ as the summation index yields

$$p_A(\lambda) = (-1)^n \sum_{k=0}^{n} (-1)^{n-k} \lambda^{n-k} a_k = \sum_{k=0}^{n} (-1)^k a_k \lambda^{n-k}.$$

Definition 7.4. *Two matrices $A, B \in \mathbb{C}^{n\times n}$ are* similar *if there exists an invertible matrix $P \in \mathbb{C}^{n\times n}$ such that $B = PAP^{-1}$. This is denoted by $A \sim B$.*

If there exists a unitary matrix U such that $B = UAU^{-1}$, then A is unitarily similar *to B. This is denoted by $A \sim_u B$.*

The matrices A, B are congruent *if $B = SAS^{\mathsf{H}}$ for some non-singular matrix S. This is denoted by $A \approx B$. If $A, B \in \mathbb{R}^{n\times n}$, we say that they are* t-congruent *if $B = SAS'$ for some invertible matrix S; this is denoted by $A \approx_t B$.*

For real matrices the notions of t-congruence and congruence are identical.

It is easy to verify that $\sim, \sim_u$ and $\approx$ are equivalence relations on $\mathbb{C}^{n\times n}$ and $\approx_t$ is an equivalence on $\mathbb{R}^{n\times n}$.

Similar matrices have the same characteristic polynomial. Indeed, suppose that $B = PAP^{-1}$. We have

$$\begin{aligned}p_B(\lambda) &= \det(\lambda I_n - B) = \det(\lambda I_n - PAP^{-1})\\ &= \det(\lambda P I_n P^{-1} - PAP^{-1}) = \det(P(\lambda I_n - A)P^{-1})\\ &= \det(P)\det(\lambda I_n - A)\det(P^{-1}) = \det(\lambda I_n - A) = p_A(\lambda),\end{aligned}$$

because $\det(P)\det(P^{-1}) = 1$. Thus, similar matrices have the same eigenvalues.

Example 7.5. Let A be the matrix

$$A = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

We have

$$p_A = \det(\lambda I_2 - A) = (\lambda - \cos\theta)^2 + \sin^2\theta = \lambda^2 - 2\lambda\cos\theta + 1.$$

The roots of this polynomial are $\lambda_1 = \cos\theta + i\sin\theta$ and $\lambda_2 = \cos\theta - i\sin\theta$, so they are complex numbers.

We regard A as a complex matrix with real entries. If we were to consider A as a real matrix, we would not be able to find real eigenvalues for A unless θ were equal to 0.

Definition 7.6. *The* algebraic multiplicity *of the eigenvalue λ of a matrix $A \in \mathbb{C}^{n\times n}$ is the multiplicity of λ as a root of the characteristic polynomial p_A of A.*

The algebraic multiplicity of λ is denoted by $\mathsf{algm}(A, \lambda)$. If $\mathsf{algm}(A, \lambda) = 1$ we say that λ is a simple eigenvalue*.*

Example 7.7. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix

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

The characteristic polynomial of A is

$$p_A(\lambda) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda - 2 \end{vmatrix} = \lambda^2 - 2\lambda + 1.$$

Therefore, A has the eigenvalue 1 with $\mathsf{algm}(A, 1) = 2$.

Example 7.8. Let $P(a) \in \mathbb{C}^{n\times n}$ be the matrix $P(a) = (a-1)I_n + J_n$. To find the eigenvalues of $P(a)$ we need to solve the equation

$$\begin{vmatrix} \lambda - a & -1 & \cdots & -1 \\ -1 & \lambda - a & \cdots & -1 \\ \vdots & \vdots & \cdots & \vdots \\ -1 & -1 & \cdots & \lambda - a \end{vmatrix} = 0.$$

By adding the first $n-1$ columns to the last and factoring out $\lambda - (a+n-1)$, we obtain the equivalent equation

$$(\lambda - (a+n-1)) \begin{vmatrix} \lambda - a & -1 & \cdots & 1 \\ -1 & \lambda - a & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ -1 & -1 & \cdots & 1 \end{vmatrix} = 0.$$

Adding the last column to the first $n-1$ columns and expanding the determinant yields the equation $(\lambda - (a+n-1))(\lambda - a + 1)^{n-1} = 0$, which shows that $P(a)$ has the eigenvalue $a+n-1$ with $\mathsf{algm}(P(a), a+n-1) = 1$ and the eigenvalue $a-1$ with $\mathsf{algm}(P(a), a-1) = n-1$.

In the special case when $a = 1$ we have $P(1) = J_{n,n}$. Thus, $J_{n,n}$ has the eigenvalue $\lambda_1 = n$ with algebraic multiplicity 1 and the eigenvalue 0 with algebraic multiplicity $n-1$.

Theorem 7.9. *The eigenvalues of Hermitian complex matrices are real numbers.*

Proof. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let λ be an eigenvalue of A. We have $A\mathbf{x} = \lambda\mathbf{x}$ for some $\mathbf{x} \in \mathbb{C}^n - \{\mathbf{0}_n\}$, so $\mathbf{x}^{\mathsf{H}} A^{\mathsf{H}} = \overline{\lambda}\mathbf{x}^{\mathsf{H}}$. Since $A^{\mathsf{H}} = A$, we have

$$\lambda \mathbf{x}^{\mathsf{H}}\mathbf{x} = \mathbf{x}^{\mathsf{H}} A\mathbf{x} = \mathbf{x}^{\mathsf{H}} A^{\mathsf{H}}\mathbf{x} = \overline{\lambda}\mathbf{x}^{\mathsf{H}}\mathbf{x}.$$

Since $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{x}^{\mathsf{H}}\mathbf{x} \neq 0$, it follows that $\overline{\lambda} = \lambda$. Thus, λ is a real number.

Corollary 7.10. *The eigenvalues of symmetric real matrices are real numbers.*

Proof. This is a direct consequence of Theorem 7.9.

Theorem 7.11. *The eigenvectors of a complex Hermitian matrix corresponding to distinct eigenvalues are orthogonal to each other.*

Proof. Let $(\lambda, \mathbf{u})$ and $(\mu, \mathbf{v})$ be two eigenpairs of the Hermitian matrix $A \in \mathbb{C}^{n\times n}$, where $\lambda \neq \mu$. Since A is Hermitian, $\lambda, \mu \in \mathbb{R}$. Since $A\mathbf{u} = \lambda\mathbf{u}$ we have $\mathbf{v}^{\mathsf{H}}A\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$. The last equality can be written as $(A\mathbf{v})^{\mathsf{H}}\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$, or as $\mu\mathbf{v}^{\mathsf{H}}\mathbf{u} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{u}$. Since $\mu \neq \lambda$, $\mathbf{v}^{\mathsf{H}}\mathbf{u} = 0$, so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

The statement clearly holds if we replace complex Hermitian matrices by real symmetric matrices.

Corollary 7.12. *The eigenvectors of a Hermitian matrix corresponding to distinct eigenvalues form a linearly independent set.*

Proof. This statement follows from Theorems 6.41 and 7.11.

The next statement is a result of Issai Schur (1875-1941), a mathematician born in Russia, who studied and worked in Germany.

Theorem 7.13. (Schur's Triangularization Theorem) *Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. There exists an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A =\sim_u T$.*

The diagonal elements of T are the eigenvalues of A; moreover, each eigenvalue λ of A occurs in the sequence of diagonal elements of T a number of $\mathsf{algm}(A, \lambda)$ times. The columns of U are unit eigenvectors of A.

Proof. The argument is by induction on n. The base case, $n = 1$, is immediate.

Suppose that the statement holds for matrices in $\mathbb{C}^{(n-1)\times(n-1)}$ and let $A \in \mathbb{C}^{n\times n}$. If $(\lambda, \mathbf{x})$ is an eigenpair of A with $\parallel \mathbf{x} \parallel_2 = 1$, let $H_{\mathbf{v}}$ be a Householder matrix that transforms $\mathbf{x}$ into $\mathbf{e}_1$. Since we also have $H_{\mathbf{v}}\mathbf{e}_1 = \mathbf{x}$, $\mathbf{x}$ is the first column of $H_{\mathbf{v}}$ and we can write $H_{\mathbf{v}} = (\mathbf{x}\ K)$, where $K \in \mathbb{C}^{n\times(n-1)}$. Consequently,

$$H_{\mathbf{v}}AH_{\mathbf{v}} = H_{\mathbf{v}}A(\mathbf{x}\ K) = H_{\mathbf{v}}(\lambda\mathbf{x}\ H_{\mathbf{v}}AK) = (\lambda\mathbf{e}_1\ H_{\mathbf{v}}AK).$$

Since $H_{\mathbf{v}}$ is Hermitian and $H_{\mathbf{v}} = (\mathbf{x}\ K)$, it follows that

$$H_{\mathbf{v}}^{\mathsf{H}} = \begin{pmatrix} \mathbf{x}^{\mathsf{H}} \\ K^{\mathsf{H}} \end{pmatrix} = H_{\mathbf{v}}.$$

Therefore,

$$H_{\mathbf{v}}AH_{\mathbf{v}} = \begin{pmatrix} \lambda & \mathbf{x}^{\mathsf{H}}AK \\ \mathbf{0}_{n-1} & K^{\mathsf{H}}AK \end{pmatrix}.$$

Since $K^{\mathsf{H}}AK \in \mathbb{C}^{(n-1)\times(n-1)}$, by the inductive hypothesis, there exists a unitary matrix W and an upper triangular matrix S such that $W^{\mathsf{H}}(K^{\mathsf{H}}AK)W = S$. Note that the matrix

$$U = H_{\mathbf{v}} \begin{pmatrix} 1 & \mathbf{0}'_{n-1} \\ \mathbf{0}_{n-1} & W \end{pmatrix}$$

is unitary and

$$U^{\mathsf{H}} A U^{\mathsf{H}} = \begin{pmatrix} \lambda & x^{\mathsf{H}} AKW \\ \mathbf{0}_{n-1} & W^{\mathsf{H}} K^{\mathsf{H}} AKW \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{x}^{\mathsf{H}} AKW \\ \mathbf{0}_{n-1} & S \end{pmatrix}.$$

The last matrix is clearly upper triangular.

Since $A \sim_u T$, A and T have the same characteristic polynomials and, therefore, the same eigenvalues, with identical multiplicities. Note that the factorization of A can be written as $A = UDU^{\mathsf{H}}$ because $U^{-1} = U^{\mathsf{H}}$. Since $AU = UD$, each column $\mathbf{u}_i$ of U is an eigenvector of A that corresponds to the eigenvalue λ_i for $1 \leqslant i \leqslant n$.

Corollary 7.14. *If $A \in \mathbb{R}^{n \times n}$ is a matrix such that $\mathsf{spec}(A) = \{0\}$, then A is nilpotent.*

Proof. By Schur's Triangularization Theorem, A is unitarily similar to a strictly upper triangular matrix, $A = UTU^{\mathsf{H}}$, so $A^n = UT^nU^{\mathsf{H}}$. Since $\mathsf{spec}(T) = \{0\}$, we have $T^n = O$, so $A^n = O$.

Corollary 7.15. *Let $A \in \mathbb{C}^{n \times n}$ and let f be a polynomial. If $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$ (including multiplicities), then $\mathsf{spec}(f(A)) = \{f(\lambda_1), \ldots, f(\lambda_n)\}$.*

Proof. By Schur's Triangularization Theorem there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n \times n}$ such that $A = UTU^{\mathsf{H}}$ and the diagonal elements of T are the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$. Therefore $f(A) = Uf(T)U^{\mathsf{H}}$, and by Theorem 5.49, the diagonal elements of $f(T)$ are $f(\lambda_1), , \ldots, f(\lambda_m)$. Since $f(A) \sim_u f(T)$, we obtain the desired conclusion because two similar matrices have the same eigenvalues with the same algebraic multiplicities.

Definition 7.16. *A matrix $A \in \mathbb{C}^{n \times n}$ is* diagonalizable *(unitarily diagonalizable) if there exists a diagonal matrix $D = \mathsf{diag}(d_1, \ldots, d_n)$ such that $A \sim D$ ($A \sim_u D$).*

Theorem 7.17. *A matrix $A \in \mathbb{C}^{n \times n}$ is diagonalizable if and only if there exists a linearly independent set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of n eigenvectors of A.*

Proof. Let $A \in \mathbb{C}^{n \times n}$ such that there exists a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of n eigenvectors of A that is linearly independent and let P be the matrix $(\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n)$ that is clearly invertible. We have:

$$\begin{aligned} P^{-1}AP &= P^{-1}(A\mathbf{v}_1 \ A\mathbf{v}_2 \ \cdots \ A\mathbf{v}_n) = P^{-1}(\lambda_1\mathbf{v}_1 \ \lambda_2\mathbf{v}_2 \ \cdots \ \lambda_n\mathbf{v}_n) \\ &= P^{-1}P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \end{aligned}$$

Therefore, we have $A = PDP^{-1}$, where

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

so $A \sim D$.

Conversely, suppose that A is diagonalizable, so $AP = PD$, where D is a diagonal matrix and P is an invertible matrix, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of the matrix P. We have $A\mathbf{v}_i = d_{ii}\mathbf{v}_i$ for $1 \leqslant i \leqslant n$, so each $\mathbf{v}_i$ is an eigenvector of A. Since P is invertible, its columns are linear independent.

Corollary 7.18. *If $A \in \mathbb{C}^{n\times n}$ is diagonalizable then the columns of any matrix P such that $D = P^{-1}AP$ is a diagonal matrix are eigenvectors of A. Furthermore, the diagonal entries of D are the eigenvalues that correspond to the columns of P.*

Proof. This statement follows from the proof of Theorem 7.17.

Corollary 7.19. *A matrix $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable if and only if there exists a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of n orthonormal eigenvectors of A.*

Proof. This statement follows from the proof of Theorem 7.17 by observing that if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set n orthonormal eigenvectors of A, then $P = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a unitary matrix that gives a unitary diagonalization of A. Conversely, if P is an unitary matrix such that $A = PDP^{-1}$ its set of columns consists of orthogonal unitary eigenvectors of A.

Corollary 7.20. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. There exists a orthonormal matrix U and a diagonal matrix T such that $A = UTU^{-1}$. The diagonal elements of T are the eigenvalues of A; moreover, each eigenvalue λ of A occurs in the sequence of diagonal elements of T a number of $\mathsf{algm}(A, \lambda)$ times.*

Proof. As the previous corollary, this follows from the proof of Theorem 7.17.

Theorem 7.21. *Let $A \in \mathbb{C}^{n\times n}$ be a block diagonal matrix,*

$$A = \begin{pmatrix} A_{11} & O & \cdots & O \\ O & A_{22} & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & A_{mm} \end{pmatrix}.$$

A is diagonalizable if and only if every matrix A_{ii} is diagonalizable for $1 \leqslant i \leqslant m$.

Proof. Suppose that A is a block diagonal matrix which is diagonalizable. Furthermore, suppose that $A_{ii} \in \mathbb{C}^{n_i \times n_i}$ for $1 \leqslant i \leqslant n$ and $\sum_{i=1}^{m} n_i = n$. There exists an invertible matrix $P \in \mathbb{C}^{n \times n}$ such that $P^{-1}AP$ is a diagonal matrix $D = \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$. Let $\mathbf{p}_1, \ldots, \mathbf{p}_n$ be the columns of P, which are eigenvectors of A. Each vector $\mathbf{p}_i$ is divided into m blocks $\mathbf{p}_i^j$ with $1 \leqslant j \leqslant m$, where $\mathbf{p}_i^j \in \mathbb{C}^{n_j}$. Thus, P can be written as

$$P = \begin{pmatrix} \mathbf{p}_1^1 & \mathbf{p}_2^1 & \cdots & \mathbf{p}_n^1 \\ \mathbf{p}_1^2 & \mathbf{p}_2^2 & \cdots & \mathbf{p}_n^2 \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{p}_1^m & \mathbf{p}_2^m & \cdots & \mathbf{p}_n^m \end{pmatrix}.$$

The equality $A\mathbf{p}_i = \lambda_i \mathbf{p}_i$ can be expressed as

$$\begin{pmatrix} A_{11} & O & \cdots & O \\ O & A_{22} & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & A_{mm} \end{pmatrix} \begin{pmatrix} \mathbf{p}_i^1 \\ \mathbf{p}_i^2 \\ \vdots \\ \mathbf{p}_i^m \end{pmatrix} = \lambda_i \begin{pmatrix} \mathbf{p}_i^1 \\ \mathbf{p}_i^2 \\ \vdots \\ \mathbf{p}_i^m \end{pmatrix},$$

which shows that $A_{jj}\mathbf{p}_i^j = \lambda_i \mathbf{p}_i^j$ for $1 \leqslant j \leqslant m$. Let $M^j = (\mathbf{p}_1^j\ \mathbf{p}_2^j\ \cdots\ \mathbf{p}_n^j) \in \mathbb{C}^{n_j \times n}$. We claim that $\mathit{rank}(M^j) = n_j$. Indeed if $\mathit{rank}(M^j)$ were less than n_j, we would have fewer that n independent rows M^j for $1 \leqslant j \leqslant m$. This, however, would imply that the rank of P is less then n, which contradicts the invertibility of P. Since there are n_j linearly independent eigenvectors of A_{jj}, it follows that each block A_{jj} is diagonalizable.

Conversely, suppose that each A_{jj} is diagonalizable, that is, there exists a invertible matrix Q_j such that $Q_j^{-1} A_{jj} Q_j$ is a diagonal matrix. Then, it is immediate to verify that the block diagonal matrix

$$Q = \begin{pmatrix} Q_1 & O & \cdots & O \\ O & Q_2 & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & Q_m \end{pmatrix}$$

is invertible and $Q^{-1}AQ$ is a diagonal matrix.

Theorem 7.22. *Let $A \in \mathbb{C}^{n \times n}$ be a matrix and let $(\lambda_1, \boldsymbol{v}_1), \ldots, (\lambda_k, \boldsymbol{v}_k)$ be k eigenpairs of A, where $\lambda_1, \ldots, \lambda_k$ are pairwise distinct. Then, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a linearly independent set.*

Proof. Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is not linearly independent. Then, there exists a linear combination of this set that equals $\mathbf{0}$, that is

$$c_1 \mathbf{v}_{i_1} + \cdots + c_r \mathbf{v}_{i_r} = \mathbf{0}_n, \tag{7.2}$$

at least one coefficient is not 0, and $r > 1$. Choose this linear combination to involve the minimal number of terms. We have

$$A(c_1\mathbf{v}_{i_1} + c_2\mathbf{v}_{i_2} + \cdots + c_r\mathbf{v}_{i_r}) = c_1\lambda_{i_1}\mathbf{v}_1 + c_2\lambda_{i_2}\mathbf{v}_{i_2} + \cdots + c_r\lambda_{i_r}\mathbf{v}_{i_r} = \mathbf{0}_n.$$

By multiplying Equality (7.2) by λ_{i_1} we have

$$c_1\lambda_{i_1}\mathbf{v}_{i_1} + c_2\lambda_{i_1}\mathbf{v}_{i_2} + \cdots + c_r\lambda_{i_1}\mathbf{v}_{i_r} = \mathbf{0}_n.$$

It follows that $c_2(\lambda_{i_2} - \lambda_{i_1})\mathbf{v}_{i_2} + \cdots + c_r(\lambda_{i_r} - \lambda_{i_1})\mathbf{v}_{i_r} = \mathbf{0}_n$. Since there exists a coefficient $\lambda_{i_p} - \lambda_{i_1}$ that is non-zero, this contradicts the minimality of r.

Corollary 7.23. *If $A \in \mathbb{C}^{n\times n}$ has n distinct eigenvalues, then A is diagonalizable.*

Proof. If $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$ consists of n complex numbers and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are corresponding eigenvectors, then, by Theorem 7.22, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent. The statement follows immediately from Theorem 7.17.

7.3 Geometric and Algebraic Multiplicities of Eigenvalues

For a matrix $A \in \mathbb{C}^{n\times n}$ let $S_{A,\lambda}$ be the subspace $\mathsf{NullSp}(\lambda I_n - A)$. We refer to $S_{A,\lambda}$ as *invariant subspace of A and λ* or as the *eigenspace* of λ.

Definition 7.24. *Let $A \in \mathbb{C}^{n\times n}$ and let $\lambda \in \mathsf{spec}(A)$. The* geometric multiplicity *of λ is the dimension $\mathsf{geomm}(A, \lambda)$ of $S_{A,\lambda}$.*

Example 7.25. The geometric multiplicity of 1 as an eigenvalue of the matrix

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

considered in Example 7.7, is 1. Indeed if $\mathbf{x}$ is an eigenvector that corresponds to this value we have $-x_2 = x_1$ and $x_1 + 2x_2 = x_2$, which means that any such eigenvector has the form $a\mathbf{1}_2$. Thus, $\dim(S_{A,1}) = 1$.

The definition of the geometric multiplicity of $\lambda \in \mathsf{spec}(A)$ implies

$$\mathsf{geomm}(A, \lambda) = \dim(\mathsf{NullSp}(A - \lambda I_n)) = n - \mathit{rank}(A - \lambda I_n). \tag{7.3}$$

Theorem 7.26. *Let $A \in \mathbb{R}^{n\times n}$. We have $0 \in \mathsf{spec}(A)$ if and only if A is a singular matrix. Moreover, in this case, $\mathsf{geomm}(A, 0) = n - \mathit{rank}(A) = \dim(\mathsf{NullSp}(A))$. If $\mathsf{algm}(A, 0) = 1$, then $\mathit{rank}(A) = n - 1$.*

Proof. The statement is an immediate consequence of Equality (7.3).

Theorem 7.27. *Let $A \in \mathbb{C}^{n\times n}$ be a square matrix and let $\lambda \in \mathsf{spec}(A)$. We have $\mathsf{geomm}(A,\lambda) \leqslant \mathsf{algm}(A,\lambda)$.*

Proof. By the definition of $\mathsf{geomm}(A,\lambda)$ we have

$$\mathsf{geomm}(A,\lambda) = \dim(\mathsf{NullSp}(\lambda I_n - A)) = n - rank(\lambda I_n - A).$$

Let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be an orthonormal basis of $S_{A,\lambda}$, where $m = \mathsf{geomm}(A,\lambda)$ and let $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m)$. We have $(\lambda I_n - A)U = O_{n,n}$, so $AU = \lambda U = U(\lambda I_m)$. Thus, by Theorem 6.57 there exists a matrix V such that we have

$$A \sim \begin{pmatrix} \lambda I_m & U^{\mathsf{H}}AV \\ O & V^{\mathsf{H}}AV \end{pmatrix},$$

where $U \in \mathbb{C}^{n\times m}$ and $V \in \mathbb{C}^{n\times(n-m)}$. Therefore, A has the same characteristic polynomial as

$$B = \begin{pmatrix} \lambda I_m & U^{\mathsf{H}}AV \\ O & V^{\mathsf{H}}AV \end{pmatrix},$$

which implies $\mathsf{algm}(A,\lambda) = \mathsf{algm}(B,\lambda)$. Since the algebraic multiplicity of λ in B is at least equal to m it follows that $\mathsf{algm}(A,\lambda) \geqslant m = \mathsf{geomm}(A,\lambda)$.

Definition 7.28. *An eigenvalue λ of a matrix A is* simple *if $\mathsf{algm}(A,\lambda) = 1$. If $\mathsf{geomm}(A,\lambda) = \mathsf{algm}(A,\lambda)$, then we refer to λ as a* semisimple eigenvalue.

The matrix A is defective *if there exists at least one eigenvalue that is not semisimple. Otherwise, A is said to be* non-defective.

A is a non-derogatory *matrix if $\mathsf{geomm}(A,\lambda) = 1$ for every eigenvalue λ.*

Note that if λ is a simple eigenvalue of A, then $\mathsf{geomm}(A,\lambda) = \mathsf{algm}(A,\lambda) = 1$, so λ is semi-simple.

Theorem 7.29. *Each eigenvalue of a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is semi-simple.*

Proof. We saw that each symmetric matrix has real eigenvalues and is orthonormally diagonalizable (by Corollary 7.20). Starting from the real Schur factorization $A = UTU^{-1}$, where U is an orthonormal matrix and $T = \mathsf{diag}(t_{11}, \ldots, t_{nn})$ is a diagonal matrix we can write $AU = UT$. If we denote the columns of U by $\mathbf{u}_1, \ldots, \mathbf{u}_n$, then we can write

$$(A\mathbf{u}_1, \ldots, A\mathbf{u}_n) = (t_{11}\mathbf{u}_1, \ldots, t_{nn}\mathbf{u}_n),$$

so $A\mathbf{u}_i = t_{ii}\mathbf{u}_i$ for $1 \leqslant i \leqslant n$. Thus, the diagonal elements of T are the eigenvalues of A and the columns of U are corresponding eigenvectors. Since these eigenvectors are pairwise orthogonal, the dimension of the invariant subspace that corresponds to an eigenvalue equals the algebraic multiplicity of the eigenvalue, so each eigenvalue is semi-simple.

Example 7.30. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix

$$A = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix},$$

where $a, b \in \mathbb{R}$ are such that $ab \neq 0$. The characteristic polynomial of A is $p_A(\lambda) = (a-\lambda)^2$, so $\mathsf{spec}(A) = \{a\}$ and $\mathsf{algm}(A, a) = 2$.

Let $\mathbf{x}$ be a characteristic vector of A that corresponds to a. We have $ax_1 + bx_2 = ax_1$ and $ax_2 = x_2$, which implies $x_2 = 0$. Thus, the invariant subspace is

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \middle| x_2 = 0 \right\},$$

which is one-dimensional, so $\mathsf{geomm}(A, a) = 1$. Thus, a is not semi-simple.

7.4 Spectra of Special Matrices

Theorem 7.31. *Let A be an block upper triangular partitioned matrix given by*

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ O & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & A_{mm} \end{pmatrix},$$

where $A_{ii} \in \mathbb{R}^{p_i \times p_i}$ for $1 \leqslant i \leqslant m$. Then, $\mathsf{spec}(A) = \bigcup_{i=1}^m \mathsf{spec}(A_{ii})$.

If A is a block lower triangular matrix

$$A = \begin{pmatrix} A_{11} & O & \cdots & O \\ A_{21} & A_{22} & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{pmatrix},$$

the same equality holds.

Proof. Let A be a block upper triangular matrix. Its characteristic equation is $\det(\lambda I_n - A) = 0$. Observe that the matrix $\lambda I_n - A$ is also an block upper triangular matrix:

$$\lambda I_n - A = \begin{pmatrix} \lambda I_{p_1} - A_{11} & O & \cdots & O \\ -A_{21} & \lambda I_{p_2} - A_{22} & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ -A_{m1} & -A_{m2} & \cdots & \lambda I_{p_m} - A_{mm} \end{pmatrix}.$$

By Theorem 5.153 the characteristic polynomial of A can be written as

$$p_A(\lambda) = \prod_{i=1}^{m} \det(\lambda I_{p_i} - A_{ii}) = \prod_{i=1}^{m} p_{A_{ii}}(\lambda).$$

Therefore, $\mathsf{spec}(A) = \bigcup_{i=1}^{m} \mathsf{spec}(A_{ii})$.

The argument for block lower triangular matrices is similar.

Corollary 7.32. *Let $A \in \mathbb{R}^{n \times n}$ be a block diagonal matrix given by*

$$A = \begin{pmatrix} A_{11} & O & \cdots & O \\ O & A_{22} & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & A_{mm} \end{pmatrix},$$

where $A_{ii} \in \mathbb{R}^{n_i \times n_i}$ for $1 \leqslant i \leqslant m$. We have $\mathsf{spec}(A) = \bigcup_{i=1}^{m} \mathsf{spec}(A_{ii})$ and $\mathsf{algm}(A, \lambda) = \sum_{i=1}^{m} \mathsf{algm}(A_i, \lambda)$. Moreover, $\boldsymbol{v} \neq \boldsymbol{0}_n$ is an eigenvector of A if and only if we can write

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_m \end{pmatrix},$$

where each vector $\boldsymbol{v}_i$ is either an eigenvector of A_i or $\boldsymbol{0}_{n_i}$ for $1 \leqslant i \leqslant m$ and there exists i such that $\boldsymbol{v}_i \neq \boldsymbol{0}_{n_i}$.

Proof. This statement follows immediately from Theorem 7.31.

Theorem 7.33. *Let $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ be an upper (lower) triangular matrix. Then, $\mathsf{spec}(A) = \{a_{ii} \mid 1 \leqslant i \leqslant n\}$.*

Proof. It is easy to see that the the characteristic polynomial of A is $p_A(\lambda) = (\lambda - a_{11}) \cdots (\lambda - a_{nn})$, which implies immediately the theorem.

Corollary 7.34. *If $A \in \mathbb{C}^{n \times n}$ is an upper triangular matrix and λ is an eigenvalue such that the diagonal entries of A that equal λ occur in $a_{i_1 i_1}, \ldots, a_{i_p i_p}$, then $S_{A,\lambda}$ is a p-dimensional subspace of $\mathbb{C}^n$ generated by $\boldsymbol{e}_{i_1}, \ldots \boldsymbol{e}_{i_p}$.*

Proof. This statement is immediate.

Corollary 7.35. *We have*

$$\mathsf{spec}(\mathsf{diag}(d_1, \ldots, d_n)) = \{d_1, \ldots, d_n\}.$$

Proof. This statement is a direct consequence of Theorem 7.33.

Note that if $\lambda \in \mathsf{spec}(A)$ we have $A\mathbf{x} = \lambda\mathbf{x}$, $A^2\mathbf{x} = \lambda^2\mathbf{x}$ and, in general, $A^k\mathbf{x} = \lambda^k\mathbf{x}$ for $k \geqslant 1$. Thus, $\lambda \in \mathsf{spec}(A)$ implies $\lambda^k \in \mathsf{spec}(A^k)$ for $k \geqslant 1$.

Theorem 7.36. *If $A \in \mathbb{C}^{n \times n}$ is a nilpotent matrix, then $\mathsf{spec}(A) = \{0\}$.*

Proof. Let $A \in \mathbb{C}^{n \times n}$ be a nilpotent matrix such that $\mathsf{nilp}(A) = k$. By a previous observation if $\lambda \in \mathsf{spec}(A)$, then $\lambda^k \in \mathsf{spec}(A^k) = \mathsf{spec}(O_{n,n}) = \{0\}$. Thus, $\lambda = 0$.

Theorem 7.37. *If $A \in \mathbb{C}^{n \times n}$ is an idempotent matrix, then $\mathsf{spec}(A) \subseteq \{0, 1\}$.*

Proof. Let $A \in \mathbb{C}^{n \times n}$ be an idempotent matrix, λ be an eigenvalue of A, and let $\mathbf{x}$ be an eigenvector of λ. We have $P^2\mathbf{x} = P\mathbf{x} = \lambda\mathbf{x}$; on another hand, $P^2\mathbf{x} = P(P\mathbf{x}) = P(\lambda\mathbf{x}) = \lambda P(\mathbf{x}) = \lambda^2\mathbf{x}$, so $\lambda^2 = \lambda$, which means that $\lambda \in \{0, 1\}$.

Theorem 7.38. *At least one eigenvalue of a stochastic matrix is equal to* 1 *and all eigenvalues lie on or inside the unit circle.*

Proof. Let $A \in \mathbb{R}^{n \times n}$ be a stochastic matrix. Then, $1 \in \mathsf{spec}(A)$ and $\mathbf{1}$ is an eigenvector that corresponds to the eigenvalue 1 as the reader can easily verify.

If λ is an eigenvalue of A and $A\mathbf{x} = \lambda\mathbf{x}$, then $\lambda x_i = \sum_{i=1}^{n} a_{ij} x_j$ for $1 \leqslant n \leqslant n$, which implies

$$|\lambda||x_i| \leqslant \sum_{i=1}^{n} a_{ij}|x_j|.$$

Since $\mathbf{x} \neq \mathbf{0}$, let x_p be a component of $\mathbf{x}$ such that $|x_p| = \max\{|x_i| \mid 1 \leqslant i \leqslant n\}$. Choosing $i = p$ we have

$$|\lambda| \leqslant \sum_{i=1}^{n} a_{ij} \frac{|x_j|}{|x_p} \leqslant \sum_{i=1}^{n} a_{ij} = 1,$$

which shows that all eigenvalues of A lie on or inside the unit circle.

Theorem 7.39. *All eigenvalues of a unitary matrix are located on the unit circle.*

Proof. Let $A \in \mathbb{R}^{n \times n}$ be an unitary matrix and let λ be an eigenvalue of A. By Theorem 6.86, if $\mathbf{x}$ is an eigenvector that corresponds to λ we have

$$\parallel \mathbf{x} \parallel = \parallel A\mathbf{x} \parallel = \parallel \lambda\mathbf{x} \parallel = |\lambda| \parallel \mathbf{x} \parallel,$$

which implies $|\lambda| = 1$.

Next, we show that unitary diagonalizability is a characteristic property of normal matrices.

Theorem 7.40. (Spectral Theorem for Normal Matrices) *A matrix $A \in \mathbb{C}^{n \times n}$ is normal if and only if there exists a unitary matrix U and a diagonal matrix D such that*

$$A = UDU^{\mathsf{H}}, \tag{7.4}$$

the columns of U are unit eigenvectors and the diagonal elements of D are the eigenvalues of A that correspond to these eigenvectors.

Proof. Suppose that A is a normal matrix. By Schur's Triangularization Theorem there exists a unitary matrix $U \in \mathbb{C}^{n \times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n \times n}$ such that $A = UTU^{-1}$. Thus, $T = U^{-1}AU = U^{\mathsf{H}}AU$ and $T^{\mathsf{H}} = U^{\mathsf{H}}A^{\mathsf{H}}U$. Therefore,

$$\begin{aligned} T^{\mathsf{H}}T &= U^{\mathsf{H}}A^{\mathsf{H}}UU^{\mathsf{H}}AU = U^{\mathsf{H}}A^{\mathsf{H}}AU \\ &\quad \text{(because } U \text{ is unitary)} \\ &= U^{\mathsf{H}}AA^{\mathsf{H}}U \\ &\quad \text{(because } A \text{ is normal)} \\ &= U^{\mathsf{H}}AUU^{\mathsf{H}}A^{\mathsf{H}}U = TT^{\mathsf{H}}, \end{aligned}$$

so T is a normal matrix. By Theorem 5.60, T is a diagonal matrix, so D's role is played by T.

We leave the proof of the converse implication to the reader.

Let $U = (\mathbf{u}_1 \cdots \mathbf{u}_n)$. Since

$$\begin{aligned} A = UDU^{\mathsf{H}} &= (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^{\mathsf{H}} \\ \vdots \\ \mathbf{u}_n^{\mathsf{H}} \end{pmatrix} \\ &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathsf{H}} + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^{\mathsf{H}}, \end{aligned} \tag{7.5}$$

it follows that $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for $1 \leqslant i \leqslant n$, which proves the statement.

The equivalent Equalities (7.4) or (7.5) are referred to as *spectral decompositions* of the normal matrix A.

Theorem 7.41. (Spectral Theorem for Hermitian Matrices) *If the matrix $A \in \mathbb{C}^{n \times n}$ is Hermitian or skew-Hermitian, A can be written as $A = UDU^{\mathsf{H}}$, where U is a unitary matrix and D is a diagonal matrix having the eigenvalues of A as its diagonal elements.*

Proof. This statement follows from Theorem 7.40 because any Hermitian or skew-Hermitian matrix is normal.

Corollary 7.42. *The rank of a Hermitian matrix is equal to the number of non-zero eigenvalues.*

Proof. The statement of the corollary obviously holds for any diagonal matrix. If A is a Hermitian matrix, by Theorem 7.41, we have $\mathit{rank}(A) = \mathit{rank}(D)$, where D is a diagonal matrix having the eigenvalues of A as its diagonal elements. This implies the statement of the corollary.

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix of rank p. By Theorem 7.41 A can be written as $A = UDU^{\mathsf{H}}$, where U is a unitary matrix, $D = \mathsf{diag}(\lambda_1, \ldots, \lambda_p, 0, \ldots, 0)$ having as non-zero diagonal elements $\lambda_1, \ldots, \lambda_p$ the eigenvalues of A and $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$. Thus, if $W \in \mathbb{C}^{n \times p}$ is a matrix that consists of the first p columns of U we can write

$$
\begin{aligned}
A &= W \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} W' \\
&= (\mathbf{u}_1 \ \ldots \ \mathbf{u}_p) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^{\mathsf{H}} \\ \vdots \\ \mathbf{u}_p^{\mathsf{H}} \end{pmatrix} \\
&= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathsf{H}} + \cdots + \lambda_p \mathbf{u}_p \mathbf{u}_p^{\mathsf{H}}.
\end{aligned}
$$

If A is not Hermitian, $\mathit{rank}(A)$ may differ from the number of non-zero-eigenvalues. For example, the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

has no non-zero eigenvalues. However, its rank is 1.

The spectral decomposition (7.5) of Hermitian matrices,

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathsf{H}} + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^{\mathsf{H}}$$

allows us to extend functions of the form $f : \mathbb{R} \longrightarrow \mathbb{R}$ to Hermitian matrices. Since the eigenvalues of a Hermitian matrix are real numbers, it makes sense to define $f(A)$ as

$$f(A) = f(\lambda_1) \mathbf{u}_1 \mathbf{u}_1^{\mathsf{H}} + \cdots + f(\lambda_n) \mathbf{u}_n \mathbf{u}_n^{\mathsf{H}}.$$

In particular, if A is positive semi-definite, we have $\lambda_i \geqslant 0$ for $1 \leqslant i \leqslant n$ and we can define

$$\sqrt{A} = \sqrt{\lambda_1} \mathbf{u}_1 \mathbf{u}_1^{\mathsf{H}} + \cdots + \sqrt{\lambda_n} \mathbf{u}_n \mathbf{u}_n^{\mathsf{H}}.$$

Definition 7.43. *Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. The triple $\mathcal{I}(A) = (n_+(A), n_-(A), n_0(A))$, where $n_+(A)$ is the number of positive eigenvalues, $n_-(A)$ is the number of negative eigenvalues, and $n_0(A)$ is the number of zero eigenvalues is the* inertia *of the matrix A.*

The number $\mathit{sig}(A) = n_+(A) - n_-(A)$ is the signature *of A.*

Example 7.44. If $A = \mathsf{diag}(4, -1, 0, 0, 1)$, then $\mathcal{I}(A) = (2, 1, 2)$ and $sig(A) = 1$.

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. By Theorem 7.41 A can be written as $A = U^{\mathsf{H}} D U$, where U is a unitary matrix and $D = \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix having the eigenvalues of A (which are real numbers) as its diagonal elements. Without loss of generality we may assume that the positive eigenvalues of A are $\lambda_1, \ldots, \lambda_{n_+}$, followed by the negative values $\lambda_{n_+ + 1}, \ldots, \lambda_{n_+ + n_-}$, and the zero eigenvalues $\lambda_{n_+ + n_- + 1}, \ldots, \lambda_n$.

Let θ_j be the numbers defined by

$$\theta_j = \begin{cases} \sqrt{\lambda_j} & \text{if } 1 \leqslant j \leqslant n_+, \\ \sqrt{-\lambda_j} & \text{if } n_+ + 1 \leqslant j \leqslant n_+ + n_-, \\ 1 & \text{if } n_+ + n_- + 1 \leqslant j \leqslant n \end{cases}$$

for $1 \leqslant j \leqslant n$. If $T = \mathsf{diag}(\theta_1, \ldots, \theta_n)$, then we can write $D = T^{\mathsf{H}} G T$, where G is a diagonal matrix, $G = (g_1, \ldots, g_n)$ defined by

$$g_j = \begin{cases} 1 & \text{if } \lambda_j > 0, \\ -1 & \text{if } \lambda_j < 0, \\ 0 & \text{if } \lambda_j = 0, \end{cases}$$

for $1 \leqslant j \leqslant n$. This allows us to write $A = U^{\mathsf{H}} D U = U^{\mathsf{H}} T^{\mathsf{H}} G T U = (TU)^{\mathsf{H}} G (TU)$. The matrix TU is nonsingular, so $A \approx G$. The matrix G defined above is the *inertia matrix* of A and these definitions show that any Hermitian matrix is congruent to its inertia matrix.

For a Hermitian matrix $A \in \mathbb{C}^{n \times n}$ let $\mathsf{S}_+(A)$ be the subspace of $\mathbb{C}^n$ generated by $n_+(A)$ orthonormal eigenvectors that correspond to the positive eigenvalues of A. Clearly, we have $\dim(\mathsf{S}_+(A)) = n_+(A)$. This notation is used in the proof of the next theorem.

Theorem 7.45. (Sylvester's Inertia Theorem) *Let A, B be two Hermitian matrices, $A, B \in \mathbb{C}^{n \times n}$. We $A \approx B$ if and only if $\mathcal{I}(A) = \mathcal{I}(B)$.*

Proof. If $\mathcal{I}(A) = \mathcal{I}(B)$, then we have $A = S^{\mathsf{H}} G S$ and $B = T^{\mathsf{H}} G T$, where both S and T are nonsingular matrices. Since $A \approx G$ and $B \approx G$, we have $A \approx B$.

Conversely, suppose that $A \approx B$, that is, $A = S^{\mathsf{H}} B S$, where S is a nonsingular matrix. We have $rank(A) = rank(B)$, so $n_0(A) = n_0(B)$. To prove that $\mathcal{I}(A) = \mathcal{I}(B)$ it suffices to show that $n_+(A) = n_+(B)$.

Let $m = n_+(A)$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be m orthonormal eigenvectors of A that correspond to the m positive eigenvalues of this matrix, and let $\mathsf{S}_+(A)$ be the subspace generated by these vectors. If $\mathbf{v} \in \mathsf{S}_+(A) - \{\mathbf{0}\}$, then we have $\mathbf{v} = a_1 \mathbf{v}_1 + \cdots + a_m \mathbf{v}_m$, so

$$\mathbf{v}^{\mathsf{H}} A \mathbf{v} = \left(\sum_{j=1}^{m} a_j \mathbf{v}_j \right)^{\mathsf{H}} A \left(\sum_{j=1}^{m} a_j \mathbf{v}_j \right) = \sum_{j=1}^{m} |a_j|^2 > 0.$$

Therefore, $\mathbf{x}^{\mathsf{H}} S^{\mathsf{H}} B S \mathbf{x} > 0$, so if $\mathbf{y} = S\mathbf{x}$, then $\mathbf{y}^{\mathsf{H}} B \mathbf{y} > 0$, which means that $\mathbf{y} \in \mathsf{S}_+(B)$. This shows that $\mathsf{S}_+(A)$ is isomorphic to a subspace of $\mathsf{S}_+(B)$, so $n_+(A) \leqslant n_+(B)$. The reverse inequality can be shown in the same manner, so $n_+(A) = n_+(B)$.

We can add an interesting detail to the full-rank decomposition of a matrix.

Corollary 7.46. *If $A \in \mathbb{C}^{m\times n}$ and $A = CR$ is the full-rank decomposition of A with rank$(A) = k$, $C \in \mathbb{C}^{m\times k}$, and $R \in \mathbb{C}^{k\times n}$, then C may be chosen to have orthogonal columns and R to have orthogonal rows.*

Proof. Since the matrix $A^{\mathsf{H}} A \in \mathbb{C}^{n\times n}$ is Hermitian, by Theorem 7.41, there exists an unitary matrix $U \in \mathbb{C}^{n\times k}$ such that $A^{\mathsf{H}} A = U^{\mathsf{H}} D U$, where $D \in \mathbb{C}^{k\times k}$ is a non-negative diagonal matrix. Let $C = AU^{\mathsf{H}} \in C^{n\times k}$ and $R = U$. Clearly, $CR = A$, and R has orthogonal rows because U is unitary. Let $\mathbf{c}_p, \mathbf{c}_q$ be two columns of C, where $1 \leqslant p, q \leqslant k$ and $p \neq q$. Since $\mathbf{c}_p = A\mathbf{u}_p$ and $\mathbf{c}_q = A\mathbf{u}_q$, where $\mathbf{u}_p, \mathbf{u}_q$ are the corresponding columns of U, we have

$$\mathbf{c}_p^{\mathsf{H}} \mathbf{c}_q = \mathbf{u}_p^{\mathsf{H}} A^{\mathsf{H}} A \mathbf{u}_q = \mathbf{u}_p^{\mathsf{H}} U^{\mathsf{H}} D U \mathbf{u}_q = \mathbf{e}_p^{\mathsf{H}} D \mathbf{e}_q = 0,$$

because $p \neq q$.

7.5 Variational Characterizations of Spectra

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. By the Spectral Theorem for Hermitian Matrices (Theorem 7.41) A can be factored as $A = UDU^{\mathsf{H}}$, where U is a unitary matrix and $D = \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$. We assume that $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. The columns of U constitute a family of orthonormal vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $(\lambda_k, \mathbf{u}_k)$ are the eigenpairs of A.

Theorem 7.47. *Let A be a Hermitian matrix, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its eigenvalues having the orthonormal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, respectively.*

Define the subspace $M = \langle \boldsymbol{u}_p, \ldots, \boldsymbol{u}_q \rangle$, where $1 \leqslant p \leqslant q \leqslant n$. If $\boldsymbol{x} \in M$ and $\parallel \boldsymbol{x} \parallel_2 = 1$, we have $\lambda_q \leqslant \boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \leqslant \lambda_p$.

Proof. If $\mathbf{x}$ is a unit vector in M, then $\mathbf{x} = a_p \mathbf{u}_p + \cdots + a_q \mathbf{u}_q$, so $\mathbf{x}^{\mathsf{H}} \mathbf{u}_i = \overline{a_i}$ for $p \leqslant i \leqslant q$. Since $\parallel \mathbf{x} \parallel_2 = 1$, we have $|a_p|^2 + \cdots + |a_q|^2 = 1$. This allows us to write:

$$\begin{aligned}
\mathbf{x}^{\mathsf{H}} A \mathbf{x} &= \mathbf{x}^{\mathsf{H}} (a_p A \mathbf{u}_p + \cdots + a_q A \mathbf{u}_q) \\
&= \mathbf{x}^{\mathsf{H}} (a_p \lambda_p \mathbf{u}_p + \cdots + a_q \lambda_q \mathbf{u}_q) \\
&= \mathbf{x}^{\mathsf{H}} (|a_p|^2 \lambda_p + \cdots + |a_q|^2 \lambda_q).
\end{aligned}$$

Since $|a_p|^2 + \cdots + |a_q|^2 = 1$, the desired inequalities follow immediately.

Corollary 7.48. *Let A be a Hermitian matrix, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its eigenvalues having the orthonormal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, respectively. The following statements hold for a unit vector $\boldsymbol{x}$:*
(i) *if $\boldsymbol{x} \in \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_i \rangle$, then $\boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \geqslant \lambda_i$;*
(ii) *if $\boldsymbol{x} \in \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{i-1} \rangle^{\perp}$, then $\boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \leqslant \lambda_i$.*

Proof. The first statement follows directly from Theorem 7.47.

For the second statement observe that $\mathbf{x} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_{i-1} \rangle^{\perp}$ is equivalent to $\mathbf{x} \in \langle \mathbf{u}_i, \ldots, \mathbf{u}_n \rangle$; again, the second inequality follows from Theorem 7.47.

Theorem 7.49. (Rayleigh-Ritz Theorem) *Let A be a Hermitian matrix and let $(\lambda_1, \boldsymbol{u}_1), \ldots, (\lambda_n, \boldsymbol{u}_n)$ be the eigenpairs of A, where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. If $\boldsymbol{x}$ is a unit vector, we have $\lambda_n \leqslant \boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \leqslant \lambda_1$.*

Proof. This statement follows from Theorem 7.47 by observing that the subspace generated by $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is the entire space $\mathbb{C}^n$.

Next we discuss an important generalization of Rayleigh-Ritz Theorem.

Let $\mathcal{S}_p^n$ be the collection of p-dimensional subspaces of $\mathbb{C}^n$. Note that $\mathcal{S}_0^n = \{\{\mathbf{0}_n\}\}$ and $\mathcal{S}_n^n = \{\mathbb{C}^n\}$.

Theorem 7.50. (Courant-Fisher Theorem) *Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. We have*

$$\begin{aligned}\lambda_k &= \max_{U \in \mathcal{S}_k^n} \min\{\boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \mid \boldsymbol{x} \in U \text{ and } \parallel \boldsymbol{x} \parallel_2 = 1\} \\ &= \min_{U \in \mathcal{S}_{n-k+1}^n} \max\{\boldsymbol{x}^{\mathsf{H}} A \boldsymbol{x} \mid \boldsymbol{x} \in U \text{ and } \parallel \boldsymbol{x} \parallel_2 = 1\}.\end{aligned}$$

Proof. Let $A = U^{\mathsf{H}} \mathsf{diag}(\lambda_1, \ldots, \lambda_n) U$ be the factorization of A provided by Theorem 7.41), where $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n)$.

If $U \in \mathcal{S}_k^n$ and $W = \langle \mathbf{u}_k, \ldots, \mathbf{u}_n \rangle \in \mathcal{S}_{n-k+1}^n$, then there is a non-zero vector $\mathbf{x} \in U \cap W$ because $\dim(U) + \dim(W) = n + 1$; we can assume that $\parallel \mathbf{x} \parallel_2 = 1$. Therefore, by Theorem 7.47 we have $\lambda_k \geqslant \mathbf{x}^{\mathsf{H}} A \mathbf{x}$, and, therefore, for any $U \in \mathcal{S}_k^n$, $\lambda_k \geqslant \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$. This implies $\lambda_k \geqslant \max_{U \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$.

The same Theorem 7.47 implies that for a unit vector $\mathbf{x} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle \in \mathcal{S}_k^n$ we have $\mathbf{x}^{\mathsf{H}} A \mathbf{x} \geqslant \lambda_k$ and $\mathbf{u}_k^{\mathsf{H}} A \mathbf{u}_k = \lambda_k$. Therefore, for $W = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle \in \mathcal{S}_k^n$ we have $\min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in W, \parallel \mathbf{x} \parallel_2 = 1\} \geqslant \lambda_k$, so $\max_{W \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in W, \parallel \mathbf{x} \parallel_2 = 1\} \geqslant \lambda_k$. The inequalities proved above yield

$$\lambda_k = \max_{U \in \mathcal{S}_k^n} \min\{\mathbf{x}^{\mathsf{H}} A \mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}.$$

For the second equality, let $U \in \mathcal{S}_{n-k+1}^n$. If $W = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle$, there is a non-zero unit vector $\mathbf{x} \in U \cap W$ because $\dim(U) + \dim(W) \geqslant n + 1$. By Theorem 7.47 we have $\mathbf{x}^{\mathsf{H}} A \mathbf{x} \leqslant \lambda_k$. Therefore, for any $U \in \mathcal{S}_{n-k+1}^n$, $\lambda_k \geqslant$

$\max\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$. This implies $\lambda_k \geqslant \min_{U \in \mathcal{S}^n_{n-k+1}} \max\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$.

Theorem 7.47 implies that for a unit vector $\mathbf{x} \in \langle \mathbf{u}_k, \ldots, \mathbf{u}_n \rangle \in \mathcal{S}^n_{n-k+1}$ we have $\lambda_k \leqslant \mathbf{x}^{\mathsf{H}} A\mathbf{x}$ and $\lambda_k = \mathbf{u}_k^{\mathsf{H}} A\mathbf{u}_k$. Thus, $\lambda_k \leqslant \max\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$. Consequently, $\lambda_k \leqslant \min_{U \in \mathcal{S}^n_{n-k+1}} \max\{\mathbf{x}^{\mathsf{H}} A\mathbf{x} \mid \mathbf{x} \in U \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$, which completes the proof of the second equality of the theorem.

An equivalent formulation of Courant-Fisher Theorem is given next.

Theorem 7.51. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. We have*

$$\begin{aligned}\lambda_k &= \max_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_{n-k}} \min\{\boldsymbol{x}^{\mathsf{H}} A\boldsymbol{x} \mid \boldsymbol{x} \perp \boldsymbol{w}_1, \ldots, \boldsymbol{x} \perp \boldsymbol{w}_{n-k} \text{ and } \parallel \boldsymbol{x} \parallel_2 = 1\}\\ &= \min_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_{k-1}} \max\{\boldsymbol{x}^{\mathsf{H}} A\boldsymbol{x} \mid \boldsymbol{x} \perp \boldsymbol{w}_1, \ldots, \boldsymbol{x} \perp \boldsymbol{w}_{k-1} \text{ and } \parallel \boldsymbol{x} \parallel_2 = 1\}.\end{aligned}$$

Proof. The equalities of the Theorem follow from the Courant-Fisher theorem taking into account that if $U \in \mathcal{S}^n_k$, then $U^{\perp} = \langle \mathbf{w}_1, \ldots, \mathbf{w}_{n-k} \rangle$ for some vectors $\mathbf{w}_1, \ldots, \mathbf{w}n - k$, and if $U \in \mathcal{S}^n_{n-k+1}$, then $U = \langle \mathbf{w}_1, \ldots \mathbf{w}_{k-1} \rangle$ for some vectors $\mathbf{w}_1, \ldots, \mathbf{w}k - 1$ in $\mathbb{C}^n$.

Definition 7.52. *Consider two non-increasing sequences of real numbers $(\lambda_1, \ldots, \lambda_n)$ and $(\mu_1, \ldots, \mu_m)$ with $m < n$. We say that the second sequence* interlace *the first if $\lambda_i \geqslant \mu_i \geqslant \lambda_{n-m+i}$ for $1 \leqslant i \leqslant m$.*

The interlacing is tight *if there exists $k \in \mathbb{N}$, $0 \leqslant k \leqslant m$ such that $\lambda_i = \mu_i$ for $0 \leqslant i \leqslant k$ and $\lambda_{n-m+1} = \mu_i$ for $k + 1 \leqslant i \leqslant m$.*

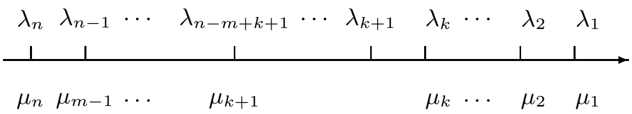

Fig. 7.1. Tight interlacing

The next statement is known as the Interlacing Theorem. The variant included here was obtained in [81].

Theorem 7.53. (Interlacing Theorem) *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, $S \in \mathbb{C}^{n\times m}$ be a matrix such that $S^{\mathsf{H}} S = I_m$, and let $B = S^{\mathsf{H}} AS \in \mathbb{C}^{m\times m}$, where $m \leqslant n$.*

Assume that A has the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ with the orthonormal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, respectively, and B has the eigenvalues $\mu_1 \geqslant \cdots \geqslant \mu_m$ with the respective eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$. The following statements hold:

(i) $\lambda_i \geqslant \mu_i \geqslant \lambda_{n-m+i}$;

(ii) *if $\mu_i = \lambda_i$ or $\mu_i = \lambda_{n-m+i}$ for some i, $1 \leqslant i \leqslant m$, then B has an eigenvector $\boldsymbol{v}$ such that $(\mu_i, \boldsymbol{v})$ is an eigenpair of B and $(\mu_i, S\boldsymbol{v})$ is an eigenpair of A;*
(iii) *if for some integer ℓ, $\mu_i = \lambda_i$ for $1 \leqslant i \leqslant \ell$ (or $\mu_i = \lambda_{n-m+i}$ for $\ell \leqslant i \leqslant m$), then $(\mu_i, S\boldsymbol{v}_i)$ are eigenpairs of A for $1 \leqslant i \leqslant \ell$ (respectively, $\ell \leqslant i \leqslant m$);*
(iv) *if the interlacing is tight $SB = AS$.*

Proof. Note that $B^{\mathsf{H}} = S^{\mathsf{H}} A^{\mathsf{H}} S = S^{\mathsf{H}} A S = B$, so B is also Hermitian.

Since $\dim(\langle \mathbf{v}_1, \ldots, \mathbf{v}_i \rangle) = i$ and $\dim(\langle S^{\mathsf{H}}\mathbf{u}_1, \ldots, S^{\mathsf{H}}\mathbf{u}_{i-1} \rangle^{\perp}) = n - i + 1$, there exist non-zero vectors in the intersection of these subspaces. Let $\mathbf{t}$ be a unit vector in $\langle \mathbf{v}_1, \ldots, \mathbf{v}_i \rangle \cap \langle S^{\mathsf{H}}\mathbf{u}_1, \ldots, S^{\mathsf{H}}\mathbf{u}_{i-1} \rangle^{\perp}$.

Since $\mathbf{t} \in \langle S^{\mathsf{H}}\mathbf{u}_1, \ldots, S^{\mathsf{H}}\mathbf{u}_{i-1} \rangle^{\perp}$ we have $\mathbf{t}^{\mathsf{H}} S^{\mathsf{H}} \mathbf{u}_\ell = (S\mathbf{t})^{\mathsf{H}} \mathbf{u}_\ell = 0$ for $1 \leqslant \ell \leqslant i - 1$. Thus, $S\mathbf{t} \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_{i-1} \rangle^{\perp}$, so, by Theorem 7.47, it follows that $\lambda_i \geqslant (S\mathbf{t})^{\mathsf{H}} A (S\mathbf{t})$. On other hand,

$$(S\mathbf{t})^{\mathsf{H}} A (S\mathbf{t}) = \mathbf{t}^{\mathsf{H}} (S^{\mathsf{H}} A S) \mathbf{t} = \mathbf{t}^{\mathsf{H}} B \mathbf{t}$$

and $\mathbf{t} \in \langle \mathbf{v}_1, \ldots, \mathbf{v}_i \rangle$, which yield $\mathbf{t}^{\mathsf{H}} B \mathbf{t} \geqslant \mu_i$, again, by Theorem7.47. Thus, we conclude that

$$\lambda_i \geqslant (S\mathbf{t})^{\mathsf{H}} A (S\mathbf{t}) = \mathbf{t}^{\mathsf{H}} B \mathbf{t} \geqslant \mu_i. \tag{7.6}$$

Note that the matrices $-A$ and $-B$ have the eigenvalues $-\lambda_n \geqslant \cdots \geqslant -\lambda_1$ and $-\mu_m \geqslant \cdots \geqslant -\mu_1$. The i^{th} eigenvalue in the list $-\lambda_n \geqslant \cdots \geqslant -\lambda_1$ is $-\lambda_{n-m+i}$, so, by applying the previous argument to the matrices $-A$ and $-B$ we obtain: $\mu_i \geqslant \lambda_{n-m+i}$, which concludes the proof of (i).

If $\lambda_i = \mu_i$, since $B\mathbf{v}_i = \mu_i \mathbf{v}_i$, it follows that $S^{\mathsf{H}} A S \mathbf{v}_i = \mu_i \mathbf{v}_i$, so $A(S\mathbf{v}_i) = \mu_i (S\mathbf{v}_i)$ because S is a unitary matrix. Thus, $(\mu_i, S\mathbf{v}_i)$ is an eigenpair of A, which proves Part (ii).

Part (iii) follows directly from Part (ii) and its proof.

Finally, if the interlacing is tight, $S\mathbf{v}_1, \ldots, S\mathbf{v}_m$ is an orthonormal set of eigenvectors of A corresponding to the eigenvalues $\mu_1, \ldots, \mu_m$, so $SB\mathbf{v}_i = \mu_i S\mathbf{v}_i = AS\mathbf{v}_i$ for $1 \leqslant i \leqslant m$. Since $SB, AS \in \mathbb{C}^{n \times m}$ and the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ form a basis in $\mathbb{C}^m$, we have $SB = AS$.

Example 7.54. Let $R = \{r_1, \ldots, r_m\}$ be a subset of $\{1, \ldots, n\}$. Define the matrix $S_R \in \mathbb{C}^{n \times m}$ by $S_R = (\mathbf{e}_{r_1} \cdots \mathbf{e}_{r_m})$.

For example, if $n = 4$ and $R = \{2, 4\}$ we have the matrix

$$S_R = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is immediate that $S_R^{\mathsf{H}} S_R = I_m$ and that $S_R^{\mathsf{H}} A S_R$ is the principal submatrix $A \begin{bmatrix} R \\ R \end{bmatrix}$, defined by the intersection of rows $r_1, \ldots, r_m$ with the columns $r_1, \ldots, r_m$.

Corollary 7.55. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let $B = A\begin{bmatrix} R \\ R \end{bmatrix} \in \mathbb{C}^{m\times m}$ be a principal submatrix of A. If A has the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and B has the eigenvalues $\mu_1 \geqslant \cdots \geqslant \mu_m$, then $\lambda_i \geqslant \mu_i \geqslant \lambda_{n-m+i}$ for $1 \leqslant i \leqslant m$.*

Proof. This statement follows immediately from Theorem 7.53, by taking $S = S_R$.

Theorem 7.56. *Let $A, B \in \mathbb{C}^{n\times}$ be two Hermitian matrices and let $E = B - A$. Suppose that the eigenvalues of A, B, E these are $\alpha_1 \geqslant \cdots \geqslant \alpha_n$, $\beta_1 \geqslant \cdots \geqslant \beta_n$, and $\epsilon_1 \geqslant \cdots \geqslant \epsilon_n$, respectively. Then, we have $\epsilon_n \leqslant \beta_i - \alpha_i \leqslant \epsilon_1$.*

Proof. Note that E is also Hermitian, so all matrices involved have real eigenvalues. By Courant-Fisher Theorem,

$$\beta_k = \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{w}_i^{\mathsf{H}}\mathbf{x} = 0 \text{ for } 1 \leqslant i \leqslant k-1\},$$

where $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}\}$. Thus,

$$\beta_k \leqslant \max_{\mathbf{x}} \mathbf{x}^{\mathsf{H}} B\mathbf{x} = \max_{\mathbf{x}} (\mathbf{x}^{\mathsf{H}} A\mathbf{x} + \mathbf{x}^{\mathsf{H}} E\mathbf{x}). \tag{7.7}$$

Let U be a unitary matrix such that $U^{\mathsf{H}} AU = \mathsf{diag}(\alpha_1, \ldots, \alpha_n)$. Choose $\mathbf{w}_i = U\mathbf{e}_i$ for $1 \leqslant i \leqslant k-1$. We have $\mathbf{w}_i^{\mathsf{H}}\mathbf{x} = \mathbf{e}_i^{\mathsf{H}} U^{\mathsf{H}}\mathbf{x} = 0$ for $1 \leqslant i \leqslant k-1$.

Define $\mathbf{y} = U^{\mathsf{H}}\mathbf{x}$. Since U is an unitary matrix, $\parallel \mathbf{y} \parallel_2 = \parallel \mathbf{x} \parallel_2 = 1$. Observe that $\mathbf{e}_i^{\mathsf{H}}\mathbf{y} = y_i = 0$ for $1 \leqslant i \leqslant k$. Therefore, $\sum_{i=k}^n y_i^2 = 1$. This, in turn implies $\mathbf{x}^{\mathsf{H}} A\mathbf{x} = \mathbf{y}^{\mathsf{H}} U^{\mathsf{H}} AU\mathbf{y} = \sum_{i=k}^n \alpha_i y_i^2 \leqslant \alpha_k$.

From the Inequality (7.7) it follows that

$$\beta_k \leqslant \alpha_k + \max_{\mathbf{x}} \mathbf{x}^{\mathsf{H}} E\mathbf{x} \leqslant \alpha_k + \epsilon_n.$$

Since $A = B - E$, by inverting the roles of A and B we have $\alpha_k \leqslant \beta_k - \epsilon_1$, or $\epsilon_1 \leqslant \beta_k - \alpha_k$, which completes the argument.

Lemma 7.57. *Let $T \in \mathbb{C}^{n\times n}$ be a upper triangular matrix and let $\lambda \in \mathsf{spec}(T)$ be an eigenvalue such that the diagonal entries that equal λ occur in $t_{i_1 i_1}, \ldots, t_{i_p i_p}$. Then, the invariant subspace $S_{T,\lambda}$ is the p-dimensional subspace generated by $\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_p}$.*

Proof. The argument is straightforward and is omitted.

Lemma 7.58. *Let $T \in \mathbb{C}^{n\times n}$ be an upper triangular matrix and let $p_T(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$ be its characteristic polynomial. Then,*

$$p_T(T) = T^n + a_1 T^{n-1} + \cdots + a_{n-1}T + a_n I_n = O_{n,n}.$$

Proof. We have

$$p_T(T) = (T - \lambda_1 I_n) \cdots (T - \lambda_n I_n).$$

Observe that for any matrix $A \in \mathbb{C}^{n\times n}$, $\lambda_j, \lambda_k \in \mathsf{spec}(A)$, and every eigenvector $\mathbf{v}$ of A in S_{A,λ_k} we have

$$(\lambda_j I_n - A)\mathbf{v} = (\lambda_j - \lambda_k)\mathbf{v}.$$

Therefore, for $\mathbf{v} \in S_{T,\lambda_k}$ we have:

$$p_T(T)\mathbf{v} = (\lambda_1 I_n - T) \cdots (\lambda_n I_n - T)\mathbf{v} = \mathbf{0},$$

because $(\lambda_k I - T)\mathbf{v} = \mathbf{0}$.

By Lemma 7.57, $p_T(T)\mathbf{e}_i = \mathbf{0}$ for $1 \leqslant i \leqslant n$, so $p_T(T) = O_{n,n}$.

Theorem 7.59. (Cayley-Hamilton Theorem) *If $A \in \mathbb{C}^{n\times n}$ is a matrix, then $p_A(A) = O_{n,n}$.*

Proof. By Schur's Triangularization Theorem there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A = UTU^{\mathsf{H}}$ and the diagonal elements of T are the eigenvalues of A. Taking into account that U is unitary we can write:

$$\begin{aligned}
p_A(A) &= (\lambda_1 I_n - A)(\lambda_2 I_n - A) \cdots (\lambda_n I_n - A) \\
&= (\lambda_1 UU^{\mathsf{H}} - UTU^{\mathsf{H}}) \cdots (\lambda_n UU^{\mathsf{H}} - UTU^{\mathsf{H}}) \\
&= U(\lambda_1 I_n - T)U^{\mathsf{H}}U(\lambda_2 I_n - T)U^{\mathsf{H}} \cdots U(\lambda_n I_n - T)U^{\mathsf{H}} \\
&= U(\lambda_1 I_n - T)(\lambda_2 I_n - T) \cdots (\lambda_n I_n - T)U^{\mathsf{H}} \\
&= U p_T(T) U^{\mathsf{H}} = O_{n,n},
\end{aligned}$$

by Lemma 7.58.

Theorem 7.60. (Ky Fan's Theorem) *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix such that $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Also, let $V \in \mathbb{C}^{n\times n}$ be a matrix, $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ whose set of columns constitutes an orthonormal set of eigenvectors of A.*

For every $q \in \mathbb{N}$ such that $1 \leqslant q \leqslant n$, the sums $\sum_{i=1}^{q} \lambda_i$ and $\sum_{i=1}^{q} \lambda_{n+1-i}$ are the maximum and minimum of $\sum_{j=1}^{q} \boldsymbol{x}_j^{\mathsf{H}} A \boldsymbol{x}_j$, where $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q\}$ is an orthonormal set of vectors in $\mathbb{C}^n$, respectively. The maximum (minimum) is achieved when $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q$ are the first (last) columns of V.

Proof. Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an orthonormal set of eigenvectors of A and let $\mathbf{x}_i = \sum_{k=1}^{n} b_{ki}\mathbf{v}_k$ be the expression of $\mathbf{x}_i$ using the columns of V as a basis for $1 \leqslant i \leqslant n$. Since each $\mathbf{x}_i$ is a unit vector we have

$$\| \mathbf{x}_i \|^2 = \mathbf{x}_i^{\mathsf{H}}\mathbf{x}_i = \sum_{k=1}^{n} |b_{ki}|^2 = 1$$

for $1 \leqslant i \leqslant n$. Also, note that

$$\mathbf{x}_i^{\mathsf{H}}\mathbf{v}_r = \left(\sum_{k=1}^{n} \overline{b_{ki}}\mathbf{v}_k^{\mathsf{H}}\right)\mathbf{v}_r = \overline{b_{ri}},$$

due to the orthonormality of the set of columns of V. We have

$$\begin{aligned}
\mathbf{x}_i^{\mathsf{H}}A\mathbf{x}_i &= \mathbf{x}_i^{\mathsf{H}}A\sum_{k=1}^{n} b_{ki}\mathbf{v}_k = \sum_{k=1}^{n} b_{ki}\mathbf{x}_i^{\mathsf{H}}A\mathbf{v}_k \\
&= \sum_{k=1}^{n} b_{ki}\mathbf{x}_i^{\mathsf{H}}\lambda_k\mathbf{v}_k = \sum_{k=1}^{n} \lambda_k b_{ki}\overline{b_{ki}} = \sum_{k=1}^{n} |b_{ki}|^2\lambda_k \\
&= \lambda_q\sum_{k=1}^{n} |b_{ki}|^2 + \sum_{k=1}^{q}(\lambda_k - \lambda_q)|b_{ki}|^2 + \sum_{k=q+1}^{n}(\lambda_k - \lambda_q)|b_{ki}|^2 \\
&\leqslant \lambda_q + \sum_{k=1}^{q}(\lambda_k - \lambda_q)|b_{ki}|^2.
\end{aligned}$$

The last inequality implies

$$\sum_{i=1}^{q} \mathbf{x}_i^{\mathsf{H}}A\mathbf{x}_i \leqslant q\lambda_q + \sum_{i=1}^{q}\sum_{k=1}^{q}(\lambda_k - \lambda_q)|b_{ki}|^2.$$

Therefore,

$$\sum_{i=1}^{q}\lambda_i - \sum_{i=1}^{q}\mathbf{x}_i^{\mathsf{H}}A\mathbf{x}_i \geqslant \sum_{i=1}^{q}(\lambda_i - \lambda_q)\left(1 - \sum_{k=1}^{q}|b_{ki}|^2\right). \tag{7.8}$$

By Inequality (6.10), we have $\sum_{k=1}^{q}|b_{ik}|^2 \leqslant \parallel \mathbf{x}_i \parallel^2 = 1$, so

$$\sum_{i=1}^{q}(\lambda_i - \lambda_q)\left(1 - \sum_{k=1}^{q}|b_{ki}|^2\right) \geqslant 0.$$

The left member of Inequality 7.8 becomes 0 when $\mathbf{x}_i = \mathbf{v}_i$, so $\sum_{i=1}^{q}\mathbf{x}_i^{\mathsf{H}}A\mathbf{x}_i \leqslant \sum_{i=1}^{q}\lambda_i$. The maximum of $\sum_{i=1}^{q}\mathbf{x}_i^{\mathsf{H}}A\mathbf{x}_i$ is obtained when $\mathbf{x}_i = \mathbf{v}_i$ for $1 \leqslant i \leqslant q$, that is, when X consists of the first q columns of V.

The argument for te minimum is similar.

Theorem 7.61. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. If A is positive semidefinite, then all its eigenvalues are non-negative; if A is positive definite then its eigenvalues are positive.*

Proof. Since A is Hermitian all its eigenvalues are real numbers. Suppose that A is positive semidefinite, that is, $\mathbf{x}^{\mathsf{H}}A\mathbf{x} \geqslant 0$ for $\mathbf{x} \in \mathbb{C}^n$. If $\lambda \in \mathsf{spec}(A)$, then $A\mathbf{v} = \lambda\mathbf{v}$ for some eigenvector $\mathbf{v} \neq \mathbf{0}$. The positive semi-definiteness of A implies $\mathbf{v}^{\mathsf{H}}A\mathbf{v} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{v} = \lambda \parallel \mathbf{v} \parallel_2^2 \geqslant 0$, which implies $\lambda \geqslant 0$. It is easy to see that if A is positive definite, then $\lambda > 0$.

Theorem 7.62. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. If A is positive semidefinite, then all its principal minors are non-negative real numbers. If A is positive definite then all its principal minors are positive real numbers.*

Proof. Since A is positive semidefinite, every sub-matrix $A\begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix}$ is a Hermitian positive semidefinite matrix by Theorem 6.110, so every principal minor is a non-negative real number. The second part of the theorem is proven similarly.

Corollary 7.63. *Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. The following statements are equivalent.*

(i) *A is positive semidefinite;*
(ii) *all eigenvalues of A are non-negative numbers;*
(iii) *there exists a Hermitian matrix $C \in \mathbb{C}^{n\times n}$ such that $C^2 = A$;*
(iv) *A is the Gram matrix of a sequence of vectors, that is, $A = B^{\mathsf{H}}B$ for some $B \in \mathbb{C}^{n\times n}$.*

Proof. (i) implies (ii): This was shown in Theorem 7.61.

(ii) implies (iii): Suppose that A is a matrix such that all its eigenvalues are the non-negative numbers $\lambda_1, \ldots, \lambda_n$. By Theorem 7.41, A can be written as $A = U^{\mathsf{H}}DU$, where U is a unitary matrix and

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Define the matrix $\sqrt{D}$ as

$$\sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}.$$

Clearly, we gave $(\sqrt{D})^2 = D$. Now we can write $A = U\sqrt{D}U^{\mathsf{H}}U\sqrt{D}U^{\mathsf{H}}$, which allows us to define the desired matrix C as $C = U\sqrt{D}U^{\mathsf{H}}$.

(iii) implies (iv): Since C is itself a Hermitian matrix, this implication is obvious.

(iv) implies (i): Suppose that $A = B^{\mathsf{H}}B$ for some matrix $B \in \mathbb{C}^{n\times k}$. Then, for $\mathbf{x} \in \mathbb{C}^n$ we have $\mathbf{x}^{\mathsf{H}}A\mathbf{x} = \mathbf{x}^{\mathsf{H}}B^{\mathsf{H}}B\mathbf{x} = (B\mathbf{x})^{\mathsf{H}}(B\mathbf{x}) = \| B\mathbf{x} \|_2^2 \geqslant 0$, so A is positive semidefinite.

7.6 Matrix Norms and Spectral Radii

Definition 7.64. *Let $A \in \mathbb{C}^{n\times n}$. The* spectral radius *of A is the number $\rho(A) = \max\{|\lambda| \mid \lambda \in \mathsf{spec}(A)\}$.*

If $(\lambda, \mathbf{x})$ is an eigenpair of A, then $|\lambda| |\!|\!| \mathbf{x} |\!|\!| = |\!|\!| A\mathbf{x} |\!|\!| \leqslant |\!|\!| A |\!|\!| |\!|\!| \mathbf{x} |\!|\!|$, so $|\lambda| \leqslant |\!|\!| A |\!|\!|$, which implies $\rho(A) \leqslant |\!|\!| A |\!|\!|$ for any matrix norm $|\!|\!| \cdot |\!|\!|$. Moreover, we can prove the following statement.

Theorem 7.65. *Let $A \in \mathbb{C}^{n\times n}$. The spectral radius $\rho(A)$ is the infimum of the set that consists of numbers of the form $|\!|\!| A |\!|\!|$, where $|\!|\!| \cdot |\!|\!|$ ranges over all matrix norms defined on $\mathbb{C}^{n\times n}$.*

Proof. Since we have shown that $\rho(A)$ is a lower bound of the set of numbers mentioned in the statement, we need to prove only that for every $\epsilon > 0$ there exists a matrix norm $|\!|\!| \cdot |\!|\!|$ such that $|\!|\!| A |\!|\!| \leqslant \rho(A) + \epsilon$.

By Schur's Triangularization Theorem there exists a unitary matrix U and an upper triangular matrix T such that $A = UTU^{-1}$ such that the diagonal elements of T are $\lambda_1, \ldots, \lambda_n$.

For $\alpha \in \mathbb{R}_{>0}$ let $S_\alpha = \mathsf{diag}(\alpha, \alpha^2, \ldots, \alpha^n)$. We have

$$S_\alpha T S_\alpha^{-1} = \begin{pmatrix} \lambda_1 & \alpha^{-1}t_{12} & \alpha^{-2}t_{12} & \cdots & \alpha^{-(n-1)}t_{1n} \\ 0 & \lambda_1 & \alpha^{-1}t_{23} & \cdots & \alpha^{-(n-2)}t_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \lambda_n \end{pmatrix}.$$

If α is sufficiently large, $|\!|\!| S_\alpha T S_\alpha^{-1} |\!|\!|_1 \leqslant \rho(A) + \epsilon$ because, in this case, the sum of the absolute values of the supradiagonal elements can be arbitrarily small. Let $M = (US_\alpha)^{-1}$. For the matrix norm $\mu_M(\cdot)$ (see Exercise 7) we have $\mu_M(A) = |\!|\!| US_\alpha A S_\alpha^{-1} U^{-1} |\!|\!|_1$. If α is sufficiently large we have $\mu_M(A) \leqslant \rho(A) + \epsilon$.

Let $A, B \in \mathbb{C}^{n\times n}$. We leave to the reader to verify that if $\mathsf{abs}(A) \leqslant \mathsf{abs}(B)$, then $\| A \|_2 \leqslant \| B \|_2$; also, $\| A \|_2 = \| \mathsf{abs}(A) \|_2$.

Theorem 7.66. *Let $A, B \in \mathbb{C}^{n\times n}$. If $\mathsf{abs}(A) \leqslant B$, then $\rho(A) \leqslant \rho(\mathsf{abs}(A)) \leqslant \rho(B)$.*

Proof. By Theorem 5.62 we have $\mathsf{abs}(A^k) \leqslant (\mathsf{abs}(A))^k \leqslant B^k$ for every $k \in \mathbb{N}$. Therefore, $\| \mathsf{abs}(A^k) \| \leqslant (\mathsf{abs}(A))^k \leqslant B^k$, so $\| A^k \|_2 \leqslant \| \mathsf{abs}(A)^k \|_2 \leqslant \| B^k \|_2$, which implies $\| A^k \|_2^{\frac{1}{k}} \leqslant \| \mathsf{abs}(A)^k \|_2^{\frac{1}{k}} \leqslant \| B^k \|_2^{\frac{1}{k}}$.

By letting k tend to ∞ we obtain the double inequality of the theorem.

Corollary 7.67. *If $A, B \in \mathbb{C}^{n\times n}$ are two matrices such that $O_{n,n} \leqslant A \leqslant B$, then $\rho(A) \leqslant \rho(B)$.*

Proof. The corollary follows immediately from Theorem 7.66 by observing that under the hypothesis, $A = \mathsf{abs}(A)$.

Theorem 7.68. *Let $A \in \mathbb{C}^{n\times n}$. We have $\lim_{k\to\infty} = O_{n,n}$ if and only if $\rho(A) < 1$.*

Proof. Suppose that $\lim_{k\to\infty} = O_{n,n}$. Let $(\lambda, \mathbf{x})$ be an eigenpair of A, so $\mathbf{x} \neq \mathbf{0}_n$ and $A\mathbf{x} = \lambda\mathbf{x}$. This implies $A^k\mathbf{x} = \lambda^k\mathbf{x}$, so $\lim_{k\to\infty} \lambda^k\mathbf{x} = \mathbf{0}_n$. Thus, $\lim_{k\to\infty} \lambda^k\mathbf{x} = \mathbf{0}_n$, which implies $\lim_{k\to\infty} \lambda^k = 0$. Thus, $|\lambda| < 1$ for every $\lambda \in \mathsf{spec}(A)$, so $\rho(A) < 1$.

Conversely, suppose that $\rho(A) < 1$. By Theorem 7.65, there exists a matrix norm $\|\!| \cdot |\!\|$ such that $\|\!| A |\!\| < 1$. Thus, $\lim_{k\to\infty} A^k = O_{n,n}$.

7.7 Singular Values of Matrices

Definition 7.69. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix. A* singular triplet *of A is a triplet $(\sigma, \boldsymbol{u}, \boldsymbol{v})$ such that $\sigma \in \mathbb{R}_{>0}$, $\boldsymbol{u} \in \mathbb{C}^n$, $\boldsymbol{v} \in \mathbb{C}^m$, $A\boldsymbol{u} = \sigma\boldsymbol{v}$ and $A^{\mathsf{H}}\boldsymbol{v} = \sigma\boldsymbol{u}$. The number σ is a* singular value *of A, $\boldsymbol{u}$ is a* left singular vector *and $\boldsymbol{v}$ is a* right singular vector*.*

For a singular triplet $(\sigma, \mathbf{u}, \mathbf{v})$ of A we have $A^{\mathsf{H}}A\mathbf{u} = \sigma A^{\mathsf{H}}\mathbf{v} = \sigma^2\mathbf{u}$ and $AA^{\mathsf{H}}\mathbf{v} = \sigma A\mathbf{u} = \sigma^2\mathbf{v}$. Therefore, σ^2 is both an eigenvalue of AA^{H} and an eigenvalue of $A^{\mathsf{H}}A$.

Example 7.70. Let A be the real matrix

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \cos\beta & \sin\beta \end{pmatrix}.$$

We have $\det(A) = \sin(\beta - \alpha)$, so the eigenvalues of $A'A$ are the roots of the equation $\lambda^2 - 2\lambda + \sin^2(\beta - \alpha) = 0$, that is, $\lambda_1 = 1 + \cos(\beta - \alpha)$ and $\lambda_2 = 1 - \cos(\beta - \alpha)$. Therefore, the singular values of A are $\sigma_1 = \sqrt{2}\left|\cos\frac{\beta-\alpha}{2}\right|$ and $\sigma_2 = \sqrt{2}\left|\sin\frac{\beta-\alpha}{2}\right|$.

It is easy to see that a unit left singular vector that corresponds to the eigenvalue $1 + \cos(\beta - \alpha)$ is

$$\mathbf{u} = \begin{pmatrix} \cos\frac{\alpha+\beta}{2} \\ \sin\frac{\alpha+\beta}{2} \end{pmatrix},$$

which corresponds to the average direction of the rows of A.

We noted that the eigenvalues of a positive semi-definite matrix are non-negative numbers. Since both AA^{H} and $A^{\mathsf{H}}A$ are positive semi-definite matrices for $A \in \mathbb{C}^{m\times n}$ (see Example 6.109), the spectra of these matrices consist of non-negative numbers $\lambda_1, \ldots, \lambda_n$. Furthermore, AA^{H} and $A^{\mathsf{H}}A$ have the same rank r and therefore, the same number r of non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$. Accordingly, the singular values of A have the form $\sqrt{\lambda_1} \geqslant \cdots \geqslant \sqrt{\lambda_r}$. We will use the notation $\sigma_i = \sqrt{\lambda_i}$ for $1 \leqslant i \leqslant r$ and will assume that $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$.

Theorem 7.71. *Let $A \in \mathbb{C}^{n \times n}$ be a matrix having the singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_n$. If λ is an eigenvalue value of A, then $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$.*

Proof. Let $\mathbf{u}$ be an unit eigenvector for the eigenvalue λ. Since $A\mathbf{u} = \lambda\mathbf{u}$ it follows that $(A^{\mathsf{H}}A\mathbf{u}, \mathbf{u}) = (A\mathbf{u}, A\mathbf{u}) = \overline{\lambda}\lambda(\mathbf{u}, \mathbf{u}) = \overline{\lambda}\lambda = |\lambda|^2$. The matrix $A^{\mathsf{H}}A$ is Hermitian and its largest and smallest eigenvalues are σ_1^2 and σ_n^2, respectively. Thus, $\sigma_n \leqslant |\lambda| \leqslant \sigma_1$.

Theorem 7.72. (SVD Theorem) *If $A \in \mathbb{C}^{m \times n}$ is a matrix and rank$(A) = r$, then A can be factored as $A = UDV^{\mathsf{H}}$, where $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary matrices, and $D = \mathsf{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{m \times n}$, where $\sigma_1 \geqslant \ldots \geqslant \sigma_r$ are real positive numbers.*

Proof. We saw that the square matrix $A^{\mathsf{H}}A \in \mathbb{C}^{n \times n}$ has the same rank r as the matrix A and is positive semidefinite. Therefore, there are r positive eigenvalues of this matrix, denoted by $\sigma_1^2, \ldots, \sigma_r^2$, where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be the corresponding pairwise orthogonal unit eigenvectors in $\mathbb{C}^n$.

We have $A^{\mathsf{H}}A\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ for $1 \leqslant i \leqslant r$. Define $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r \ \mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$ by completing the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to an orthogonal basis

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$$

for $\mathbb{C}^n$. If $V_1 = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r)$ and $V_2 = (\mathbf{v}_{r+1} \ \cdots \ \mathbf{v}_n)$, we can write $V = (V_1 \ V_2)$.

The equalities involving the eigenvectors can now be written as $A^{\mathsf{H}}AV_1 = V_1E^2$, where $E = \mathsf{diag}(\sigma_1, \ldots, \sigma_r)$.

Define $U_1 = AV_1E^{-1} \in \mathbb{C}^{m \times r}$. We have $U_1^{\mathsf{H}} = S^{-1}V_1^{\mathsf{H}}A^{\mathsf{H}}$, so

$$U_1^{\mathsf{H}}U_1 = S^{-1}V_1^{\mathsf{H}}A^{\mathsf{H}}AV_1E^{-1} = E^{-1}V_1^{\mathsf{H}}V_1E^2E^{-1} = I_r,$$

which shows that the columns of U_1 are pairwise orthogonal unit vectors. Consequently, $U_1^{\mathsf{H}}AV_1E^{-1} = I_r$, so $U_1^{\mathsf{H}}AV_1 = E$.

If $U_1 = (\mathbf{u}_1 \ \cdots, \mathbf{u}_r)$, let $U_2 = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m)$ be the matrix whose columns constitute the extension of the set $\{\mathbf{u}_1 \ \cdots, \mathbf{u}_r\}$ to an orthogonal basis of $\mathbb{C}^m$. Define $U \in \mathbb{C}^{m \times m}$ as $U = (U_1 \ U_2)$. Note that

$$\begin{aligned} U^{\mathsf{H}}AV &= \begin{pmatrix} U_1^{\mathsf{H}} \\ U_2^{\mathsf{H}} \end{pmatrix} A(V_1 \ V_2) = \begin{pmatrix} U_1^{\mathsf{H}}AV_1 & U_1^{\mathsf{H}}AV_2 \\ U_2^{\mathsf{H}}AV_1 & U_2^{\mathsf{H}}AV_2 \end{pmatrix} \\ &= \begin{pmatrix} U_1^{\mathsf{H}}AV_1 & U_1^{\mathsf{H}}AV_2 \\ U_2^{\mathsf{H}}AV_1 & U_2^{\mathsf{H}}AV_2 \end{pmatrix} = \begin{pmatrix} U_1^{\mathsf{H}}AV_1 & O \\ O & O \end{pmatrix} = \begin{pmatrix} E & O \\ O & O \end{pmatrix}, \end{aligned}$$

which is the desired decomposition.

Corollary 7.73. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix such that rank$(A) = r$. If $\sigma_1 \geqslant \cdots \geqslant \sigma_r$ are non-zero singular values, then*

$$A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathsf{H}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathsf{H}}, \tag{7.9}$$

where $(\sigma_i, \boldsymbol{u}_i, \boldsymbol{v}_i)$ are singular triplets of A for $1 \leqslant i \leqslant r$.

Proof. This follows directly from Theorem 7.72.

The value of a unitarily invariant norm of a matrix depends only on its singular values.

Corollary 7.74. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix and let $A = UDV^{\mathsf{H}}$ be the singular value decomposition of A. If $\| \cdot \|$ is a unitarily invariant norm, then*

$$\| A \| = \| D \| = \| \mathsf{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \| .$$

Proof. This statement is a direct consequence of Theorem 7.72 because the matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary.

As we saw in Theorem 6.86, $||| \cdot |||_2$ and $\| \cdot \|_F$ are unitarily invariant. Therefore, the Frobenius norm can be written as

$$\| A \|_F = \sqrt{\sum_{i=1}^{r} \sigma_r^2}.$$

and $|||A|||_2 = \sigma_1$.

Theorem 7.75. *Let A and B be two matrices in $\mathbb{C}^{m\times n}$. If $A \sim_u B$, then they have the same singular values.*

Proof. Suppose that $A \sim_u B$, that is, $A = W_1^{\mathsf{H}} B W_2$ for some unitary matrices W_1 and W_2. If A has the SVD $A = U^{\mathsf{H}} \mathsf{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) V$, then

$$B = W_1 A W_2^{\mathsf{H}} = (W_1 U^{\mathsf{H}}) \mathsf{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)(V W_2^{\mathsf{H}}).$$

Since $W_1 U^{\mathsf{H}}$ and $V W_2^{\mathsf{H}}$ are both unitary matrices, it follows that the singular values of B are the same as the singular values of A.

Let $\mathbf{v} \in \mathbb{C}^n$ be an eigenvector of the matrix $A^{\mathsf{H}} A$ that corresponds to a non-zero, positive eigenvalue σ^2, that is, $A^{\mathsf{H}} A \mathbf{v} = \sigma^2 \mathbf{v}$.

Define $\mathbf{u} = \frac{1}{\sigma} A \mathbf{v}$. We have $A\mathbf{v} = \sigma \mathbf{u}$. Also,

$$A^{\mathsf{H}} \mathbf{u} = A^{\mathsf{H}} \left(\frac{1}{\sigma} A \mathbf{v} \right) = \sigma \mathbf{v}.$$

This implies $AA^{\mathsf{H}} \mathbf{u} = \sigma^2 \mathbf{u}$, so $\mathbf{u}$ is an eigenvector of AA^{H} that corresponds to the same eigenvalue σ^2.

Conversely, if $\mathbf{u} \in \mathbb{C}^m$ is an eigenvector of the matrix AA^{H} that corresponds to a non-zero, positive eigenvalue σ^2, we have $AA^{\mathsf{H}}\mathbf{u} = \sigma^2\mathbf{u}$. Thus, if $\mathbf{v} = \frac{1}{\sigma}A\mathbf{u}$ we have $A\mathbf{v} = \sigma\mathbf{u}$ and $\mathbf{v}$ is an eigenvector of $A^{\mathsf{H}}A$ for the eigenvalue σ^2.

The Courant-Fisher Theorem (Theorem 7.50) allows the formulation of a similar result for singular values.

Theorem 7.76. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix such that $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$ is the non-increasing sequence of singular values of A. For $1 \leqslant k \leqslant r$ we have*

$$\sigma_k = \min_{\dim(S)=n-k+1} \max\{\| A\boldsymbol{x} \|_2 \mid \boldsymbol{x} \in S \text{ and } \| \boldsymbol{x} \|_2 = 1\}$$
$$\sigma_k = \max_{\dim(T)=k} \min\{\| A\boldsymbol{x} \|_2 \mid \boldsymbol{x} \in T \text{ and } \| \boldsymbol{x} \|_2 = 1\},$$

where S and T range over subspaces of $\mathbb{C}^n$.

Proof. We give the argument only for the second equality of the theorem; the first can be shown in a similar manner.

We saw that σ_k equals the square root of k^{th} largest absolute value of the eigenvalue $|\lambda_k|$ of the matrix $A^{\mathsf{H}}A$. By Courant-Fisher Theorem, we have

$$\begin{aligned}\lambda_k &= \max_{dim(T)=k} \min_{\mathbf{x}}\{\mathbf{x}^{\mathsf{H}}A^{\mathsf{H}}A\mathbf{x} \mid \mathbf{x} \in T \text{ and } \| \mathbf{x} \|_2 = 1\} \\ &= \max_{dim(T)=k} \min_{\mathbf{x}}\{\| A\mathbf{x} \|_2^2 \mid \mathbf{x} \in T \text{ and } \| \mathbf{x} \|_2 = 1\},\end{aligned}$$

which implies the second equality of the theorem.

The equalities established in Theorem 7.76 can be rewritten as

$$\begin{aligned}\sigma_k &= \min_{\mathbf{w}_1,\ldots,\mathbf{w}_{k-1}} \max\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{k-1} \text{ and } \| \mathbf{x} \|_2 = 1\} \\ &= \max_{\mathbf{w}_1,\ldots,\mathbf{w}_{n-k}} \min\{\| A\mathbf{x} \|_2 \mid \mathbf{x} \perp \mathbf{w}_1, \ldots, \mathbf{x} \perp \mathbf{w}_{n-k} \text{ and } \| \mathbf{x} \|_2 = 1\}.\end{aligned}$$

Corollary 7.77. *The smallest singular value of a matrix $A \in \mathbb{C}^{m\times n}$ equals*

$$\min\{\| A\boldsymbol{x} \|_2 \mid \boldsymbol{x} \in \mathbb{C}^n \text{ and } \| \boldsymbol{x} \|_2 = 1\}.$$

The largest singular value of a matrix $A \in \mathbb{C}^{m\times n}$ equals

$$\max\{\| A\boldsymbol{x} \|_2 \mid \boldsymbol{x} \in \mathbb{C}^n \text{ and } \| \boldsymbol{x} \|_2 = 1\}.$$

Proof. The corollary is a direct consequence of Theorem 7.76.

The SVD allows us to find the best approximation of of a matrix by a matrices of limited rank. The central result of this section is Theorem 7.79.

Lemma 7.78. *Let $A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathsf{H}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathsf{H}}$ be the SVD of a matrix $A \in \mathbb{R}^{m\times n}$, where $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$. For every k, $1 \leqslant k \leqslant r$ the matrix $B(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{H}}$ has rank k.*

Proof. The null space of the matrix $B(k)$ consists of those vectors $\mathbf{x}$ such that $\sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}} \mathbf{x} = \mathbf{0}$. The linear independence of the vectors $\mathbf{u}_i$ and the fact that $\sigma_i > 0$ for $1 \leqslant i \leqslant r$ implies the equalities $\mathbf{v}_i^{\mathsf{H}} \mathbf{x} = \mathbf{0}$ for $1 \leqslant i \leqslant r$. Thus,

$$\mathsf{NullSp}(B(k)) = \mathsf{NullSp}\left((\mathbf{v}_1 \ \cdots \ \mathbf{v}_k)\right).$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent it follows that $\dim(\mathsf{NullSp}(B(k)) = n - k$, which implies $\mathit{rank}(B(k)) = k$ for $1 \leqslant k \leqslant r$.

Theorem 7.79. (Eckhart-Young Theorem) *Let $A \in \mathbb{C}^{m \times n}$ be a matrix whose sequence of non-zero singular values is $(\sigma_1, \ldots, \sigma_r)$. Assume that $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$ and that A can be written as*

$$A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathsf{H}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathsf{H}}.$$

Let $B(k) \in \mathbb{C}^{m \times n}$ be the matrix defined by

$$B(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{H}}.$$

If $r_k = \inf\{\|A - X\|_2 \mid X \in \mathbb{C}^{m \times n}$ and $\mathit{rank}(X) \leqslant k\}$, then

$$\|A - B(k)\|_2 = r_k = \sigma_{k+1},$$

for $1 \leqslant k \leqslant r$, where $\sigma_{r+1} = 0$ and $B(k)$ is the best approximation of A among the matrices of rank no larger than k in the sense of the norm $\|\cdot\|_2$.

Proof. Observe that

$$A - B(k) = \sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}},$$

and the largest singular value of the matrix $\sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{H}}$ is σ_{k+1}. Since σ_{k+1} is the largest singular value of $A - B(k)$ we have $\|A - B(k)\|_2 = \sigma_{k+1}$ for $1 \leqslant k \leqslant r$.

We prove now that for every matrix $X \in \mathbb{C}^{m \times n}$ such that $\mathit{rank}(X) \leqslant k$, we have $\|A - X\|_2 \geqslant \sigma_{k+1}$. Since $\dim(\mathsf{NullSp}(X)) = n - \mathit{rank}(X)$, it follows that $\dim(\mathsf{NullSp}(X)) \geqslant n - k$. If T is the subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$, we have $\dim(T) = k+1$. Since $\dim(\mathsf{NullSp}(X)) + \dim(T) > n$, the intersection of these subspaces contains a non-zero vector and, without loss of generality, we can assume that this vector is a unit vector $\mathbf{x}$.

We have $\mathbf{x} = a_1 \mathbf{v}_1 + \cdots a_k \mathbf{v}_k + a_{k+1} \mathbf{v}_{k+1}$ because $\mathbf{x} \in T$. The orthogonality of $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$ implies $\| \mathbf{x} \|_2^2 = \sum_{i=1}^{k+1} |a_i|^2 = 1$.

Since $\mathbf{x} \in \mathsf{NullSp}(X)$, we have $X\mathbf{x} = \mathbf{0}$, so

$$(A - X)\mathbf{x} = A\mathbf{x} = \sum_{i=1}^{k+1} a_i A \mathbf{v}_i = \sum_{i=1}^{k+1} a_i \sigma_i \mathbf{u}_i.$$

Thus, we have

$$\|(A - X)\mathbf{x}\|_2^2 = \sum_{i=1}^{k+1} |\sigma_i a_i|^2 \geqslant \sigma_{k+1}^2 \sum_{i=1}^{k+1} |a_i|^2 = \sigma_{k+1}^2,$$

because $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are also orthonormal. This implies $\|A - X\|_2 \geqslant \sigma_{k+1} = \|A - B(k)\|_2$.

It is interesting to observe that the matrix $B(k)$ provides an optimal approximation of A not only with respect to $\|\cdot\|_2$ but also relative to the Frobenius norm.

Theorem 7.80. *Using the notations introduced in Theorem 7.79, $B(k)$ is the best approximation of A among matrices of rank no larger than k in the sense of the Frobenius norm.*

Proof. Note that $\| A - B(k) \|_F^2 = \| A \|_F^2 - \sum_{i=1}^k \sigma_i^2$. Let X be a matrix of rank k, which can be written as $X = \sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i^{\mathsf{H}}$. Without loss of generality we may assume that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are orthonormal. If this is not the case, we can use the Gram-Schmidt algorithm to express then as linear combinations of orthonormal vectors, replace these expressions in $\sum_{i=1}^k \mathbf{x}_i \mathbf{y}_i^{\mathsf{H}}$ and rearrange the terms. Now, the Frobenius norm of $A - X$ can be written as

$$\begin{aligned}
\| A - X \|_F^2 &= \mathit{trace}\left(\left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}^{\mathsf{H}}\right)^{\mathsf{H}} \left(A - \sum_{i=1}^k \mathbf{x}_i \mathbf{y}^{\mathsf{H}}\right)\right) \\
&= \mathit{trace}\left(A^{\mathsf{H}} A + \sum_{i=1}^k (\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)(\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)^{\mathsf{H}} - \sum_{i=1}^k A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}} A\right).
\end{aligned}$$

Taking into account that $\sum_{i=1}^k (\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)(\mathbf{y}_i - A^{\mathsf{H}}\mathbf{x}_i)^{\mathsf{H}}$ is a real non-negative number and that $\sum_{i=1}^k A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}} A = \| A\mathbf{x}_i \|_F^2$ we have

$$\| A - X \|_F^2 \geq \mathit{trace}\left(A^{\mathsf{H}} A - \sum_{i=1}^k A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}} A\right) = \| A \|_F^2 - \mathit{trace}\left(\sum_{i=1}^k A^{\mathsf{H}}\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}} A\right).$$

Let $A = U\mathsf{diag}(\sigma_1, \ldots, \sigma_n)V^{\mathsf{H}}$ be the singular value decomposition of A. If $V = (V_1\ V_2)$, where V_1 has k columns $\mathbf{v}_1, \ldots, \mathbf{v}_k$, $D_1 = \mathsf{diag}(\sigma_1, \ldots, \sigma_k)$ and $D_2 = \mathsf{diag}(\sigma_{k+1}, \ldots, \sigma_n)$, then we can write

$$\begin{aligned}
A^{\mathsf{H}} A = V D^{\mathsf{H}} U^{\mathsf{H}} U D V^{\mathsf{H}} &= (V_1\ V_2) \begin{pmatrix} D_1^2 & O \\ O & D_2^2 \end{pmatrix} \begin{pmatrix} V_1^{\mathsf{H}} \\ V_2^{\mathsf{H}} \end{pmatrix} \\
&= V_1 D_1^2 V_1^{\mathsf{H}} + V_2 D_2^2 V_2^{\mathsf{H}}.
\end{aligned}$$

and $A^{\mathsf{H}} A = V D^2 V^{\mathsf{H}}$. These equalities allow us to write:

$$\begin{aligned}
\parallel A\mathbf{x}_i \parallel_F^2 &= \mathit{trace}(\mathbf{x}_i^{\mathsf{H}} A^{\mathsf{H}} A\mathbf{x}_i) \\
&= \mathit{trace}\left(\mathbf{x}_i^{\mathsf{H}} V_1 D_1^2 V_1^{\mathsf{H}} \mathbf{x}_i + \mathbf{x}_i^{\mathsf{H}} V_2 D_2^2 V_2^{\mathsf{H}} \mathbf{x}_i\right) \\
&= \parallel D_1 V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 + \parallel D_2 V_2^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 \\
&= \sigma_k^2 + \left(\parallel D_1 V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2\right) \\
&\quad - \left(\sigma_k^2 \parallel V_2^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \parallel D_2 V_2^{\mathsf{H}} \mathbf{x}_i \parallel_F^2)\right) - \sigma_k^2 (1 - \parallel V^{\mathsf{H}} \mathbf{x}_i \parallel).
\end{aligned}$$

Since $\parallel V^{\mathsf{H}} \mathbf{x}_i \parallel_F^1 = 1$ (because $\mathbf{x}_i$ is an unit vector and V is an unitary matrix) and $\sigma_k^2 \parallel V_2^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \parallel D_2 V_2^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 \geqslant 0$, it follows that

$$\parallel A\mathbf{x}_i \parallel_F^2 \leqslant \sigma_k^2 + \left(\parallel D_1 V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2\right).$$

Consequently,

$$\begin{aligned}
\sum_{i=1}^{k} \parallel A\mathbf{x}_i \parallel_F^2 &\leqslant k\sigma_k^2 + \sum_{i=1}^{k} \left(\parallel D_1 V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2 - \sigma_k^2 \parallel V_1^{\mathsf{H}} \mathbf{x}_i \parallel_F^2\right) \\
&= k\sigma_k^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} (\sigma_j^2 - \sigma_k^2) |\mathbf{v}_j^{\mathsf{H}} \mathbf{x}_i|^2 \\
&= \sum_{j=1}^{k} \left(\sigma_k^2 + (\sigma_j^2 - \sigma_k^2) \sum_{i=1}^{k} |\mathbf{v}_j \mathbf{x}_i|^2\right) \\
&\leqslant \sum_{j=1}^{k} (\sigma_k^2 + (\sigma_j^2 - \sigma_k^2)) = \sum_{j=1}^{k} \sigma_j^2,
\end{aligned}$$

which concludes the argument.

Definition 7.81. *Let $A \in \mathbb{C}^{m\times n}$. The* numerical rank *of A is the function $\mathsf{nr}_A : [0,\infty) \longrightarrow \mathbb{N}$ given by*

$$\mathsf{nr}_A(d) = \min\{\mathit{rank}(B) \mid \|A - B\|_2 \leqslant d\}$$

for $d \geqslant 0$.

Theorem 7.82. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix having the sequence of non-zero singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$. Then, $\mathsf{nr}_A(d) = k < r$ if and only if $\sigma_k > d \geqslant \sigma_{k+1}$.*

Proof. Let d be a number such that $\sigma_k > d \geqslant \sigma_{k+1}$. Equivalently, by Eckhart-Young Theorem, we have

$$\|A - B(k-1)\|_2 > d \geqslant \|A - B(k)\|_2,$$

Since $\|A - B(k-1)\|_2 = \min\{\|A - X\|_2 \mid \mathit{rank}(X) = k-1\} > d$, it follows that $\min\{\mathit{rank}(B) \mid \|A - B\|_2 \leqslant d\} = k$, so $\mathsf{nr}_A(d) = k$.

Conversely, suppose that $\mathsf{nr}_A(d) = k$. This means that the minimal rank of a matrix B such that $\|A - B\|_2 \leqslant d$ is k. Therefore, $\|A - B(k-1)\|_2 > d$. On another hand, $d \geqslant \|A - B(k)\|_2$ because there exists a matrix C of rank k such that $d \geqslant \|A - C\|_2$, so $d \geqslant \|A - B(k)\|_2 = \sigma_{k+1}$. Thus, $\sigma_k > d \geqslant \sigma_{k+1}$.

Exercises and Supplements

1. Prove that if $(a, \mathbf{x})$ is an eigenpair of a matrix $A \in \mathbb{C}^{n\times n}$ if and only if $(a-b, \mathbf{x})$ is an eigenpair of the matrix $A - bI_n$.
2. Let $A \in \mathbb{R}^{n\times n}$ be a matrix and let $(a, \mathbf{x})$ be an eigenpair of A and $(a, \mathbf{y})$ be an eigenpair of A' such that $\mathbf{x}'\mathbf{y} = 1$. If $L = \mathbf{x}\mathbf{y}'$, prove that
   a) every non-zero eigenvalue of $A - aL$ is also an eigenvalue of A and every eigenpair $(\lambda, \mathbf{t})$ of $A - aL$ is an eigenpair of A;
   b) if $a \neq 0$ is an eigenvalue of A with $\mathsf{geomm}(A, a) = 1$, then a is not an eigenvalue of $A - aL$.

   Solution: Let λ a non-zero eigenvalue of $A - aL$. We have $(A - aL)\mathbf{t} = \lambda\mathbf{t}$ for some $\mathbf{t} \neq \mathbf{0}_n$. By Part (d) of Exercise 24 of Chapter 5, $L(A - aL) = O_{n,n}$, so $\lambda L\mathbf{t} = \mathbf{0}_n$, so $L\mathbf{t} = \mathbf{0}$, which implies $A\mathbf{t} = \lambda\mathbf{t}$.

   For the second part suppose that $a \neq 0$ were an eigenvalue of $A - aL$ and let $(a, \mathbf{w})$ be an eigenpair of this matrix. By the first part, $A\mathbf{w} = a\mathbf{w}$. Since $\mathsf{geomm}(A, a) = 1$, there exists $b \in \mathbb{C} - \{0\}$ such that $\mathbf{w} = b\mathbf{x}$. This allows us to write

$$\begin{aligned} a\mathbf{w} &= (A - aL)\mathbf{w} = (A - aL)b\mathbf{x} = ab\mathbf{x} - abL\mathbf{x} = ab\mathbf{x} - ab(\mathbf{x}\mathbf{y}')\mathbf{x} \\ &= ab\mathbf{x} - ab\mathbf{x}(\mathbf{y}'\mathbf{x}) = \mathbf{0}_n, \end{aligned}$$

   because $\mathbf{y}'\mathbf{x} = 1$. Since $a \neq 0$ and $\mathbf{x} \neq \mathbf{0}_n$, this is impossible.
3. Let $A \in \mathbb{R}^{n\times n}$ be a matrix, $(\lambda, \mathbf{x})$ be an eigenpair of A and $(\lambda, \mathbf{y})$ be an eigenpair of A' such that:
   (i) $\mathbf{x}'\mathbf{y} = 1$;
   (ii) λ be an eigenvalue of A with $|\lambda| = \rho(A)$ and λ is unique with this property.
   If $L = \mathbf{x}\mathbf{y}'$, and $\theta \in \mathsf{spec}(A)$ is such that $|\theta| < \rho(A)$ and $|\theta|$ is maximal with this property, prove that:
   a) $\rho(A - \lambda L) \leqslant |\theta| < \rho(A)$;
   b) $\left(\frac{1}{\lambda}A\right)^m = L + \left(\frac{1}{\lambda}A - L\right)^m$ and $\lim_{m\to\infty} \left(\frac{1}{\lambda}A\right)^m = L$.

   Solution: By Part (a) of Supplement 2, every non-zero eigenvalue of $A - \lambda L$ is an eigenvalue of A. Therefore, either $\rho(A - \lambda L) = 0$ or $\rho(A - \lambda L) = |\lambda'|$ for some $\lambda' \in \mathsf{spec}(A)$. Therefore, in either case, $\rho(A - \lambda L) \leqslant |\theta| < \rho(A)$.

   Since $(A - \lambda L)^m = A^m - \lambda^m L$, we have $\left(\frac{1}{\lambda}A\right)^m = L + \left(\frac{1}{\lambda}A - L\right)^m$. Note that $\rho\left(\frac{1}{\lambda}A - L\right) = \frac{\rho(A-\lambda L)}{\rho(A)} \leqslant \frac{|\theta|}{\rho(A)} < 1$. Therefore, by Theorem 7.68, $\lim_{m\to\infty} \left(\frac{1}{\lambda}A\right)^m = L$.
4. Let $A \in \mathbb{R}^{n\times n}$, where n is an odd number. Prove that A has at least one real eigenvalue.
5. Prove that the eigenvalues of an upper triangular (or lower triangular) matrix are its diagonal entries.

Let $\mathsf{s}_k : \mathbb{C}^n \longrightarrow \mathbb{C}$ be the k^{th} symmetric function of n arguments defined by

$$\mathsf{s}_k^n(z_1, \ldots, z_n) = \sum_{i_1,\ldots,i_k} \left\{ \prod_{j=1}^{k} \mathbf{z}_{i_j} \;\middle|\; 1 \leqslant i_1 < \cdots < i_k \leqslant n \right\},$$

for $z_1, \ldots, z_n \in \mathbb{C}$. For example, we have

$$\begin{aligned}
\mathsf{s}_1^3(z_1, z_2, z_3) &= z_1 + z_2 + z_3,\\
\mathsf{s}_2^3(z_1, z_2, z_3) &= z_1 z_2 + z_1 z_3 + z_2 z_3,\\
\mathsf{s}_3^3(z_1, z_2, z_3) &= z_1 z_2 z_3.
\end{aligned}$$

6. Prove that

$$\begin{aligned}(t - z_1)\cdots(t - z_n) = t^n - \mathsf{s}_1^n(z_1, \ldots, z_n)t^{n-1} + \mathsf{s}_2^n(z_1, \ldots, z_n)t^{n-2} - \cdots\\ +(-1)^n\, \mathsf{s}_n^n(z_1, \ldots, z_n).\end{aligned}$$

Solution: The equality follows by observing that the coefficient of t^{n-k} in $(t - z_1)\cdots(t - z_n)$ equals $(-1)^k\, \mathsf{s}_k^n(z_1, \ldots, z_n)$.

7. Let $M \in \mathbb{C}^{n\times n}$ be an invertible matrix and let $\|\!| \cdot |\!\|$ be a matrix norm. Prove that the mapping $\mu_M : \mathbb{C}^{n\times n} \longrightarrow \mathbb{R}_{\geqslant 0}$ given by $\mu_M(A) = \|\!| M^{-1}AM |\!\|$ is a matrix norm.
8. Let $A \in \mathbb{C}^{n\times n}$ be a matrix. Prove that:
   a) if $\lambda \in \mathsf{spec}(A)$, then $1 + \lambda \in \mathsf{spec}(I_n + A)$;
   b) $\mathsf{algm}(I_n + A, 1 + \lambda) = \mathsf{algm}(A, \lambda)$;
   c) $\rho(I_n + A) \leqslant 1 + \rho(A)$.

   Solution: Suppose that $\lambda \in \mathsf{spec}(A)$ and $\mathsf{algm}(A, \lambda) = k$. Then λ is root of multiplicity k of $p_A(\lambda) = \det(\lambda I_n - A)$. Since $p_{I_n+A}(\lambda) = \det(\lambda I_n - I_n - A)$, it follows that $p_{I_n+A}(1 + \lambda) = \det(\lambda I_n - A) = p_A(\lambda)$. Thus $1 + \lambda \in \mathsf{spec}(I_n + A)$ and $\mathsf{algm}(I_n + A, 1 + \lambda) = \mathsf{algm}(A, \lambda)$.

   We have $\rho(I_n + A) = \max\{|1 + \lambda| \mid \lambda \in \mathsf{spec}(A)\} \leqslant 1 + \max\{|\lambda| \mid \lambda \in \mathsf{spec}(A) = 1 + \rho(A)$.
9. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. Prove that
$$\left\|\!\left| A - \frac{\lambda_1 + \lambda_n}{2} I_n \right|\!\right\|_2 = \frac{\lambda_1 - \lambda_n}{2}.$$
10. Let $A \in \mathbb{R}^{n\times n}$ be a matrix and let $c \in \mathbb{R}$. Prove that for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n - \{\mathbf{0}_n\}$ we have $\mathsf{ral}_A(\mathbf{x}) - \mathsf{ral}_A(\mathbf{y}) = \mathsf{ral}_B(\mathbf{x}) - \mathsf{ral}_B(\mathbf{y})$, where $B = A + cI_n$.
11. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix having the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in $\mathbb{R}^n - \{\mathbf{0}_n\}$. Prove that
$$|\mathsf{ral}_A(\mathbf{x}) - \mathsf{ral}_A(\mathbf{y})| \leqslant (\lambda_1 - \lambda_n)\sin\angle(\mathbf{x}, \mathbf{y}).$$

    Solution: Assume that $\| \mathbf{x} \|=\| \mathbf{y} \|= 1$ and let $B = A - \frac{\lambda_1+\lambda_n}{2} I_n$. By Exercise 10, we have
$$|\mathsf{ral}_A(\mathbf{x}) - \mathsf{ral}_A(\mathbf{y})| = |\mathsf{ral}_B(\mathbf{x}) - \mathsf{ral}_B(\mathbf{y})| == |\mathbf{x}'B\mathbf{x} - \mathbf{y}'B\mathbf{y}| = |B(\mathbf{x} - \mathbf{y})'(\mathbf{x} + \mathbf{y})|.$$
    By Cauchy-Schwarz Inequality we have
$$|B(\mathbf{x} - \mathbf{y})'(\mathbf{x} + \mathbf{y})| \leqslant 2 \parallel B \parallel \frac{\| \mathbf{x} - \mathbf{y} \| \, \| \mathbf{x} + \mathbf{y} \|}{2} = (\lambda_1 - \lambda_n)\sin\angle(\mathbf{x}, \mathbf{y}).$$
12. Prove that if A is a unitary matrix and $1 \notin \mathsf{spec}(A)$, then there exists a skew-Hermitian S such that $A = (I_n - S)(I_n + S)^{-1}$.
13. Let $f : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$ be a function such that $f(AB) = f(BA)$ for $A, B \in \mathbb{R}^{n\times n}$. Prove that if $A \sim B$, then $f(A) = f(B)$.

14. Let $a, b \in \mathbb{C} - \{0\}$ and let $B_r(\lambda, a) \in \mathbb{C}^{r\times r}$ be the matrix defined by

$$B_r(\lambda, a) = \begin{pmatrix} \lambda & a & 0 & \cdots & 0 \\ 0 & \lambda & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix} \in \mathbb{C}^{r\times r}.$$

Prove that
 a) $B_n(\lambda, a) \sim B_n(\lambda, b)$;
 b) $B_r(\lambda)$ is given by

$$(B_r(\lambda)^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \binom{k}{r-1}\lambda^{k-r+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{r-2}\lambda^{k-r+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^k \end{pmatrix};$$

 c) if $|\lambda| < 1$, then $\lim_{k\to\infty} B_r(\lambda)^k = O$.

15. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric real matrix. Prove that

$$\nabla \mathsf{ral}_A(\mathbf{x}) = \frac{2}{\mathbf{x}'\mathbf{x}}(A\mathbf{x} - \mathsf{ral}_A(\mathbf{x})\mathbf{x}).$$

Also, show that the eigenvectors of A are the stationary points of the function $\mathsf{ral}_A(\mathbf{x})$.

16. Let $A, B \in \mathbb{C}^{n\times n}$ be two Hermitian matrices. Prove that AB is a Hermitian matrix if and only if $AB = BA$.

17. Let $A \in \mathbb{R}^{3\times 3}$ be a symmetric matrix. Prove that if $trace(A) \neq 0$, the sum of principal minors of order 2 equals 0, and $\det(A) = 0$, then $rank(A) = 1$.

 Solution: The characteristic polynomial of A is $p_A(\lambda) = \lambda^3 - trace(A)\lambda^2 = 0$. Thus, $\mathsf{spec}(A) = \{trace(A), 0\}$, where $\mathsf{algm}(A, 0) = 2$, so $rank(A) = 1$.

18. Let $A \in \mathbb{R}^{3\times 3}$ be a symmetric matrix. Prove that if the sum of principal minors of order 2 does not equal 0 but $\det(A) = 0$, then $rank(A) = 2$.

19. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, $\mathbf{u} \in \mathbb{C}^n$ be a vector and let a be a complex number. Define the Hermitian matrix B as

$$B = \begin{pmatrix} A & \mathbf{u} \\ \mathbf{u}^{\mathsf{H}} & a \end{pmatrix}.$$

Let $\alpha_1 \leqslant \cdots \leqslant \alpha_n$ be the eigenvalues of A and let $\beta_1 \leqslant \cdots \leqslant \beta_n \leqslant \beta_{n+1}$ be the eigenvalues of B. Prove that

$$\beta_1 \leqslant \alpha_1 \leqslant \beta_2 \leqslant \cdots \leqslant \beta_n \leqslant \alpha_n \leqslant \beta_{n+1}.$$

 Solution: Since $B \in \mathbb{C}^{(n+1)\times(n+1)}$, by Courant-Fisher Theorem we have

$$\begin{aligned}\beta_{k+1} &= \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W\rangle^{\perp}\} \\ &= \max_Z \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z\rangle^{\perp}\},\end{aligned}$$

where W ranges of sets of k non-zero arbitrary vectors, and let Z be a subset of $\mathbb{C}^n$ that consists of $n - k$ non-zero arbitrary vectors in $\mathbb{C}^{n+1}$.

Let U be a set of k non-zero vectors in $\mathbb{C}^n$ and let Y be a set of $n-k-1$ vectors in $\mathbb{C}^n$. Define the subsets W_U and Z_Y of $\mathbb{C}^{n+1}$ as

$$W_U = \left\{ \begin{pmatrix} \mathbf{u} \\ 0 \end{pmatrix} \middle| \mathbf{u} \in U \right\}$$

and

$$Z_Y = \left\{ \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix} \middle| \mathbf{y} \in Y \right\} \cup \{\mathbf{e}_{n+1}\}.$$

By restricting the sets W and Z to sets of the form W_U and Z_Y we obtain the double inequality

$$\begin{aligned} &\max_{Z_Y} \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z_Y \rangle^{\perp}\} \\ &\leqslant \beta_{k+1} \leqslant \min_{W_U} \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W_U \rangle^{\perp}\}. \end{aligned}$$

Note that, if $\mathbf{x} \in \langle Z_Y \rangle^{\perp}$, then we have $\mathbf{x} \perp \mathbf{e}_{n+1}$, so $x_{n+1} = 0$. Therefore,

$$\mathbf{x}^{\mathsf{H}} B\mathbf{x} = (\mathbf{y}^{\mathsf{H}} 0) \begin{pmatrix} A & \mathbf{u} \\ \mathbf{u}^{\mathsf{H}} & a \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix} = \mathbf{y}^{\mathsf{H}} A\mathbf{y}.$$

Consequently,

$$\begin{aligned} &\max_{Z_Y} \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z_Y \rangle^{\perp}\} \\ &= \max_{Y} \min_{\mathbf{y}} \{\mathbf{y}^{\mathsf{H}} A\mathbf{y} \mid \parallel \mathbf{y} \parallel_2 = 1 \text{ and } \mathbf{y} \in \langle Y \rangle^{\perp}\} = \alpha_k. \end{aligned}$$

This allows us to conclude that $\alpha_k \leqslant \beta_{k+1}$ for $1 \leqslant k \leqslant n$.

On another hand, if $\mathbf{x} \in \langle W_U \rangle^{\perp}$ and

$$\mathbf{x} = \begin{pmatrix} \mathbf{u} \\ 0 \end{pmatrix}$$

then $\mathbf{x}^{\mathsf{H}} B\mathbf{x} = \mathbf{u}^{\mathsf{H}} A\mathbf{u}$ and $\parallel \mathbf{x} \parallel_2 = \parallel \mathbf{u} \parallel_2$. Now we can write

$$\begin{aligned} &\min_{W_U} \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} B\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W_U \rangle^{\perp}\} \\ &= \min_{U} \max_{\mathbf{u}} \{\mathbf{u}^{\mathsf{H}} A\mathbf{u} \mid \parallel \mathbf{u} \parallel_2 = 1 \text{ and } \mathbf{u} \in \langle U \rangle^{\perp}\} = \alpha_{k+1}, \end{aligned}$$

so $\beta_{k+1} \leqslant \alpha_{k+1}$ for $1 \leqslant k \leqslant n-1$.

20. Let $A, B \in \mathbb{C}^{n \times n}$ be two matrices such that $AB = BA$. Prove that A and B have a common eigenvector.

Solution: Let $\lambda \in \mathsf{spec}(A)$ and let $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be a basis for $\mathsf{NullSp}(A - \lambda I_n)$. Observe that the matrices $A - \lambda I_n$ and B commute because

$$(A - \lambda I_n)B = AB - \lambda B \text{ and } B(A - \lambda I_n) = BA - \lambda B.$$

Therefore, we have

$$(A - \lambda I_n)B\mathbf{x}_i = B(A - \lambda I_n)\mathbf{x}_i = \mathbf{0},$$

so $(A - \lambda I_n)BX = O_{n,n}$, where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Consequently, $ABX = \lambda BX$. Let $\mathbf{y}_1, \ldots, \mathbf{y}_m$ be the columns of the matrix BX. The last equality implies that

$A\mathbf{y}_i = \lambda\mathbf{y}_i$, so $\mathbf{y}_i \in \mathsf{NullSp}(A - \lambda I_n)$. Since X is a basis of $\mathsf{NullSp}(A - \lambda I_n)$ it follows that each $\mathbf{y}_i$ is a linear combination of the columns of X so there exists a matrix P such that $(\mathbf{y}_1 \cdots \mathbf{y}_m) = (\mathbf{x}_1 \cdots \mathbf{x}_k)P$, which is equivalent to $BX = XP$. Let $\mathbf{w}$ be an eigenvector of P. We have $P\mathbf{w} = \mu\mathbf{w}$. Consequently, $BX\mathbf{w} = XP\mathbf{w} = \mu X\mathbf{w}$, which proves that $X\mathbf{w}$ is an eigenvector of B. Also, $A(X\mathbf{w}) = A(BX\mathbf{w}) = (\lambda BX)\mathbf{w} = \lambda\mu X\mathbf{w}$, so $X\mathbf{w}$ is also an eigenvector of A.

21. Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ be two matrices. Prove that the set of non-zero eigenvalues of the matrices $AB \in \mathbb{C}^{m\times m}$ and $BA \in \mathbb{C}^{n\times n}$ are the same and $\mathsf{algm}(AB, \lambda) = \mathsf{algm}(BA, \lambda)$ for each such eigenvalue.

Solution: Consider the following straightforward equalities:

$$\begin{pmatrix} I_m & -A \\ O_{n,m} & \lambda I_n \end{pmatrix}\begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} = \begin{pmatrix} \lambda I_m - AB & O_{m,n} \\ -\lambda B & \lambda I_n \end{pmatrix}$$
$$\begin{pmatrix} -I_m & O_{m,n} \\ -B & \lambda I_n \end{pmatrix}\begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} = \begin{pmatrix} -\lambda I_m & -A \\ O_{n,m} & \lambda I_n - BA \end{pmatrix}.$$

Observe that

$$\det\left(\begin{pmatrix} I_m & -A \\ O_{n,m} & \lambda I_n \end{pmatrix}\begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix}\right) = \det\left(\begin{pmatrix} -I_m & O_{m,n} \\ -B & \lambda I_n \end{pmatrix}\begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix}\right),$$

and therefore,

$$\det\begin{pmatrix} \lambda I_m - AB & O_{m,n} \\ -\lambda B & \lambda I_n \end{pmatrix} = \det\begin{pmatrix} -\lambda I_m & -A \\ O_{n,m} & \lambda I_n - BA \end{pmatrix}.$$

The last equality amounts to $\lambda^n p_{AB}(\lambda) = \lambda^m p_{BA}(\lambda)$. Thus, for $\lambda \neq 0$ we have $p_{AB}(\lambda) = p_{BA}(\lambda)$, which gives the desired conclusion.

22. Let $\mathbf{a} \in \mathbb{C}^n - \{\mathbf{0}_n\}$. Prove that the matrix $\mathbf{a}\mathbf{a}^{\mathsf{H}} \in \mathbb{C}^{n\times n}$ has one eigenvalue distinct from 0, and this eigenvalue is equal to $\parallel \mathbf{a} \parallel^2$.

23. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix such that $a_{ij} \in \{0, 1\}$ for $1 \leqslant i, j \leqslant n$. If $d = \frac{|\{a_{ij} \mid a_{ij}=1\}|}{n^2}$, prove that $n\sqrt{d} \geqslant \lambda_1 \geqslant nd$, where λ_1 is the largest eigenvalue of A.

Solution: By Rayleigh-Ritz Theorem (Theorem 7.49) we have $\lambda_1 \mathbf{1}_n'\mathbf{1}_n \geqslant \mathbf{1}_n' A\mathbf{1}_n$. Since $\mathbf{1}_n'\mathbf{1}_n = n$ and $\mathbf{1}_n' A\mathbf{1}_n = n^2 d$ it follows that $\lambda_1 \geqslant nd$. On another hand we have $\sum_{i=1}^n \lambda_i^2 \leqslant \parallel A \parallel_F^2 = n^2 d$, so $\lambda_1 \leqslant n\sqrt{d}$.

24. Let $U = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \in \mathbb{C}^{n\times r}$ be a matrix having an orthonormal set of columns. For $A \in \mathbb{C}^{n\times n}$ define the matrix $A_U \in \mathbb{C}^{r\times r}$ by $A_U = U^{\mathsf{H}} AU$.
    a) Prove that if A is a Hermitian matrix, then A_U is also a Hermitian matrix.
    b) If A is a Hermitian matrix having the eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ and A_U has the eigenvalues $\mu_1 \leqslant \cdots \leqslant \mu_r$, prove that
    $$\lambda_k \leqslant \mu_k \leqslant \lambda_{k+n-r}.$$
    c) Prove that
    $$\sum_{i=1}^{r} \lambda_i = \min\{\mathit{trace}(A_U) \mid U^{\mathsf{H}}U = I_r\}$$
    $$\sum_{i=n-r+1}^{n} \lambda_i = \max\{\mathit{trace}(A_U) \mid U^{\mathsf{H}}U = I_r\}.$$

Solution: Observe that $(A_U)^{\mathsf{H}} = U^{\mathsf{H}} A^{\mathsf{H}} U = U^{\mathsf{H}} A U = A_U$, so A_U is indeed Hermitian and its eigenvalues $\mu_1, \ldots, \mu_r$ are real numbers.

Extend the set of columns $\mathbf{u}_1, \ldots, \mathbf{u}_r$ of U to an orthonormal basis

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$$

and let $W \in \mathbb{C}^{n\times n}$ be the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Since W is a unitary matrix, $\mathsf{spec}(W^{\mathsf{H}} AW) = \mathsf{spec}(A)$ and A_U is a principal submatrix of $\mathsf{spec}(W^{\mathsf{H}} AW)$. The second part follows from Theorem 7.53.

The first equality of the third part follows from the fact that the second part implies

$$\sum_{i=1}^{r} \lambda_i \leqslant \sum_{i=1}^{r} \mu_i = \mathit{trace}(A_U),$$

where $A_U \in \mathbb{C}^{r\times r}$. If the columns $\mathbf{u}_1, \ldots, \mathbf{u}_r$ of U are chosen as orthonormal eigenvectors the above inequality becomes an equality. In this case we have $U^{\mathsf{H}} U = I_r$ and the first equality of the third part follows. The argument for the second equality is similar.

25. Let $A, B \in \mathbb{C}^{n\times n}$ be two Hermitian matrices, where $\mathsf{spec}(A) = \{\xi_1, \ldots, \xi_n\}$, $\mathsf{spec}(B) = \{\zeta_1, \ldots, \zeta_n\}$, and $\mathsf{spec}(A + B) = \{\lambda_1, \ldots, \lambda_n\}$. Also, suppose that $\xi_1 \leqslant \cdots \leqslant \xi_n$, $\zeta_1 \leqslant \cdots \leqslant \zeta_n$, and $\lambda_1 \leqslant \cdots \leqslant \lambda_n$. Prove that for $r \leqslant n$ we have

$$\sum_{i=1}^{r} \lambda_i \leqslant \sum_{i=1}^{r} \xi_i + \sum_{i=1}^{r} \zeta_i$$

and

$$\sum_{i=n-r+1}^{n} \lambda_i \geqslant \sum_{i=n-r+1}^{n} \xi_i + \sum_{i=n-r+1}^{n} \zeta_i.$$

Solution: Supplement 24 implies that

$$\begin{aligned}
\sum_{i=1}^{r} \lambda_i &= \min\{\mathit{trace}((A + B)_U) \mid U^{\mathsf{H}} U = I_r\} \\
&\geqslant \min\{\mathit{trace}(A_U) \mid U^{\mathsf{H}} U = I_r\} + \min\{\mathit{trace}(B_U) \mid U^{\mathsf{H}} U = I_r\} \\
&= \sum_{i=1}^{r} \xi_i + \sum_{i=1}^{r} \zeta_i.
\end{aligned}$$

For the second part we can write

$$\begin{aligned}
\sum_{i=n-r+1}^{n} \lambda_i &= \max\{\mathit{trace}((A + B)_U) \mid U^{\mathsf{H}} U = I_r\} \\
&\leqslant \max\{\mathit{trace}(A_U) \mid U^{\mathsf{H}} U = I_r\} + \max\{\mathit{trace}(B_U) \mid U^{\mathsf{H}} U = I_r\} \\
&= \sum_{i=n-r+1}^{n} \xi_i + \sum_{i=n-r+1}^{n} \zeta_i.
\end{aligned}$$

26. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Prove that A is diagonalizable if and only if $(a - d)^2 + 4bc \neq 0$.

27. Let $A \in \mathbb{C}^{n \times n}$ be a matrix. Prove that the following statements are equivalent:
   a) A is a rank 1 matrix;
   b) A has exactly one non-zero eigenvalue λ with $\mathsf{algm}(A, \lambda) = 1$;
   c) There exist $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n - \{\mathbf{0}\}$ such that $A = \mathbf{x}\mathbf{y}^{\mathsf{H}}$ and $\mathbf{x}^{\mathsf{H}}\mathbf{y}$ is an eigenvalue of A.
28. Prove that the characteristic polynomial of the companion matrix of a polynomial p is p itself.
29. Let $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{k \times k}$, and $X \in \mathbb{C}^{n \times k}$ be three matrices such that $AX = XB$. Prove that
   a) $\mathsf{Ran}(X)$ is an invariant subspace of A;
   b) if $\mathbf{v}$ is an eigenvector of B, then $X\mathbf{v}$ is an eigenvector of A;
   c) if $\mathit{rank}(X) = k$, then $\mathsf{spec}(B) \subseteq \mathsf{spec}(A)$.

The next supplements present a result known as *Weyl's Theorem* (Supplement 30) and several of its important consequences. For a Hermitian matrix $A \in \mathbb{C}^{n \times n}$ we denote its eigenvalues arranged in increasing order as $\lambda_1(A) \leqslant \cdots \leqslant \lambda_n(A)$.

30. Let A and B be two Hermitian matrices in $\mathbb{C}^{n \times n}$. Prove that

$$\lambda_1(B) \leqslant \lambda_k(A + B) - \lambda_k(A) \leqslant \lambda_n(B)$$

for $1 \leqslant k \leqslant n$.

Solution: By the Rayleigh-Ritz Theorem we have

$$\lambda_1(B) \leqslant \mathbf{x}^{\mathsf{H}} B \mathbf{x} \leqslant \lambda_n(B),$$

for $\mathbf{x} \neq \mathbf{0}$ and $\parallel \mathbf{x} \parallel_2 = 1$. Then, by the Courant-Fisher Theorem,

$$\lambda_k(A + B) = \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A + B)\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W \rangle^{\perp}\},$$

where the minimum is taken over sets W that contain $k - 1$ vectors. Since $\mathbf{x}^{\mathsf{H}}(A + B)\mathbf{x} = \mathbf{x}^{\mathsf{H}} A \mathbf{x} + \mathbf{x}^{\mathsf{H}} B \mathbf{x}$, it follows that

$$\begin{aligned} \lambda_k(A + B) &\geqslant \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} A \mathbf{x} + \lambda_n(B) \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W \rangle^{\perp}\} \\ &= \lambda_k(A) + \lambda_n(B). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \lambda_k(A + B) &\leqslant \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}} A \mathbf{x} + \lambda_1(B) \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W \rangle^{\perp}\} \\ &= \lambda_k(A) + \lambda_1(B). \end{aligned}$$

31. Let A and E be Hermitian matrices in $\mathbb{C}^{n \times n}$. Prove that $|\lambda_p(A + E) - \lambda_p(A)| \leqslant \rho(E) = \|E\|_2$ for $1 \leqslant p \leqslant n$.

Solution: By Weyl's inequalities (Supplement 30) we have

$$\lambda_1(E) \leqslant \lambda_p(A + E) - \lambda_p(A) \leqslant \lambda_n(E)$$

By Definition 7.64, this implies $|\lambda_p(A + E) - \lambda_p(A)| \leqslant \rho(E) = \|E\|_2$ because A is Hermitian.

32. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let $\mathbf{w} \in \mathbb{C}^n$. Then, $\lambda_k(A+\mathbf{w}\mathbf{w}^{\mathsf{H}}) \leqslant \lambda_{k+1}(A) \leqslant \lambda_{k+2}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})$ and $\lambda_k(A) \leqslant \lambda_{k+1}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}}) \leqslant \lambda_{k+2}(A)$ for $1 \leqslant k \leqslant n-2$.
Solution: Let W ranging over the subsets of $\mathbb{C}^n$ that consist of $n-k-2$ vectors. By Courant-Fisher Theorem (Theorem 7.50),
$$\begin{aligned}
&\lambda_{k+2}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\\
&= \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W\rangle^{\perp}\}\\
&\geqslant \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1, \mathbf{x} \in \langle W\rangle^{\perp} \text{ and } \mathbf{x} \perp \mathbf{w}\}\\
&= \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1, \mathbf{x} \in \langle W\rangle^{\perp} \text{ and } \mathbf{x} \perp \mathbf{w}\}\\
&\quad \text{(because } \mathbf{x}^{\mathsf{H}}\mathbf{w} = \mathbf{w}^{\mathsf{H}}\mathbf{x} = 0)\\
&\geqslant \min_{W_1} \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W_1\rangle^{\perp}\} = \lambda_{k+1}(A),
\end{aligned}$$
where W_1 ranges over the sets that contain $n-k-1$ vectors.
For $2 \leqslant k \leqslant n-2$, the same Courant-Fisher Theorem yields
$$\lambda_k(A+\mathbf{w}\mathbf{w}^{\mathsf{H}}) = \max_Z \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z\rangle^{\perp}\},$$
where Z is a set that contains $k-1$ vectors. This implies
$$\begin{aligned}
&\lambda_k(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\\
&\leqslant \max_Z \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+\mathbf{w}\mathbf{w}^{\mathsf{H}})\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z\rangle^{\perp} \text{ and } \mathbf{x} \perp \mathbf{w}\}\\
&= \max_Z \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z\rangle^{\perp} \text{ and } \mathbf{x} \perp \mathbf{w}\}\\
&\leqslant \max_{Z_1} \min_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle Z_1\rangle^{\perp}\} = \lambda(k+1, A),
\end{aligned}$$
where Z_1 ranges over sets that contain k vectors.

33. Let A and B be two Hermitian matrices in $\mathbb{C}^{n\times n}$. If $\mathit{rank}(B) \leqslant r$, prove that $\lambda_k(A+B) \leqslant \lambda_{k+r}(A) \leqslant \lambda_{k+2r}(A+B)$ for $1 \leqslant k \leqslant n-2r$ and $\lambda_k(A) \leqslant \lambda_{k+r}(A+B) \leqslant \lambda_{k+2r}(A)$.
Solution: If B is a Hermitian matrix of rank no larger than r, then $B = U\mathsf{diag}(\beta_1, \ldots, \beta_r, 0, \ldots, 0)U^{\mathsf{H}}$, where $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n)$ is a unitary matrix. This amounts to $B = \beta_1\mathbf{u}_1\mathbf{u}_1' + \cdots + \beta_r\mathbf{u}_r\mathbf{u}_r'$. Conversely, every Hermitian matrix of rank no larger than r can be written in this form.
Let W range over the subsets of $\mathbb{R}^n$ that consist of $n-k-2r$ vectors. We have
$$\begin{aligned}
&\lambda_{k+2r}(A+B)\\
&= \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+B)\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W\rangle^{\perp}\}\\
&\geqslant \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}(A+B)\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1, \mathbf{x} \in \langle W \cup \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}\rangle^{\perp}\}\\
&= \min_W \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1, \mathbf{x} \in \langle W \cup \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}\rangle^{\perp}\}\\
&\quad \text{(because } \mathbf{x}^{\mathsf{H}}\mathbf{u}_i = \mathbf{u}_i^{\mathsf{H}}\mathbf{x} = 0)\\
&\geqslant \min_{W_1} \max_{\mathbf{x}} \{\mathbf{x}^{\mathsf{H}}A\mathbf{x} \mid \parallel \mathbf{x} \parallel_2 = 1 \text{ and } \mathbf{x} \in \langle W_1\rangle^{\perp}\} = \lambda_{k+r}(A),
\end{aligned}$$

where W_1 ranges over the subsets of $\mathbb{R}^n$ that contain $n-k-r$ vectors.

The proof of the remaining inequalities follows the same pattern as above and generalize the results and proofs of Supplement 32.

34. Let A be a Hermitian matrix in $\mathbb{C}^{n\times n}$ such that $A = UDU^{\mathsf{H}}$, where $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_n)$ is an unitary matrix and $D = (\lambda_1(A), \ldots, \lambda_n(A))$. If $A_j = \sum_{i=j+1}^{n} \lambda_i(A)\mathbf{u}_i\mathbf{u}_i'$ for $0 \leqslant j \leqslant n-1$, prove that the largest eigenvalue of the matrix $A - A_{n-k}$ is $\lambda_{n-k}(A)$.

    Solution: Since $A - A_{n-k} = \sum_{i=1}^{n-k} \lambda_i(A)\mathbf{u}_i\mathbf{u}_i'$ the statement follows immediately.

35. Let A and B be two Hermitian matrices in $\mathbb{C}^{n\times n}$. Using the same notations as in Supplement 30 prove that for any i, j such that $1 \leqslant i, j \leqslant n$ and $i+j \geqslant n+1$ we have
$$\lambda_{j+k-n}(A+B) \leqslant \lambda_j(A) + \lambda_k(B).$$
    Also, if $i+j \leqslant n+1$, then
$$\lambda_j(A) + \lambda_k(B) \leqslant \lambda_{j+k-1}(A+B).$$

The *field of values of a matrix* $A \in \mathbb{C}^{n\times n}$ is the set of numbers $F(A) = \{\mathbf{x}A\mathbf{x}^{\mathsf{H}} \mid \mathbf{x} \in \mathbb{C}^n \text{ and } \parallel \mathbf{x} \parallel_2 = 1\}$.

36. Prove that $\mathsf{spec}(A) \subseteq F(A)$ for any $A \in \mathbb{C}^{n\times n}$.
37. If $U \in \mathbb{C}^{n\times n}$ is a unitary matrix and $A \in \mathbb{C}^{n\times n}$, prove that $F(UAU^{\mathsf{H}}) = F(A)$.
38. Prove that $A \sim B$ implies $f(A) \sim f(B)$ for every $A, B \in \mathbb{C}^{n\times n}$ and every polynomial f.
39. Let $A \in \mathbb{C}^{n\times n}$ be a matrix such that $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$. Prove that $\sum_{i=1}^n |\lambda_i|^2 \leqslant \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2$; furthermore, prove that A is normal if and only if $\sum_{i=1}^n |\lambda_i|^2 = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2$.
40. Let $A \in \mathbb{C}^{n\times n}$ such that $A \geqslant O_{n,n}$. Prove that if $\mathbf{1}_n$ is an eigenvector of A, then $\rho(A) = \|A\|_\infty$ and if $\mathbf{1}_n$ is an eigenvector of A', then $\rho(A) = \|A\|_1$.

    Solution: If $\mathbf{1}_n$ is an eigenvector of A, then $A\mathbf{1}_n = \lambda\mathbf{1}_n$, so $\sum_{j=1}^n a_{ij} = \lambda$ for every i, $1 \leqslant i \leqslant n$. This means that all rows of A have the same sum λ and, therefore, $\lambda = \|A\|_\infty$, as we saw in Example 6.84. This implies $\rho(A) = \|A\|_\infty$.

    The argument for the second part is similar.
41. Prove that the matrix $A \in \mathbb{C}^{n\times n}$ is normal if and only if there exists a polynomial p such that $p(A) = A^{\mathsf{H}}$.
42. Prove that the matrix $A \in \mathbb{C}^{n\times n}$ is normal if and only if there exist $B, C \in \mathbb{C}^{n\times n}$ such that $A = B + iC$ and $BC = CB$.
43. Let $A \in \mathbb{C}^{n\times n}$ be a matrix and let $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$. Prove that
    a) $\sum_{p=1}^n |\lambda_p|^2 \leqslant \parallel A \parallel_F^2$;
    b) the equality $\sum_{p=1}^n |\lambda_p|^2 = \parallel A \parallel_F^2$ holds if and only if A is normal.

    Solution: By Schur's Triangularization Theorem there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A = U^{\mathsf{H}}TU$ and the diagonal elements of T are the eigenvalues of A. Thus,
$$\parallel A \parallel_F^2 = \parallel T \parallel_F^2 = \sum_{p=1}^n |\lambda_p|^2 + \sum_{i<j} |t_{ij}|^2,$$
    which implies the desired inequality.

The equality of the second part follows from the Spectral Theorem for Normal Matrices. The converse implication can be obtained noting that by the first part, the equality of the second part implies $t_{ij} = 0$ for $i < j$, which means that T is actually a diagonal matrix.

44. Let A and B be two normal matrices in $\mathbb{C}^{n\times n}$. Prove that if AB is a normal matrix, then so is BA.
Solution: By Supplement 21 the matrices AB and BA have the same non-zero eigenvalues. Since A and B are normal we have $A^{\mathsf{H}}A = AA^{\mathsf{H}}$ and $B^{\mathsf{H}}B = BB^{\mathsf{H}}$. Thus, we can write
$$\begin{aligned}\| AB \|_F^2 = \mathit{trace}((AB)^{\mathsf{H}}AB) &= \mathit{trace}(B^{\mathsf{H}}A^{\mathsf{H}}AB) = \mathit{trace}(B^{\mathsf{H}}AA^{\mathsf{H}}B)\\ &\quad\text{(because } A \text{ is a normal matrix)}\\ &= \mathit{trace}((B^{\mathsf{H}}A)(A^{\mathsf{H}}B)) = \mathit{trace}((A^{\mathsf{H}}B)(B^{\mathsf{H}}A))\\ &\quad\text{(by the third part of Theorem 5.51)}\\ &= \mathit{trace}(A^{\mathsf{H}}(BB^{\mathsf{H}})A)) = \mathit{trace}(A^{\mathsf{H}}(B^{\mathsf{H}}B)A))\\ &\quad\text{(because } B \text{ is a normal matrix)}\\ &= \mathit{trace}((BA)^{\mathsf{H}}BA) = \| BA \|_F^2 .\end{aligned}$$
Since AB is a normal matrix, if $\mathsf{spec}(AB) = \{\lambda_1, \ldots, \lambda_p\}$, we have $\sum_{p=1}^{n} |\lambda_p|^2 = \| AB \|_F^2 = \| BA \|_F^2$.
Taking into account the equalities shown above, it follows that BA is a normal matrix by Supplement 43.
45. Let $A \in \mathbb{C}^{n\times n}$ be a matrix with $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$. Prove that A is normal if and only if its singular values are $|\lambda_1|, \ldots, |\lambda_n|$.
46. Let A be a non-negative matrix in $\mathbb{C}^{n\times n}$ and let $\mathbf{u} = A\mathbf{1}_n$ and $\mathbf{v} = A'\mathbf{1}_n$. Prove that
$$\max\{\min u_i, \min v_j\} \leqslant \rho(A) \leqslant \min\{\max u_i, \max v_j\}.$$
Solution: Note that $|\!|\!| A |\!|\!|_\infty = \max u_i$ and $|\!|\!| A |\!|\!|_1 = \max v_j$. By Theorem 7.65, we have $\rho(A) \leqslant \min\{\max u_i, \max v_j\}$.
Let $a = \min u_i$. If $a > 0$ define the non-negative matrix $B \in \mathbb{R}^{n\times n}$ as $b_{ij} = \frac{a a_{ij}}{u_i}$. We have $A \geqslant B \geqslant O_{n,n}$. By Corollary 7.67 we have $\rho(A) \geqslant \rho(B) = a$; the same equality, $\rho(A) \geqslant a$ holds trivially when $a = 0$. In a similar manner, one could prove that $\min v_j \leqslant \rho(A)$, so $\max\{\min u_i, \min v_j\} \leqslant \rho(A)$.
47. Let A be a non-negative matrix in $\mathbb{C}^{n\times n}$ and let $\mathbf{x} \in \mathbb{R}^n$ be a vector such that $\mathbf{x} > \mathbf{0}_n$. Prove that
a) $\min\left\{\frac{(A\mathbf{x})_j}{x_j} \mid 1 \leqslant j \leqslant n\right\} \leqslant \rho(A) \leqslant \max\left\{\frac{(A\mathbf{x})_j}{x_j} \mid 1 \leqslant j \leqslant n\right\}$;
b) $\min\left\{x_j \sum_{i=1}^{n} \frac{a_{ij}}{x_j} \mid 1 \leqslant j \leqslant n\right\} \leqslant \rho(A) \leqslant \max\left\{x_j \sum_{i=1}^{n} \frac{a_{ij}}{x_j} \mid 1 \leqslant j \leqslant n\right\}$.
Solution: Define the diagonal matrix $S = \mathsf{diag}(x_1, \ldots, x_n)$. Its inverse is $S^{-1} = \mathsf{diag}(\frac{1}{x_1}, \ldots, \frac{1}{x_n})$. The matrix $B = S^{-1}AS$ is non-negative and we have $b_{ij} = \frac{x_i a_{ij}}{x_j}$. This implies
$$\mathbf{u} = B\mathbf{1}_n = \begin{pmatrix} x_1 \sum_{j=1}^{n} \frac{a_{1j}}{x_j} \\ \vdots \\ x_n \sum_{j=1}^{n} \frac{a_{nj}}{x_j} \end{pmatrix} \text{ and } \mathbf{v} = B'\mathbf{1}_n = \begin{pmatrix} \frac{\sum_{i=1}^{n} x_i a_{1i}}{x_1} \\ \vdots \\ \frac{\sum_{i=1}^{n} x_i a_{ni}}{x_n} \end{pmatrix} = \begin{pmatrix} \frac{(A\mathbf{x})_1}{x_1} \\ \vdots \\ \frac{(A\mathbf{x})_n}{x_n} \end{pmatrix}.$$
Thus, by applying the inequalities of Supplement 46 we obtain the desired inequalities.

48. Let $A \in \mathbb{R}^{n\times n}$ and $\mathbf{x} \in \mathbb{R}^n$ such that $A \geqslant O_{n,n}$ and $\mathbf{x} > 0$. Prove that if $a, b \in \mathbb{R}_{\geqslant 0}$ are such that $a\mathbf{x} \leqslant A\mathbf{x} \leqslant b\mathbf{x}$, then $a \leqslant \rho(A) \leqslant b$.

    Solution: Since $a\mathbf{x} \leqslant A\mathbf{x}$ we have $a \leqslant \min_{1\leqslant i\leqslant n} \frac{(A\mathbf{x})_i}{x_i}$, so $a \leqslant \rho(A)$ by Supplement 47. Similarly, we hace $\max_{1\leqslant i\leqslant n} \frac{(A\mathbf{x})_i}{x_i} \leqslant b$, so $\rho(A) \leqslant b$.
49. Let $A \in \mathbb{C}^{n\times n}$ be a matrix such that $A \geqslant O_{n,n}$. Prove that if there exists $k \in \mathbb{N}$ such that $A^k > O_{n,n}$, then $\rho(A) > 0$.
50. Let $A \in \mathbb{C}^{n\times n}$ be a matrix such that $A \geqslant O_{n,n}$. If $A \neq O_{n,n}$ and there exists an eigenvector $\mathbf{x}$ of A such that $\mathbf{x} > \mathbf{0}_n$, prove that $\rho(A) > 0$.
51. Prove that $A \in \mathbb{C}^{n\times n}$ is positive semidefinite if and only if there is a set $U = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq \mathbb{C}^n$ such that $A = \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i^{\mathsf{H}}$. Furthermore, prove that A is positive definite if and only if there exists a linearly independent set U as above.
52. Prove that A is positive definite if and only if A^{-1} is positive definite.
53. Prove that if $A \in \mathbb{C}^{n\times n}$ is a positive semidefinite matrix, then A^k is positive semidefinite for every $k \geqslant 1$.
54. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let $p_A(\lambda) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_m\lambda^{n-m}$ be its characteristic polynomial, where $c_m \neq 0$. Then, A is positive semidefinite if and only if $c_i \neq 0$ for $0 \leqslant k \leqslant m$ (where $c_0 = 1$) and $c_jc_{j+1} < 0$ for $0 \leqslant j \leqslant m-1$.
55. Let $A \in \mathbb{C}^{n\times n}$ be a positive semidefinite matrix. Prove that for every $k \geqslant 1$ there exists a positive semidefinite matrix B having the same rank as A such that
    a) $B^k = A$;
    b) $AB = BA$;
    c) B can be expressed as a polynomial in A.

    Solution: Since A is Hermitian, its eigenvalues are real nonnegative numbers and, by the Spectral Theorem for Hermitian matrices, there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that $A = U^{\mathsf{H}}\mathsf{diag}(\lambda_1, \ldots, \lambda_n)U$. Let $B = U^{\mathsf{H}}\mathsf{diag}(\lambda_1^{\frac{1}{k}}, \ldots, \lambda_n^{\frac{1}{k}})U$, where $\lambda_i^{\frac{1}{k}}$ is a non-negative root of order k of λ_i. Thus, $B^k = A$, B is clearly positive semidefinite, $\mathit{rank}(B) = \mathit{rank}(A)$, and $AB = BA$.

    Let
    $$p(x) = \sum_{j=1}^n \lambda_j^{\frac{1}{k}} \prod_{k=1, k\neq j}^n \frac{x - \lambda_k}{\lambda_j - \lambda_k}$$
    be a Lagrange interpolation polynomial such that $p(\lambda_j) = \lambda_j^{\frac{1}{k}}$ (see Exercise 67). Then,
    $$p(\mathsf{diag}(\lambda_1, \ldots, \lambda_n)) = \mathsf{diag}(\lambda_1^{\frac{1}{k}}, \ldots, \lambda_n^{\frac{1}{k}}),$$
    so
    $$\begin{aligned} p(A) &= p(U^{\mathsf{H}}\mathsf{diag}(\lambda_1, \ldots, \lambda_n)U) = U^{\mathsf{H}}p(\mathsf{diag}(\lambda_1, \ldots, \lambda_n))U \\ &= U^{\mathsf{H}}\mathsf{diag}(\lambda_1^{\frac{1}{k}}, \ldots, \lambda_n^{\frac{1}{k}})U = B. \end{aligned}$$
56. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Prove that there exists $b \in \mathbb{R}$ such that $A + b(\mathbf{1}\mathbf{1}' - I_n)$ is positive semi-definite, where $\mathbf{1} \in \mathbb{R}^n$.

    Solution: We need to find b such that for every $\mathbf{x} \in \mathbb{R}^n$ we will have
    $$\mathbf{x}'(A + b(\mathbf{1}\mathbf{1}' - I_n))\mathbf{x} \geqslant 0.$$

We have $\mathbf{x}'(A + b(\mathbf{1}\mathbf{1}' - I_n))\mathbf{x} = \mathbf{x}'A\mathbf{x} + b\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{x} - b\mathbf{x}'\mathbf{x} \geqslant 0$, which amounts to

$$\mathbf{x}'A\mathbf{x} + b\left(\left(\sum_{i=1}^{n} x_i\right)^2 - \parallel \mathbf{x} \parallel_2^2\right) \geqslant 0.$$

Since A is symmetric, by Rayleigh-Ritz Theorem, we have $\mathbf{x}'A\mathbf{x} \geqslant \lambda_1 \parallel \mathbf{x} \parallel_2^2$, where λ_1 is the least eigenvalue of A. Therefore, it suffices to take $b \leqslant \lambda_1$ to satisfy the equality for every $\mathbf{x}$.

57. If $A \in \mathbb{R}^{n\times n}$ is a positive definite matrix prove that there exist $c, d > 0$ such that $c \parallel \mathbf{x} \parallel_2^2 \leqslant \mathbf{x}'A\mathbf{x} \leqslant d \parallel \mathbf{x} \parallel_2^2$, for every $\mathbf{x} \in \mathbb{R}^n$.
58. Let $A = \mathsf{diag}(A_1, \ldots, A_p)$ and $B = (B_1, \ldots, B_q)$ be two block-diagonal matrices. Prove that $sep_F(A, B) = \min\{sep_F(A_i, B_j) \mid 1 \leqslant i \leqslant p \text{ and } 1 \leqslant j \leqslant q\}$.
59. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Prove that if for any $\lambda \in \mathsf{spec}(A)$ we have $\lambda > -a$, then the matrix $A + aI$ is positive-semidefinite.
60. Let $A \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times n}$ be two matrices that have the eigenvalues $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$, respectively. Prove that:
    a) if A and B are positive definite, then so is $A \otimes B$;
    b) if $m = n$ and A, B are symmetric positive definite, the Hadamard product $A \odot B$ is positive definite.

    Solution: For the second part recall that the Hadamard product $A \odot B$ of two square matrices of the same format is a principal submatrix of $A \otimes B$. Then, apply Theorem 7.53.
61. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Prove that if A is positive semidefinite, then all its eigenvalues are non-negative; if A is positive definite then its eigenvalues are positive.

    Solution: Since A is Hermitian, all its eigenvalues are real numbers. Suppose that A is positive semidefinite, that is, $\mathbf{x}^{\mathsf{H}}A\mathbf{x} \geqslant 0$ for $\mathbf{x} \in \mathbb{C}^n$. If $\lambda \in \mathsf{spec}(A)$, then $A\mathbf{v} = \lambda\mathbf{v}$ for some eigenvector $\mathbf{v} \neq \mathbf{0}$. The positive semi-definiteness of A implies $\mathbf{v}^{\mathsf{H}}A\mathbf{v} = \lambda\mathbf{v}^{\mathsf{H}}\mathbf{v} = \lambda \parallel \mathbf{v} \parallel_2^2 \geqslant 0$, which implies $\lambda \geqslant 0$. It is easy to see that if A is positive definite, then $\lambda > 0$.
62. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Prove that if A is positive semidefinite, then all its principal minors are non-negative real numbers; if A is positive definite then all its principal minors are positive real numbers.

    Solution: Since A is positive semidefinite, every sub-matrix $A\begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix}$ is a Hermitian positive semidefinite matrix by Theorem 6.110, so every principal minor is a non-negative real number. The second part is proven similarly.
63. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Prove that he following statements are equivalent:
    a) A is positive semidefinite;
    b) all eigenvalues of A are non-negative numbers;
    c) there exists a Hermitian matrix $C \in \mathbb{C}^{n\times n}$ such that $C^2 = A$;
    d) A is the Gram matrix of a sequence of vectors, that is, $A = B^{\mathsf{H}}B$ for some $B \in \mathbb{C}^{n\times n}$.

    Solution: (a) implies (b): This is stated in Exercise 61.

    (b) implies (c): Suppose that A is a matrix such that all its eigenvalues are the non-negative numbers $\lambda_1, \ldots, \lambda_n$. By Theorem 7.41, A can be written as $A = U^{\mathsf{H}}DU$, where U is a unitary matrix and

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Define the matrix $\sqrt{D}$ as

$$\sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}.$$

Clearly, we gave $(\sqrt{D})^2 = D$. Now we can write $A = U\sqrt{D}U^{\mathsf{H}}U\sqrt{D}U^{\mathsf{H}}$, which allows us to define the desired matrix C as $C = U\sqrt{D}U^{\mathsf{H}}$.

(c) implies (d): Since C is itself a Hermitian matrix, this implication is obvious.

(d) implies (a): Suppose that $A = B^{\mathsf{H}}B$ for some matrix $B \in \mathbb{C}^{n\times k}$. Then, for $\mathbf{x} \in \mathbb{C}^n$ we have $\mathbf{x}^{\mathsf{H}}A\mathbf{x} = \mathbf{x}^{\mathsf{H}}B^{\mathsf{H}}B\mathbf{x} = (B\mathbf{x})^{\mathsf{H}}(B\mathbf{x}) = \| B\mathbf{x} \|_2^2 \geqslant 0$, so A is positive semidefinite.

64. Let $A \in \mathbb{R}^{n\times n}$ be a real matrix that is symmetric and positive semidefinite such that $A\mathbf{1}_n = \mathbf{0}_n$. Prove that $2\max_{1\leqslant i\leqslant n}\sqrt{a_{ii}} \leqslant \sum_{j=1}^n \sqrt{a_{jj}}$.

    Solution: By Supplement 63(d), A is the Gram matrix of a sequence of vectors $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$, so $A = B'B$. Since $A\mathbf{1}_n = \mathbf{0}_n$, it follows that $(B\mathbf{1}_n)'(B\mathbf{1}_n) = 0$, so $B\mathbf{1}_n = \mathbf{0}_n$. Thus, $\sum_{i=1}^n \mathbf{b}_i = \mathbf{0}_n$. Then, we have $\| \mathbf{b}_i \|_2 = \| -\sum_{j\neq i}\mathbf{b}_j \|_2 \leqslant \sum_{j\neq i} \| \mathbf{b}_j \|_2$, which implies $2\max_{1\leqslant i\leqslant n} \| \mathbf{b}_i \|_2 \leqslant \sum_{j=1}^n \| \mathbf{b}_j \|_2$. This, is equivalent to the inequality to be shown.

65. Let $A \in \mathbb{R}^{n\times n}$ be a matrix and let $(\lambda, \mathbf{x})$ be an eigenpair of A. Prove that
    a) $2\Im(\lambda) = \mathbf{x}^{\mathsf{H}}(A - A')\mathbf{x}$;
    b) if $\alpha = \frac{1}{2}\max\{|a_{ij} - a_{ji}| \mid 1 \leqslant i, j \leqslant n\}$, then
    $2|\Im(\lambda)| \leqslant \alpha \sum \left\{ |\bar{x}_i x_j - x_i \bar{x}_j| \,\middle|\, 1 \leqslant i, j \leqslant n, i \neq j \right\}$;
    c) $|\Im(\lambda)| \leqslant \alpha\sqrt{\frac{n(n-1)}{2}}$.

    The inequality of Part (c) is known as *Bendixon Inequality*.

    Hint: apply the results of Exercise 5 of Chapter 6.

66. Let $A \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times n}$ be two matrices that have the eigenvalues $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_n$, respectively. Prove that the Kronecker product $A \otimes B$ has the eigenvalues $\lambda_1\mu_1, \ldots, \lambda_1\mu_n, \ldots, \lambda_m\mu_1, \ldots, \lambda_m\mu_n$.

    Solution: Suppose that $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ and $B\mathbf{u}_j = \mu_j\mathbf{u}_j$. Then, $(A \otimes B)(\mathbf{v}_i \otimes \mathbf{u}_j) = (A\mathbf{v}_i) \otimes (B\mathbf{u}_j) = \lambda_i\mu_j(\mathbf{v}_i \times \mathbf{u}_j)$.

67. Let $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{m\times m}$. Prove that $trace(A \otimes B) = trace(A)trace(B) = trace(B \otimes A)$ and $\det(A \otimes B) = (\det(A))^m(\det(B))^n = \det(B \otimes A)$.

68. Let $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{m\times m}$ be two matrices. If $\mathsf{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $\mathsf{spec}(B) = \{\mu_1, \ldots, \mu_m\}$, prove that $\mathsf{spec}(A \oplus B) = \{\lambda_i + \mu_j \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$ and $\mathsf{spec}(A \ominus B) = \{\lambda_i - \mu_j \mid 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$.

    Solution: Let $\mathbf{x}$ and $\mathbf{y}$ be two eigenvectors of A and B that correspond to the eigenvalues λ and μ, respectively. Since $A \oplus B = (A \otimes I_m) + (I_n \otimes B)$ we have

$$\begin{aligned}(A \oplus B)(\mathbf{x} \otimes \mathbf{y}) &= (A \otimes I_m)(\mathbf{x} \otimes \mathbf{y}) + (I_n \otimes B)(\mathbf{x} \otimes \mathbf{y}) \\ &= (A\mathbf{x} \otimes \mathbf{y}) + (\mathbf{x} \otimes B\mathbf{y}) \\ &= \lambda(\mathbf{x} \otimes \mathbf{y}) + \mu(\mathbf{x} \otimes \mathbf{y}) \\ &= (\lambda + \mu)(\mathbf{x} \otimes \mathbf{y}).\end{aligned}$$

By replacing B by $-B$ we obtain the spectrum of $A \ominus B$.

69. Let $A \in \mathbb{C}^{n \times n}$ be matrix and let $r_i = \sum\{a_{ij} \mid 1 \leqslant j \leqslant n \text{ and } j \neq i\}$, for $1 \leqslant i \leqslant n$. Prove that $\mathsf{spec}(A) \subseteq \bigcup_{i=1}^{n}\{z \in \mathbb{C} \mid |z - a_{ii}| \leqslant r_i\}$.
A disk of the form $D_i(A) = \{z \in \mathbb{C} \mid |z - a_{ii}| \leqslant r_i\}$ is called a *Gershgorin disk.*
Solution: Let $\lambda \in \mathsf{spec}(A)$ and let suppose that $A\mathbf{x} = \lambda\mathbf{x}$, where $\mathbf{x} \neq \mathbf{0}$. Let p be such that $|x_p| = \max\{|x_i| \mid 1 \leqslant i \leqslant n\}$. Then, $\sum_{j=1}^{n} a_{pj}x_j = \lambda x_p$, which is the same as $\sum_{j=1,j\neq p}^{n} a_{pj}x_j = (\lambda - a_{pp})x_p$. This, in turn, implies
$$\begin{aligned}|x_p||\lambda - a_{pp}| &= \left|\sum_{j=1,j\neq p}^{n} a_{pj}x_j\right| \leqslant \sum_{j=1,j\neq p}^{n} |a_{pj}||x_j| \\ &\leqslant |x_p| \sum_{j=1,j\neq p}^{n} |a_{pj}| = |x_p|r_p.\end{aligned}$$
Therefore, $|\lambda - a_{pp}| \leqslant r_p$ for some p.

70. If $A, B \in \mathbb{R}^{m \times m}$ is a symmetric matrices and $A \sim B$, prove that $\mathcal{I}(A) = \mathcal{I}(B)$.

71. If A is a symmetric block diagonal matrix, $A = \mathsf{diag}(A_1, \ldots, A_k)$, then $\mathcal{I}(A) = \sum_{i=1}^{k} \mathcal{I}(A_i)$.

72. Let $A = \begin{pmatrix} B & \mathbf{c} \\ \mathbf{c}' & b \end{pmatrix} \in \mathbb{R}^{m \times m}$ be a symmetric matrix, where $B \in \mathbb{R}^{(m-1)\times(m-1)}$ such that there exists $\mathbf{u} \in \mathbb{R}^{m-1}$ for which $B\mathbf{u} = \mathbf{0}_{m-1}$ and $\mathbf{c}'\mathbf{u} \neq 0$. Prove that $\mathcal{I}(A) = \mathcal{I}(B) + (1, 1, -1)$.
Solution: It is clear that $\mathbf{u} \neq \mathbf{0}_{m-1}$ and we may assume that $u_1 \neq 0$. We can write
$$\mathbf{u} = \begin{pmatrix} u_1 \\ \mathbf{v} \end{pmatrix}, B = \begin{pmatrix} b_{11} & \mathbf{d}' \\ \mathbf{d} & D \end{pmatrix}, \text{ and } \mathbf{c} = \begin{pmatrix} c_1 \\ \mathbf{e} \end{pmatrix},$$
where $\mathbf{v} \in \mathbb{R}^{m-2}$, $D \in \mathbb{R}^{(m-2)\times(m-2)}$, and $\mathbf{d}, \mathbf{e} \in \mathbb{R}^{m-2}$. Define $k = \mathbf{c}'\mathbf{u} = c_1u_1 + \mathbf{e}'\mathbf{v} \neq 0$ and
$$P = \begin{pmatrix} u_1 & \mathbf{v}' & 0 \\ \mathbf{0}_{m-2} & I_{m-2} & \mathbf{0}_{m-2} \\ 0 & \mathbf{0}'_{m-2} & 1 \end{pmatrix}.$$
The equality $B\mathbf{u} = \mathbf{0}_{m-1}$ can be written as $b_{11}u_1 + \mathbf{d}'\mathbf{v} = 0$ and $\mathbf{d}u_1 + D\mathbf{v} = \mathbf{0}_{m-2}$. With these notations, A can be written as
$$A = \begin{pmatrix} b_{11} & \mathbf{d}' & c_1 \\ \mathbf{d} & D & \mathbf{e} \\ \mathbf{c}_1 & \mathbf{e}' & b \end{pmatrix}$$
and we have
$$PAP' = \begin{pmatrix} 0 & \mathbf{0}_{m-2} & k \\ \mathbf{0}'_{m-2} & D & \mathbf{e} \\ k & \mathbf{e}' & b \end{pmatrix}$$
For

$$Q = \begin{pmatrix} 1 & \mathbf{0}' & 0 \\ \frac{1}{k}\mathbf{e} & I_{m-2} & \mathbf{0}_{m-2} \\ -\frac{2}{k}b & \mathbf{0}'_{m-2} & 1 \end{pmatrix}$$

we have

$$(QP)A(QP)' = \begin{pmatrix} 0 & \mathbf{0}'_{m-2} & k \\ \mathbf{0} & D & \mathbf{0}_{m-2} \\ k & \mathbf{0}'_{m-2} & 0 \end{pmatrix}.$$

Let R be the permutation matrix

$$R = \begin{pmatrix} 0 & \mathbf{0}'_{m-2} & 1 \\ 1 & \mathbf{0}'_{m-2} & 0 \\ \mathbf{0}_{m-2} & I_{m-2} & \mathbf{0}_{m-2} \end{pmatrix}.$$

Observe that

$$R(QP)A(QP)'R' = \begin{pmatrix} 0 & k & \mathbf{0}'_{m-2} \\ k & 0 & \mathbf{0}'_{m-2} \\ \mathbf{0}_{m-2} & \mathbf{0}_{m-2} & D \end{pmatrix},$$

which implies that $\mathcal{I}(A) = \mathcal{I}(D) + (1,1,0)$ because the eigenvalues of the matrix $\begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix}$ are k and $-k$.

On the other hand, if

$$S = \begin{pmatrix} u_1 & \mathbf{v} \\ \mathbf{0}_{m-2} & I_{m-2} \end{pmatrix},$$

we have $SDS' = \begin{pmatrix} 0 & \mathbf{0}'_{m-2} \\ \mathbf{0}_{m-2} & D \end{pmatrix}$, which implies $\mathcal{I}(B) = \mathcal{I}(D) + (0,0,1)$. This yields $\mathcal{I}(A) = \mathcal{I}(B) + (1,1,-1)$.

73. Let $A \in \mathbb{C}^{n\times n}$ and let $U \in \mathbb{C}^{n\times n}$ an unitary matrix and T an upper-triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A = U^{\mathsf{H}}TU$ whose existence follows by Schur's Triangularition Theorem. Let $B = \frac{1}{2}\left(A + A^{\mathsf{H}}\right)$ and $C = \frac{1}{2}\left(A - A^{\mathsf{H}}\right)$ be the matrices introduced in Exercise 18 of Chapter 5.
Prove that
$$\sum_{i=1}^{n} \Re(\lambda_i)^2 \leqslant \| B \|_F^2 \text{ and } \sum_{i=1}^{n} \Im(\lambda_i)^2 \leqslant \| C \|_F^2 .$$

Bibliographical Comments

The proof of Theorem 7.56 is given in [193]. Hoffman-Wielandt theorem was shown in [94]. Ky Fan's Theorem appeared in [67].

Supplement 23 is a result of Juhasz which appears in [104]; Supplement 11 originated in [111], and Supplement 72 is a result of Fiedler [71].

As before, the two volumes [95] and [96] are highly recommended for a deep understanding of spectral aspects of linear algebra.