Review of Basic Linear Algebra for MA3291

Wong Yan Loi

Chapter 1

Vector spaces and Linear Transformations

Notations: $\mathbb{R}$ is the set of real numbers, $\mathbb{C}$ is the set of complex numbers, $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$. $\mathcal{M}_{n\times n}(\mathbb{R})$ is the set of $n \times n$ matrices over $\mathbb{R}$, $\mathcal{M}_{n\times n}(\mathbb{C})$ is the set of $n \times n$ matrices over $\mathbb{C}$, $\mathcal{M}_{n\times n}(\mathbb{K})$ is the set of $n \times n$ matrices over $\mathbb{K}$. $\boldsymbol{O}_n$ is the $n \times n$ zero matrix. $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

1.1 Basis and Dimension

Let V be a vector space over $\mathbb{K}$.

Definition 1.1. *A set $\mathcal{S} \subset V$ is said to be a linearly independent set over $\mathbb{K}$ if for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathcal{S}$, $\alpha_1\boldsymbol{x}_1 + \cdots + \alpha_k\boldsymbol{x}_k = \boldsymbol{0}$ implies $\alpha_1 = \cdots = \alpha_k = 0$.*

Definition 1.2. *A set $\mathcal{S} \subset V$ is called a spanning set for V if for any $\boldsymbol{x} \in V$, there exist $\alpha_1, \ldots \alpha_k \in \mathbb{K}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathcal{S}$ such that $\boldsymbol{x} = \alpha_1\boldsymbol{x}_1 + \cdots + \alpha_k\boldsymbol{x}_k$.*

Definition 1.3. *A set $\mathcal{E} \subset V$ is called a basis for V over $\mathbb{K}$ if $\mathcal{E}$ is a linearly independent spanning set.*

Definition 1.4. *A vector space V over $\mathbb{K}$ is called a finite dimensional vector space if there exists a finite basis for V over $\mathbb{K}$.*

Proposition 1.1. *Let V be a finite dimensional vector space over $\mathbb{K}$. Suppose $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ and $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m\}$ are two bases of V. Then $m = n$.*

Proof. For each $j = 1, \ldots, n$, $\boldsymbol{e}_j = \sum_{i=1}^m a_{ij}\boldsymbol{f}_i$. Similarly, for each $j = 1, \ldots, m$, $\boldsymbol{f}_j = \sum_{i=1}^n b_{ij}\boldsymbol{e}_i$. Let $\boldsymbol{E} = (\boldsymbol{e}_1 \ldots, \boldsymbol{e}_n)$ and $\boldsymbol{F} = (\boldsymbol{f}_1 \ldots, \boldsymbol{f}_m)$ be row vectors, and let $\boldsymbol{A} = (a_{ij})_{m\times n}$ and $\boldsymbol{B} = (b_{ij})_{n\times m}$. The above relations can be written as $\boldsymbol{E} = \boldsymbol{F}\boldsymbol{A}$ and $\boldsymbol{F} = \boldsymbol{E}\boldsymbol{B}$. Substituting one into the other, we obtain $\boldsymbol{E} = \boldsymbol{E}\boldsymbol{B}\boldsymbol{A}$ and $\boldsymbol{F} = \boldsymbol{F}\boldsymbol{A}\boldsymbol{B}$, or equivalently $\boldsymbol{E}(\boldsymbol{B}\boldsymbol{A} - \boldsymbol{I}_n) = \boldsymbol{0}_n$ and $\boldsymbol{F}(\boldsymbol{A}\boldsymbol{B} - \boldsymbol{I}_m) = \boldsymbol{0}_m$, where $\boldsymbol{0}_n$ and $\boldsymbol{0}_m$ are row vectors consisting of n and m zero vectors of V respectively. Since $\boldsymbol{E}$ and $\boldsymbol{F}$ are independent sets of vectors, we must have $\boldsymbol{B}\boldsymbol{A} - \boldsymbol{I}_n = \boldsymbol{O}_n$ and $\boldsymbol{A}\boldsymbol{B} - \boldsymbol{I}_m = \boldsymbol{O}_m$. Thus $n = \mathrm{tr}(\boldsymbol{I}_n) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{I}_m) = m$.

Thus the number of elements of any basis of a finite dimensional vector space V is always the same. This number is called the *dimension* of V over $\mathbb{K}$. We use $\dim_{\mathbb{K}}(V)$ to denote the dimension of V over $\mathbb{K}$. When the scalar field is clear, we simply denote it by $\dim(V)$.

1.2 Linear Transformation

Let V and W be vectors spaces over $\mathbb{K}$.

Definition 1.5. *A function $T : V \longrightarrow W$ is called a linear transformation if for any $\alpha, \beta \in \mathbb{K}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$, $T(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha T(\boldsymbol{x}) + \beta T(\boldsymbol{y})$.*

Definition 1.6. *A linear transformation $T : V \longrightarrow W$ is called an isomorphism if T is bijective. In this case, the inverse $T^{-1} : W \longrightarrow V$ is also a linear transformation. If there is an isomorphism between the vector spaces V and W, then V and W are said to be isomorphic. We shall denote two isomorphic vector spaces by the notation $V \cong W$.*

Remark. Clearly, if $V \cong W$, then $\dim(V) = \dim(W)$.

Proposition 1.2. *Let $T : V \longrightarrow W$ be a linear transformation. Let $\mathcal{E} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ and $\mathcal{F} = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m\}$ be bases of V and W respectively. For each $j = 1, \ldots, n$ let $T(\boldsymbol{e}_j) = \sum_{i=1}^{m} a_{ij}\boldsymbol{f}_i$. Let $\boldsymbol{A} = (a_{ij})_{m\times n}$. The matrix $\boldsymbol{A}$ is called the matrix representing T relative to the bases $\mathcal{E}$ and $\mathcal{F}$.*

Let $\boldsymbol{x} \in V$ and $\boldsymbol{y} \in W$ be such that $T(\boldsymbol{x}) = \boldsymbol{y}$. Express $\boldsymbol{x} = \sum_{j=1}^{n} x_j\boldsymbol{e}_j$, and $\boldsymbol{y} = \sum_{i=1}^{m} y_i\boldsymbol{f}_i$. Let $\boldsymbol{X} = (x_1, \ldots, x_n)^\top$ and $\boldsymbol{Y} = (y_1, \ldots, y_m)^\top$ be column vectors. Then $\boldsymbol{AX} = \boldsymbol{Y}$.

Proof. $T(\boldsymbol{x}) = \sum_{j=1}^{n} x_j T(\boldsymbol{e}_j) = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij}\boldsymbol{f}_i\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}x_j\right)\boldsymbol{f}_i$. Since $T(\boldsymbol{x}) = \boldsymbol{y}$, we have $\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}x_j\right)\boldsymbol{f}_i = \sum_{i=1}^{m} y_i\boldsymbol{f}_i$. Equating coefficients, we obtain $\sum_{j=1}^{n} a_{ij}x_j = y_i$ for $i = 1, \ldots, m$. Therefore, $\boldsymbol{AX} = \boldsymbol{Y}$.

Example 1.1. Let $\boldsymbol{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \boldsymbol{E}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \boldsymbol{E}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \boldsymbol{E}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Then $\mathcal{E} = \{\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \boldsymbol{E}_4\}$ is a basis of $\mathcal{M}_{2\times 2}(\mathbb{R})$. Fix $\boldsymbol{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $\mathcal{M}_{2\times 2}(\mathbb{R})$. Define $R_{\boldsymbol{B}} : \mathcal{M}_{2\times 2}(\mathbb{R}) \longrightarrow \mathcal{M}_{2\times 2}(\mathbb{R})$ by $R_{\boldsymbol{B}}(\boldsymbol{X}) = \boldsymbol{XB}$. The matrix representing $R_{\boldsymbol{B}}$ relative to the basis $\mathcal{E}$ is $\begin{pmatrix} \boldsymbol{B}^\top & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}^\top \end{pmatrix} = \boldsymbol{I} \otimes \boldsymbol{B}^\top$. *Similarly, if we define $L_{\boldsymbol{A}}(\boldsymbol{X}) = \boldsymbol{AX}$, then the matrix representing $L_{\boldsymbol{A}}$ relative to the basis $\mathcal{E}$ is $\boldsymbol{A} \otimes \boldsymbol{I}$.*

By the associativity of matrix multiplication, $L_{\boldsymbol{A}} \circ R_{\boldsymbol{B}} = R_{\boldsymbol{B}} \circ L_{\boldsymbol{A}}$, and the matrix representing this linear transformation relative to $\mathcal{E}$ is $(\boldsymbol{A} \otimes \boldsymbol{I})(\boldsymbol{I} \otimes \boldsymbol{B}^\top) = (\boldsymbol{I} \otimes \boldsymbol{B}^\top)(\boldsymbol{A} \otimes \boldsymbol{I}) = \boldsymbol{A} \otimes \boldsymbol{B}^\top$.

Let V be an n-dimensional vector space over $\mathbb{K}$. Let $\mathcal{E}$ and $\mathcal{F}$ be bases of V over $\mathbb{K}$. For $j = 1, \ldots, n$, let $\boldsymbol{f}_j \in \mathcal{F}$. We may write $\boldsymbol{f}_j = \sum_{i=1}^{n} p_{ij}\boldsymbol{e}_i$. Let $\boldsymbol{P} = (p_{ij})$. Then $\boldsymbol{P}$ is invertible and is the matrix of change of basis from $\mathcal{F}$ to $\mathcal{E}$. That is $\boldsymbol{X}_{\mathcal{E}} = \boldsymbol{P}\boldsymbol{X}_{\mathcal{F}}$.

Proposition 1.3. *Let $T : V \longrightarrow V$ be a linear transformation. Let the matrices representing T with respect to the bases $\mathcal{E}$ and $\mathcal{F}$ be $\boldsymbol{A}_{\mathcal{E}}$ and $\boldsymbol{A}_{\mathcal{F}}$ respectively. Then $\boldsymbol{A}_{\mathcal{F}} = \boldsymbol{P}^{-1}\boldsymbol{A}_{\mathcal{E}}\boldsymbol{P}$.*

Proof. Let $T(\boldsymbol{x}) = \boldsymbol{y}$ so that $\boldsymbol{A}_{\mathcal{E}}\boldsymbol{X}_{\mathcal{E}} = \boldsymbol{Y}_{\mathcal{E}}$ and $\boldsymbol{A}_{\mathcal{F}}\boldsymbol{X}_{\mathcal{F}} = \boldsymbol{Y}_{\mathcal{F}}$. Then $\boldsymbol{X}_{\mathcal{E}} = \boldsymbol{P}\boldsymbol{X}_{\mathcal{F}}$ and $\boldsymbol{Y}_{\mathcal{E}} = \boldsymbol{P}\boldsymbol{Y}_{\mathcal{F}}$. Therefore,

$$\boldsymbol{P}\boldsymbol{A}_{\mathcal{F}}\boldsymbol{X}_{\mathcal{F}} = \boldsymbol{P}\boldsymbol{Y}_{\mathcal{F}} = \boldsymbol{Y}_{\mathcal{E}} = \boldsymbol{A}_{\mathcal{E}}\boldsymbol{X}_{\mathcal{E}} = \boldsymbol{A}_{\mathcal{E}}\boldsymbol{P}\boldsymbol{X}_{\mathcal{F}}$$

Thus $\boldsymbol{P}\boldsymbol{A}_{\mathcal{F}} = \boldsymbol{A}_{\mathcal{E}}\boldsymbol{P}$. That is $\boldsymbol{A}_{\mathcal{F}} = \boldsymbol{P}^{-1}\boldsymbol{A}_{\mathcal{E}}\boldsymbol{P}$.

Definition 1.7. *Two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{K})$ are said to be similar over $\mathbb{K}$ if there exists an invertible matrix $\boldsymbol{P} \in \mathcal{M}_{n\times n}(\mathbb{K})$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{B}$.*

Remark. Any two matrices representing a linear transformation $T : V \longrightarrow V$ are similar over $\mathbb{K}$.

Definition 1.8. *Let $T : V \longrightarrow W$ be a linear transformation. The null space of T or the kernel of T, denoted by $\ker T$, is the subspace $\{\boldsymbol{x} \in V \mid T(\boldsymbol{x}) = \boldsymbol{0}\}$. The image of T, denoted by $\operatorname{im} T$ is subspace $\{T(\boldsymbol{x}) |\ \boldsymbol{x} \in V\}$ in W. The nullity of T, denoted by $\operatorname{nullity}(T)$, is the dimension of the null space of T.*

Theorem 1 (rank-nullity theorem)**.** *Let $\boldsymbol{A} : \mathbb{K}^n \longrightarrow \mathbb{K}^m$ be a linear transformation, where $\boldsymbol{A}$ is an $m \times n$ matrix over $\mathbb{K}$. Let $\boldsymbol{E}$ be the reduced row-echelon form of $\boldsymbol{A}$. Then*
(i) $\operatorname{nullity}(\boldsymbol{A}) + \dim(\operatorname{im} \boldsymbol{A}) = n$.
(ii) $\dim(\operatorname{im} \boldsymbol{A}) =$ *number of nonzero rows of* $\boldsymbol{E}$.

Proof. (i) Each row operation applied to $\boldsymbol{A}$ corresponds to a change of basis in the codomain $\mathbb{K}^m$ so that the linear transformation is represented by the resulting matrix. Hence $\operatorname{im} \boldsymbol{A} \cong \operatorname{im} \boldsymbol{E} =$ column space of $\boldsymbol{E}$. Thus $\dim(\operatorname{im} \boldsymbol{A}) =$ number of pivotal columns of $\boldsymbol{E}$. Note that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ if and only if $\boldsymbol{E}\boldsymbol{x} = \boldsymbol{0}$. Hence the null space of $\boldsymbol{A}$ = the null space of $\boldsymbol{E}$. Thus $\operatorname{nullity}(\boldsymbol{A}) = \operatorname{nullity}(\boldsymbol{E}) =$ number of non-pivotal columns of $\boldsymbol{E}$. Therefore, $\operatorname{nullity}(\boldsymbol{A}) + \dim(\operatorname{im} \boldsymbol{A}) = n$.
(ii) $\dim(\operatorname{im} \boldsymbol{A}) =$ number of pivotal columns of $\boldsymbol{E} =$ number of nonzero rows of $\boldsymbol{E}$.

Remark. Alternatively, $\operatorname{im} \boldsymbol{A} \cong \operatorname{im} \boldsymbol{E}$ can be proved as follow. Each row operation corresponds to multiplying to $\boldsymbol{A}$ on the left by an elementary matrix. Therefore, $\boldsymbol{E} = \boldsymbol{R}\boldsymbol{A}$, where $\boldsymbol{R}$ is the product of all the elementary matrices applied to $\boldsymbol{A}$. Since each elementary matrix is invertible, $\boldsymbol{R}$ is invertible, and hence an isomorphism. Consequently, $\operatorname{im} \boldsymbol{E} = \operatorname{im}(\boldsymbol{R}\boldsymbol{A}) = \boldsymbol{R}(\operatorname{im} \boldsymbol{A}) \cong \operatorname{im} \boldsymbol{A}$.

Remark. The row rank of a matrix $\boldsymbol{A}$ is the dimension of the row space of $\boldsymbol{A}$, or equivalently, the maximum number of linearly independent row vectors of $\boldsymbol{A}$. The column rank of $\boldsymbol{A}$ is the dimension of the column space of $\boldsymbol{A}$, or equivalently, the maximum number of linearly independent column vectors of $\boldsymbol{A}$. If a matrix $\boldsymbol{E}$ is in reduced row-echelon form, then the number of nonzero rows of $\boldsymbol{E}$ is the same as the number of pivotal columns of $\boldsymbol{E}$. Thus the row rank of $\boldsymbol{E}$ equals the column rank of $\boldsymbol{E}$. For any matrix $\boldsymbol{A}$, the elementary row operations do not change the row space of $\boldsymbol{A}$ so that the row rank of $\boldsymbol{A}$ is the same as the row rank of its reduced row-echelon form. Also each elementary row operation can be taken as an invertible linear transformation applied to the column space of $\boldsymbol{A}$ so that the column rank of $\boldsymbol{A}$ is the same as the column rank of its reduced row-echelon form. This means that the row rank of $\boldsymbol{A}$ is the same as the column rank of $\boldsymbol{A}$. We define the rank of $\boldsymbol{A}$, denoted by $\operatorname{rank}(\boldsymbol{A})$ to be either the row rank or the column rank of $\boldsymbol{A}$. Thus we may state the rank-nullity theorem as

$$\operatorname{nullity}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{A}) = n.$$

Proposition 1.4. *Let $\boldsymbol{A} \in \mathcal{M}_{m\times n}(\mathbb{K})$ and $\boldsymbol{B} \in \mathcal{M}_{n\times \ell}(\mathbb{K})$. Then*

$$\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \min(\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})).$$

Proof. Let $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{jk})$. Let A_j be the jth column of $\boldsymbol{A}$, for $j = 1, \ldots, n$. Then $\boldsymbol{AB} = (\sum_{j=1}^{n} b_{j1}A_j, \ldots, \sum_{j=1}^{n} b_{j\ell}A_j)$. Thus each column of $\boldsymbol{AB}$ is a linear combination of the columns of $\boldsymbol{A}$. Hence the column rank of $\boldsymbol{AB}$ is less than or equal to the column rank of $\boldsymbol{A}$. Similarly, the row rank of $\boldsymbol{AB}$ is less than or equal to the row rank of $\boldsymbol{B}$.

Example 1.2. *Let $\boldsymbol{A} \in \mathcal{M}_{m\times n}(\mathbb{K})$ and $\boldsymbol{B} \in \mathcal{M}_{n\times m}(\mathbb{K})$, where $m \geq n$. Suppose* $\operatorname{rank}(\boldsymbol{AB}) = n$ *and* $(\boldsymbol{AB})^2 = \boldsymbol{AB}$. *Find* $\boldsymbol{BA}$.

Solution. We have $n = \operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}((\boldsymbol{AB})^2) = \operatorname{rank}(\boldsymbol{A}(\boldsymbol{BA})\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{BA}) \leq n$. Thus the $n\times n$ matrix $\boldsymbol{BA}$ has rank n and hence is non-singular. Then $(\boldsymbol{BA})^2(\boldsymbol{BA}-\boldsymbol{I}) = \boldsymbol{B}(\boldsymbol{AB})^2\boldsymbol{A} - (\boldsymbol{BA})^2 = \boldsymbol{B}(\boldsymbol{AB})\boldsymbol{A} - (\boldsymbol{BA})^2 = \boldsymbol{O}_n$. As $(\boldsymbol{BA})^2$ is non-singular, we have $\boldsymbol{BA} - \boldsymbol{I} = \boldsymbol{O}_n$ so that $\boldsymbol{BA} = \boldsymbol{I}_n$.

Proposition 1.5. [Sylvester's Inequality] *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{K})$. Then*

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{AB}) + n.$$

Proof. It is equivalent to prove that $\operatorname{nullity}(\boldsymbol{A}) + \operatorname{nullity}(\boldsymbol{B}) \geq \operatorname{nullity}(\boldsymbol{AB})$. Denote the restriction of the linear transformation $\boldsymbol{B}$ on $\ker \boldsymbol{AB}$ by $\tilde{\boldsymbol{B}}$. Note that $\ker \boldsymbol{B} \subseteq \ker \boldsymbol{AB}$. Thus $\ker \tilde{\boldsymbol{B}} = \ker \boldsymbol{B}$. Therefore, $\operatorname{nullity}(\tilde{\boldsymbol{B}}) = \operatorname{nullity}(\boldsymbol{B})$.

Also $\operatorname{im} \tilde{\boldsymbol{B}} \subseteq \ker \boldsymbol{A}$. [$\boldsymbol{x} \in \operatorname{im} \tilde{\boldsymbol{B}} \Rightarrow \boldsymbol{x} = \tilde{\boldsymbol{B}}(\boldsymbol{v})$ for some $\boldsymbol{v} \in \ker \boldsymbol{AB} \Rightarrow \boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}(\tilde{\boldsymbol{B}}(\boldsymbol{v})) = (\boldsymbol{AB})(\boldsymbol{v}) = \boldsymbol{0}$]. Therefore $\operatorname{rank}(\tilde{\boldsymbol{B}}) \leq \operatorname{nullity}(\boldsymbol{A})$.

Applying the rank-nullity theorem [Theorem 1 (i)] to $\tilde{\boldsymbol{B}}$, we have

$$\operatorname{nullity}(\tilde{\boldsymbol{B}}) + \operatorname{rank}(\tilde{\boldsymbol{B}}) = \dim(\ker \boldsymbol{AB}).$$

That is

$$\operatorname{nullity}(\boldsymbol{B}) + \operatorname{nullity}(\boldsymbol{A}) \geq \operatorname{nullity}(\boldsymbol{AB}).$$

1.3 Characteristic Polynomial

Definition 1.9. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$. The characteristic polynomial of $\boldsymbol{A}$, denoted by $\chi_{\boldsymbol{A}}(x)$, is the polynomial* $\det(x\boldsymbol{I} - \boldsymbol{A})$.

Remark. $\chi_{\boldsymbol{A}}(x)$ is a monic polynomial of degree n.

Definition 1.10. *The characteristic polynomial of a linear transformation $T : V \longrightarrow V$ is the characteristic polynomial of any matrix representing T.*

In view of proposition 1.3, this is well defined.

Definition 1.11. *A scalar $\lambda \in \mathbb{K}$ is called an eigenvalue of T if there exists a nonzero vector $\boldsymbol{x} \in V$ such that $T(\boldsymbol{x}) = \lambda\boldsymbol{x}$. The vector $\boldsymbol{x}$ is called an eigenvector of T associated to the eigenvalue λ.*

Remark. If V has a basis $\mathcal{E}$ consisting eigenvectors of T, then the matrix representing T relative to the basis $\mathcal{E}$ is a diagonal matrix consisting of the eigenvalues of T.

Proposition 1.6. *A scalar λ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial $\chi_{\boldsymbol{A}}(x)$, where $\boldsymbol{A}$ is a matrix representing T.*

Proof. Suppose λ is an eigenvalue of T. Thus there exists $\boldsymbol{x} \neq \boldsymbol{0}$ such that $T\boldsymbol{x} = \lambda\boldsymbol{x}$. Let $\boldsymbol{A}$ be the matrix representing T with respect to a basis $\mathcal{B}$ of V. Let $\boldsymbol{X}$ be the column vector representing $\boldsymbol{x}$. Thus $\boldsymbol{AX} = \lambda\boldsymbol{X}$. That is $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = \boldsymbol{0}$. Since $\boldsymbol{x} \neq \boldsymbol{0}$, the column vector $\boldsymbol{X}$ is non-zero. Therefore the coefficient matrix $\lambda\boldsymbol{I} - \boldsymbol{A}$ must be singular. Thus $\chi_{\boldsymbol{A}}(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$. Therefore, λ is a root of the characteristic polynomial $\chi_{\boldsymbol{A}}(x)$. The converse is similar.

Theorem 2. *Cayley-Hamilton Theorem. Let $\chi_{\boldsymbol{A}}(x)$ be the characteristic polynomial of an $n \times n$ matrix $\boldsymbol{A}$. Then $\chi_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{O}_n$.*

Proof. Let $\chi_{\boldsymbol{A}}(x) = a_0 + a_1x + \cdots + a_nx^n$. The result will be proved by using the relation:

$$(x\boldsymbol{I}_n - \boldsymbol{A})\mathrm{adj}(x\boldsymbol{I}_n - \boldsymbol{A}) = \det(x\boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{I}_n. \tag{1.1}$$

The cofactors of $x\boldsymbol{I}_n - \boldsymbol{A}$ are of degree at most $n-1$ in x, so that the same is true for the entries of $\mathrm{adj}(x\boldsymbol{I}_n - \boldsymbol{A})$. Hence we may represent $\mathrm{adj}(x\boldsymbol{I}_n - \boldsymbol{A})$ as a matrix polynomial:

$$\mathrm{adj}(x\boldsymbol{I}_n - \boldsymbol{A}) = \boldsymbol{C}_0 + \boldsymbol{C}_1x + \cdots + \boldsymbol{C}_{n-1}x^{n-1}.$$

By (1.1), we have

$$(x\boldsymbol{I}_n - \boldsymbol{A})(\boldsymbol{C}_0 + \boldsymbol{C}_1x + \cdots + \boldsymbol{C}_{n-1}x^{n-1}) = (a_0 + a_1x + \cdots + a_nx^n)\boldsymbol{I}_n \tag{1.2}$$

In (1.2), we may equate corresponding coefficients and obtain the following set of equations:

$$\begin{aligned}
-\boldsymbol{AC}_0 &= a_0\boldsymbol{I}_n \\
\boldsymbol{C}_0 - \boldsymbol{AC}_1 &= a_1\boldsymbol{I}_n \\
\boldsymbol{C}_1 - \boldsymbol{AC}_2 &= a_2\boldsymbol{I}_n \\
&\vdots \\
\boldsymbol{C}_{n-2} - \boldsymbol{AC}_{n-1} &= a_{n-1}\boldsymbol{I}_n \\
\boldsymbol{C}_{n-1} &= a_n\boldsymbol{I}_n
\end{aligned}$$

Multiplying these equations on the left by $\boldsymbol{I}_n, \boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{n-1}, \boldsymbol{A}^n$ respectively, and adding, we get

$$\boldsymbol{O}_n = a_0\boldsymbol{I}_n + a_1\boldsymbol{A} + \cdots + a_n\boldsymbol{A}^n.$$

Example 1.3. The Cayley-Hamilton Theorem can be used to compute $\boldsymbol{A}^n$ effectively. Let $\boldsymbol{A} = \begin{pmatrix} \frac{1}{2} & \frac{5}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. For a 2×2 matrix $\boldsymbol{A}$, its characteristic polynomial is given by

$$\chi_{\boldsymbol{A}}(x) = x^2 - \mathrm{tr}(\boldsymbol{A})x + \det(\boldsymbol{A}).$$

Thus in this example, $\chi_{\boldsymbol{A}}(x) = x^2 - x - 1$. By the Cayley-Hamilton Theorem, $\boldsymbol{A}^2 - \boldsymbol{A} - \boldsymbol{I}_2 = \boldsymbol{O}_2$, or $\boldsymbol{A}^2 = \boldsymbol{A} + \boldsymbol{I}_2$. Therefore, $\boldsymbol{A}^4 = (\boldsymbol{A} + \boldsymbol{I}_2)^2 = 3\boldsymbol{A} + 2\boldsymbol{I}_2$, $\boldsymbol{A}^8 = (3\boldsymbol{A} + 2\boldsymbol{I}_2)^2 = 21\boldsymbol{A} + 13\boldsymbol{I}_2$, $\boldsymbol{A}^9 = 34\boldsymbol{A} + 21\boldsymbol{I}_2$, etc. For instance $\boldsymbol{A}^9 = \begin{pmatrix} 38 & 85 \\ 17 & 38 \end{pmatrix}$.

Example 1.4. Let $\boldsymbol{A} = \begin{pmatrix} \frac{1}{2} & -\frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Let's find $\boldsymbol{A}^{2015}$.

Note that $\boldsymbol{A}^6 = \boldsymbol{I}_2$. Thus $\boldsymbol{A}^{2015} = \boldsymbol{A}^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

1.4 Minimal Polynomial

Definition 1.12. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$. The minimal polynomial of $\boldsymbol{A}$, denoted by $m_{\boldsymbol{A}}(x)$, is the (unique) monic polynomial of least degree such that $m_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{O}_n$.*

Definition 1.13. *The minimal polynomial of a linear transformation $T : V \longrightarrow V$ is the minimal polynomial of any matrix representing T.*

In view of proposition 1.3, this is well defined. We shall focus on a square matrix $\boldsymbol{A}$.

Proposition 1.7. *Let $m_{\boldsymbol{A}}(x)$ be the minimal polynomial of $\boldsymbol{A}$.*
(i) If $p(x)$ is a polynomial such that $p(\boldsymbol{A}) = \boldsymbol{O}_n$, then $m_{\boldsymbol{A}}(x)$ divides $p(x)$.
(ii) λ is an eigenvalue of $\boldsymbol{A}$ if and only if λ is a root of $m_{\boldsymbol{A}}(x)$.

Proof. (i) Write $p(x) = q(x)m_{\boldsymbol{A}}(x) + r(x)$, where $r(x) = 0$, or $\deg(r(x)) < \deg(m_{\boldsymbol{A}}(x))$. If $r(x) \neq 0$, then by substituting $x = \boldsymbol{A}$ into the above equation, we get $r(\boldsymbol{A}) = 0$ with $\deg(r(x)) < \deg(m_{\boldsymbol{A}}(x))$, which contradicts the minimality of the degree of $m_{\boldsymbol{A}}(x)$. Thus $r(x) = 0$ and $p(x) = q(x)m_{\boldsymbol{A}}(x)$ which means $m_{\boldsymbol{A}}(x)$ divides $p(x)$.

(ii) Let λ be an eigenvalue of $\boldsymbol{A}$. Write $m_{\boldsymbol{A}}(x) = (x - \lambda)q(x) + c$, where c is a scalar. Thus $\boldsymbol{O} = (\boldsymbol{A} - \lambda\boldsymbol{I})q(\boldsymbol{A}) + c\boldsymbol{I}$. If $c \neq 0$, then $(\boldsymbol{A} - \lambda\boldsymbol{I})q(\boldsymbol{A})/(-c) = \boldsymbol{I}$. This implies $\boldsymbol{A}-\lambda\boldsymbol{I}$ is non-singular which contradicts the fact that λ is an eigenvalue of $\boldsymbol{A}$. Consequently, $m_{\boldsymbol{A}}(x) = (x-\lambda)q(x)$, and λ is a root of $m_{\boldsymbol{A}}(x)$. Conversely, suppose λ is a root of $m_{\boldsymbol{A}}(x)$. Thus $m_{\boldsymbol{A}}(x) = (x - \lambda)q(x)$. Then $\boldsymbol{O} = (\boldsymbol{A} - \lambda\boldsymbol{I})q(\boldsymbol{A})$. If $\boldsymbol{A} - \lambda\boldsymbol{I}$ is non-singular, then this gives $q(\boldsymbol{A}) = \boldsymbol{O}$. But $\deg(q(x)) < \deg(m_{\boldsymbol{A}}(x))$ which contradicts the minimality of the degree of $m_{\boldsymbol{A}}(x)$. Therefore $\boldsymbol{A} - \lambda\boldsymbol{I}$ is singular. This means λ is an eigenvalue of $\boldsymbol{A}$.

Remark. By 1.7, the minimal polynomial of $\boldsymbol{A}$ divides the characteristics polynomial of $\boldsymbol{A}$.

Example 1.5. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 12 & -1 \\ 0 & 1 & 0 \\ 0 & 4 & -1 \end{pmatrix}$. We have $\chi_{\boldsymbol{A}}(x) = (x-1)^2(x+1)$. Thus $m_{\boldsymbol{A}}(x) = (x-1)(x+1)$ or $(x-1)^2(x+1)$. Direct computation gives $(\boldsymbol{A}-\boldsymbol{I})(\boldsymbol{A}+\boldsymbol{I}) = \boldsymbol{0}$. Therefore, $m_{\boldsymbol{A}}(x) = (x - 1)(x + 1)$.

Definition 1.14. *$\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ is said to be diagonalizable if $\boldsymbol{A}$ is similar to a diagonal matrix.*

This is equivalent to the existence of a basis of $\mathbb{K}^n$ consisting of eigenvectors of $\boldsymbol{A}$.

Theorem 3. *$\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ is diagonalizable if and only if $m_{\boldsymbol{A}}(x) = (x-\lambda_1)\cdots(x-\lambda_k)$, where $\lambda_1, \ldots, \lambda_k$ are all the distinct eigenvalues of $\boldsymbol{A}$.*

Proof. Assume $\boldsymbol{A}$ is diagonalizable. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of $\boldsymbol{A}$. Then $\boldsymbol{A}$ is similar to the diagonal matrix $\boldsymbol{D} = \lambda_1\boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k\boldsymbol{I}_{n_k}$. Clearly the diagonal matrix $(\boldsymbol{D} - \lambda_1\boldsymbol{I})\cdots(\boldsymbol{D} - \lambda_k\boldsymbol{I})$ equals to the zero matrix $\boldsymbol{O}_n$. Thus $\boldsymbol{D}$ and hence $\boldsymbol{A}$ satisfies the polynomial $(x - \lambda_1)\cdots(x - \lambda_k)$. Since the minimal polynomial of $\boldsymbol{A}$ must contain each $(x - \lambda_i)$ as a factor, the polynomial $(x - \lambda_1)\cdots(x - \lambda_k)$ is the minimal polynomial of $\boldsymbol{A}$.

Conversely, let $m_{\boldsymbol{A}}(x) = (x - \lambda_1)\cdots(x - \lambda_k)$, where $\lambda_1, \ldots, \lambda_k$ are all the distinct eigenvalues of $\boldsymbol{A}$. Let $q_i(x) = m_{\boldsymbol{A}}(x)/(x - \lambda_i)$ for $i = 1, \ldots, k$. By resolving $1/m_{\boldsymbol{A}}(x)$ into

partial fractions and finding the common denominator, we can find scalars c_i, $i = 1, \ldots, k$. such that $c_1 q_1(x) + \cdots + c_k q_k(x) = 1$. Thus

$$\boldsymbol{I} = c_1 q_1(\boldsymbol{A}) + \cdots + c_k q_k(\boldsymbol{A}). \tag{1.3}$$

Let $V_i = \{\boldsymbol{x} \in \mathbb{K}^n \mid (\boldsymbol{A} - \lambda_i \boldsymbol{I})(\boldsymbol{x}) = \boldsymbol{0}\}$, for $i = 1, \ldots, k$. The subspace V_i is the eigenspace corresponding to the eigenvalue λ_i. Let $\boldsymbol{x} \in \mathbb{K}^n$. By (1.3), $\boldsymbol{x} = c_1 q_1(\boldsymbol{A})(\boldsymbol{x}) + \cdots + c_k q_k(\boldsymbol{A})(\boldsymbol{x})$. Since $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\left(c_i q_i(\boldsymbol{A})(\boldsymbol{x})\right) = c_i(\boldsymbol{A} - \lambda_i \boldsymbol{I}) q_i(\boldsymbol{A})(\boldsymbol{x}) = c_i m_{\boldsymbol{A}}(\boldsymbol{A})(\boldsymbol{x}) = \boldsymbol{0}$, we see that $c_i q_i(\boldsymbol{A})(\boldsymbol{x}) \in V_i$. This shows that $\mathbb{K}^n = V_1 + \cdots + V_k$. Next we shall show that $V_1 + \cdots + V_k$ is a direct sum. To do so, we need to show $V_i \cap (V_1 + \cdots + \hat{V}_i + \cdots + V_k) = \{\boldsymbol{0}\}$, for $i = 1, \ldots, k$. Here $\hat{V}_i$ means the term V_i is omitted. For simplicity, let's prove $V_1 \cap (V_2 + \cdots + V_k) = \{\boldsymbol{0}\}$. The other cases are similar. Let $\boldsymbol{v}_1 \in V_1 \cap (V_2 + \cdots + V_k)$. Thus $\boldsymbol{v}_1 \in V_1$ and there exists $-\boldsymbol{v}_i \in V_i$, $i = 2, \ldots, k$ such that $\boldsymbol{v}_1 = -(\boldsymbol{v}_2 + \cdots + \boldsymbol{v}_k)$. That is $\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k = \boldsymbol{0}$. Applying $\boldsymbol{A}, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^{k-1}$ to this equation, we have

$$\begin{array}{rcl} \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k & = & \boldsymbol{0} \\ \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_k \boldsymbol{v}_k & = & \boldsymbol{0} \\ \vdots & \vdots & \vdots \\ \lambda_1^{k-1} \boldsymbol{v}_1 + \cdots + \lambda_k^{k-1} \boldsymbol{v}_k & = & \boldsymbol{0} \end{array}$$

We may write this system in the form of a matrix equation:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \cdots & \lambda_k^{k-1} \end{pmatrix} \begin{pmatrix} c\boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_k \end{pmatrix} = \begin{pmatrix} c\boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix}$$

In this system of equations in $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, the determinant of the coefficient matrix is the Vandermonde determinant, which has value $\prod_{1 \le i < j \le k} (\lambda_i - \lambda_j)$. It is nonzero as $\lambda_1, \ldots, \lambda_k$ are all distinct. Thus the coefficient matrix is non-singular. We can multiply the above equation on the left by the inverse of the coefficient matrix and conclude that $\boldsymbol{v}_i = \boldsymbol{0}$, for all i. Therefore, $V_1 \cap (V_2 + \cdots + V_k) = \{\boldsymbol{0}\}$. Consequently, $\mathbb{K}^n = V_1 \oplus \cdots \oplus V_k$. For $i = 1, \ldots, k$, choose a basis for V_i and put them together in this order to form a basis of $\mathbb{K}^n$. Let $\boldsymbol{P}$ be the matrix form by this basis. Then $\boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P} = \lambda_1 \boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k \boldsymbol{I}_{n_k}$.

Example 1.6. Let $\boldsymbol{A} \in \mathcal{M}_{n \times n}(\mathbb{K})$ be such that $\boldsymbol{A}^2 = \boldsymbol{A}$. Suppose $\boldsymbol{A} \neq \boldsymbol{O}$ and $\boldsymbol{A} \neq \boldsymbol{I}_n$. Then $m_{\boldsymbol{A}}(x) = x(x-1)$. Thus $\boldsymbol{A}$ is diagonalizable.

Chapter 2

Complex Inner Product Spaces

Definition 2.1. *Let V be a vector space over $\mathbb{C}$. An inner product on V is a function $\langle\ ,\ \rangle : V \times V \to \mathbb{C}$ satisfying the following properties. For all $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}' \in V$ and for all $\lambda \in \mathbb{C}$,*

I1. $\langle \boldsymbol{x} + \boldsymbol{x}', \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}', \boldsymbol{y} \rangle$,
I2. $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{y}' \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{y}' \rangle$,
I3. $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\lambda} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$,
I4. $\langle \boldsymbol{x}, \lambda \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$,
I5. $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}$,
I6. *If* $\boldsymbol{x} \neq \boldsymbol{0}$, *then* $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$.

The vector space V when equipped with the inner product $\langle\ ,\ \rangle$ said to be an complex inner product space. Note that in our definition, $\langle\ ,\ \rangle$ is conjugate linear in the first variable.

We focus only on $\mathbb{C}^n$. Let $\boldsymbol{x} \in \mathbb{C}^n$. Here $\boldsymbol{x}$ is represented as a column vector, or an $n \times 1$ matrix. Denote the conjugate transpose of $\boldsymbol{x}$ by $\boldsymbol{x}^*$. The inner product on $\mathbb{C}^n$ is given by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y}$. Note that the inner product is conjugate linear in the first variable: $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\lambda} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. The *length* of a vector $\boldsymbol{x} \in \mathbb{C}^n$ is given by $|\boldsymbol{x}|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^* \boldsymbol{x}$. A basis $\{\boldsymbol{e}_1 \ldots, \boldsymbol{e}_n\}$ of $\mathbb{C}^n$ is called an orthonormal basis if $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \delta_{ij}$ for all $i, j = 1, \ldots, n$.

$\mathcal{M}_{n \times n}(\mathbb{C})$ is a complex inner product space with inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \operatorname{tr}(\boldsymbol{X}^* \boldsymbol{Y})$, which is just the Euclidean inner product when we identity $\mathcal{M}_{n \times n}(\mathbb{C})$ with $\mathbb{C}^{n^2}$.

Definition 2.2. *Let $\boldsymbol{A} \in \mathcal{M}_{n \times n}(\mathbb{C})$. $\boldsymbol{A}$ is said to be hermitian if $\boldsymbol{A} = \boldsymbol{A}^*$. $\boldsymbol{A}$ is said to be unitary if $\boldsymbol{A}^{-1} = \boldsymbol{A}^*$.*
Let $\boldsymbol{A} \in \mathcal{M}_{n \times n}(\mathbb{R})$. $\boldsymbol{A}$ is said to be symmetric if $\boldsymbol{A} = \boldsymbol{A}^\top$. $\boldsymbol{A}$ is said to be orthogonal if $\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$.

Remark. $\boldsymbol{A} = (a_{ij})$ is hermitian if and only if $a_{ji} = \overline{a}_{ij}$ for all $i, j = 1, \ldots, n$. $\boldsymbol{A}$ is unitary if and only if the column vectors of $\boldsymbol{A}$ form an orthonormal basis of $\mathbb{C}^n$.

Proposition 2.1. *The eigenvalues of a hermitian matrix are all real.*

Proof. Since $\boldsymbol{A}$ is hermitian, $(\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x})^* = \boldsymbol{x}^* \boldsymbol{A}^* \boldsymbol{x} = \boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}$ so that $\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}$ is real. If λ is an eigenvalue of $\boldsymbol{A}$ with eigenvector $\boldsymbol{x}$ of unit length, then $\lambda = \lambda \boldsymbol{x}^* \boldsymbol{x} = \boldsymbol{x}^* (\lambda \boldsymbol{x}) = \boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}$ is real. Thus the eigenvalues of a hermitian matrix are real.

Remark. The eigenvalues of a symmetric matrix are all real.

Proposition 2.2. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. There exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}$ is in upper triangular form.*

Proof. Use induction on n. The result is clearly true for $n = 1$. Let $n \geq 1$. Suppose the result is true for matrices of size $n \times n$. Let $\boldsymbol{A}$ be an $(n + 1) \times (n + 1)$ matrix over $\mathbb{C}$. Let λ be an eigenvalue of $\boldsymbol{A}$ with associated eigenvector $\boldsymbol{x}$ of unit length. Thus $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^*(\lambda\boldsymbol{x}) = \lambda\boldsymbol{x}^*\boldsymbol{x} = \lambda|\boldsymbol{x}|^2 = \lambda$. Let $\boldsymbol{P}$ be a unitary matrix whose first column is $\boldsymbol{x}$. This can be achieved by extending $\{\boldsymbol{x}\}$ to an orthonormal basis of $\mathbb{C}^{n+1}$ and putting them together to form $\boldsymbol{P}$. Then

$$\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda & * \\ \boldsymbol{0} & \boldsymbol{B} \end{pmatrix}.$$

By induction hypothesis, there exists an $n \times n$ unitary matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q}^*\boldsymbol{B}\boldsymbol{Q}$ is in upper triangular form. Let $\boldsymbol{U} = \boldsymbol{P}(\boldsymbol{I}_1 \oplus \boldsymbol{Q})$. Then $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}$ is in upper triangular form.

Definition 2.3. *A matrix $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ is said to be normal if $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.*

Definition 2.4. *Two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$ are said to be unitarily similar if there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{B}$. (Here $\boldsymbol{U}^{-1} = \boldsymbol{U}^*$ as $\boldsymbol{U}$ is unitary).*

Theorem 4. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then $\boldsymbol{A}$ is normal if and only if $\boldsymbol{A}$ is unitarily similar to a diagonal matrix.*

Solution. Suppose $\boldsymbol{A}$ is normal. Let $\boldsymbol{U}$ be a unitary matrix such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}$ is in upper triangular form. Then $\boldsymbol{T}\boldsymbol{T}^* = (\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})(\boldsymbol{U}^*\boldsymbol{A}^*\boldsymbol{U}) = \boldsymbol{U}^*\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{U} = \boldsymbol{U}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{U} = (\boldsymbol{U}^*\boldsymbol{A}^*\boldsymbol{U})(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}) = \boldsymbol{T}^*\boldsymbol{T}$. Let $\boldsymbol{T} = (t_{ij})$. Since $\boldsymbol{T}$ is upper triangular, $t_{ij} = 0$ for all $i > j$. Equating the $(1, 1)$-entry on both sides of $\boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{T}^*\boldsymbol{T}$, we get $|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2$. Thus $t_{1j} = 0$ for $j = 2, \ldots n$. Equating the $(2, 2)$-entry, we have $|t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2 = |t_{12}|^2 + |t_{22}|^2$. Since $t_{12} = 0$, we conclude that $t_{2j} = 0$ for $j = 3, \ldots, n$. Continuing in this way, we deduce that $t_{ij} = 0$ for $i < j$. Consequently $\boldsymbol{T}$ is a diagonal matrix. The converse is easy to verify.

Remark. One can prove the forward implication by induction on n. Let V be the eigenspace of $\boldsymbol{A}^*$ corresponding to an eigenvalue λ. Let $V^\perp$ be the orthogonal complement of V in $\mathbb{C}^n$. One can show that V and $V^\perp$ are subspaces of $\mathbb{C}^n$ invariant under $\boldsymbol{A}$ by using the property that $\boldsymbol{A}$ is normal. That is $\boldsymbol{A}(V) \subseteq V$ and $\boldsymbol{A}(V^\perp) \subseteq V^\perp$. Pick orthonormal bases $\mathcal{B}_1$ of V and $\mathcal{B}_2$ of $V^\perp$ so that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a basis of $\mathbb{C}^n$. Let $\boldsymbol{P}$ be the unitary matrix formed by the orthonormal basis $\mathcal{B}$. Then $\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P} = \boldsymbol{A}_1 \oplus \boldsymbol{A}_2$, where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are normal and of size small than n. The result follows by applying induction hypothesis to $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$.

Example 2.1. If $\boldsymbol{A}$ is hermitian or unitary, then $\boldsymbol{A}$ is normal, and hence unitarily diagonalizable. Similarly, if $\boldsymbol{A}$ is real and symmetric, then $\boldsymbol{A}$ is diagonalizable by orthogonal matrices.

2.1 Dual Spaces

Definition 2.5. *Let V be an n-dimensional vector space over $\mathbb{C}$. The set of all linear transformations from V to $\mathbb{C}$ is a vector space over $\mathbb{C}$. It is called the dual space of V, and is denoted by $\mathrm{Hom}(V, \mathbb{C})$, or V^*. Elements of V^* are called linear functionals.*

Proposition 2.3. *Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ be a basis of V. For each $i = 1, \ldots, n$, let $\boldsymbol{e}_i^* : V \longrightarrow \mathbb{C}$ be the linear functional defined by $\boldsymbol{e}_i^*(\boldsymbol{e}_j) = \delta_{ij}$, extended by linearity. Then $\{\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*\}$ is a basis of V^*.*

Proof. Let $\boldsymbol{f} \in V^*$, and let $\boldsymbol{f}(\boldsymbol{e}_j) = \alpha_j \in \mathbb{C}$, for $j = 1, \ldots, n$. We claim that $\boldsymbol{f} = \sum_{i=1}^n \alpha_i \boldsymbol{e}_i^*$. This is because

$$\left(\sum_{i=1}^n \alpha_i \boldsymbol{e}_i^*\right)(\boldsymbol{e}_j) = \sum_{i=1}^n \alpha_i \boldsymbol{e}_i^*(\boldsymbol{e}_j) = \sum_{i=1}^n \alpha_i \delta_{ij} = \alpha_j = \boldsymbol{f}(\boldsymbol{e}_j).$$

Thus both $\boldsymbol{f}$ and $\sum_{i=1}^n \alpha_i \boldsymbol{e}_i^*$ agree on the basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ of V, and hence they are equal. This shows that every element of V^* is a linear combination of $\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*$.
To show that $\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*$ are linearly independent, suppose $\sum_{i=1}^n \lambda_i \boldsymbol{e}_i^* = \boldsymbol{0}$. Here $\boldsymbol{0}$ is the zero function. Then $\lambda_j = \left(\sum_{i=1}^n \lambda_i \boldsymbol{e}_i^*\right)(\boldsymbol{e}_j) = 0$ for all $j = 1, \ldots, n$. Thus $\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*$ are linearly independent.

Remark. The correspondence between V and V^* defined by sending a basis of V to its dual basis is not *natural*. If V is a complex inner product space, then the correspondence can be defined by $\boldsymbol{v} \mapsto \langle \boldsymbol{v}, \ \rangle$ which is a *conjugate linear isomorphism* $\alpha_V : V \longrightarrow V^*$. Recall that for a complex inner product space, the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is conjugate linear in the first variable. That is $\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\alpha} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. For example in $\mathbb{C}^n$, a vector is denoted as an $n \times 1$ column matrix $\boldsymbol{x}$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y}$ is conjugate linear in the first variable.

Proposition 2.4. *Let $T : V \longrightarrow W$ be a linear transformation. Then $T^t : W^* \longrightarrow V^*$ defined by $T(f) = f \circ T$ is a linear transformation.*

Proposition 2.5. *Let $\mathcal{E} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ and $\mathcal{F} = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m\}$ be bases of V and W respectively. Thus $\mathcal{E}^* = \{\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*\}$ and $\mathcal{F}^* = \{\boldsymbol{f}_1^*, \ldots, \boldsymbol{f}_m^*\}$ are bases of V^* and W^* respectively. Let $\boldsymbol{A}$ be the matrix representing T with respect to the bases $\mathcal{E}$ of V and $\mathcal{F}$ of W. Show that $\boldsymbol{A}^\top$ is the matrix representing T^t with respect to the bases $\mathcal{F}^*$ of W^* and $\mathcal{E}^*$ of V^*.*

Proof. For each $j = 1, \ldots, n$, let $T(\boldsymbol{e}_j) = \sum_{k=1}^m a_{kj} \boldsymbol{f}_k$. Thus $\boldsymbol{A} = (a_{ij})_{m \times n}$. Let $T^t(\boldsymbol{f}_j^*) = \sum_{k=1}^n b_{kj} \boldsymbol{e}_k^*$. That is the matrix representing T^t is $(b_{ij})_{n \times m}$. By the definition of T^t, we have $\boldsymbol{f}_j^* \circ T = \sum_{k=1}^n b_{kj} \boldsymbol{e}_k^*$. Applying both sides to $\boldsymbol{e}_i$, we have
LHS equals $\left(\boldsymbol{f}_j^* \circ T\right) \boldsymbol{e}_i = \boldsymbol{f}_j^*(T(\boldsymbol{e}_i)) = \boldsymbol{f}_j^* \left(\sum_{k=1}^m a_{ki} \boldsymbol{f}_k\right) = a_{ji}$.
RHS equals $\left(\sum_{k=1}^n b_{kj} \boldsymbol{e}_k^*\right) \boldsymbol{e}_i = b_{ij}$. Thus the matrix representing T^t is $A^\top$.

Remark. If V is a complex inner product space, then the correspondence $\alpha_V : V \longrightarrow V^*$ defined by $\boldsymbol{v} \mapsto \langle \boldsymbol{v}, \ \rangle$ is a conjugate linear isomorphism. Then T^* defined by $T^* = \alpha_V^{-1} \circ T^t \circ \alpha_W : W \longrightarrow V$ is called the *adjoint* of T. If $\mathcal{E} = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ and $\mathcal{F} = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m\}$ are orthonormal bases of V and W respectively, then the matrix representing T^* relative to the dual bases is $\boldsymbol{A}^*$. The definition that $T^* = \alpha_V^{-1} \circ T^t \circ \alpha_W$ can be written as $\alpha_V \circ T^* = T^t \circ \alpha_W$. Applying both sides to $\boldsymbol{w} \in W$ and then to $\boldsymbol{v} \in V$, we obtain the equation

$$\langle T^*(\boldsymbol{w}), \boldsymbol{v} \rangle = \langle \boldsymbol{w}, T(\boldsymbol{v}) \rangle.$$

Here the LHS is the inner product in V and the RHS is the inner product in W. Note that by the definition of T^t, we have $T^t(\alpha_W(\boldsymbol{w})) = T^t\left(\langle \boldsymbol{w}, \ \rangle\right) = \langle \boldsymbol{w}, T(\ ) \rangle$. Let $T^*(\boldsymbol{f}_i) = \sum_{k=1}^n c_{ki} \boldsymbol{e}_k$ and $T(\boldsymbol{e}_j) = \sum_{k=1}^m a_{kj} \boldsymbol{f}_k$. Then

$$\langle T^*(\boldsymbol{f}_i), \boldsymbol{e}_j \rangle = \langle \boldsymbol{f}_i, T(\boldsymbol{e}_j) \rangle \Rightarrow \langle \sum_{k=1}^n c_{ki} \boldsymbol{e}_k, \boldsymbol{e}_j \rangle = \langle \boldsymbol{f}_i, \sum_{k=1}^m a_{kj} \boldsymbol{f}_k \rangle \Rightarrow \overline{c}_{ji} = a_{ij}.$$

Therefore $(c_{ij}) = (\overline{a}_{ji})$. It follows either from this or the equation $\langle T^*(\boldsymbol{w}), \boldsymbol{v} \rangle = \langle \boldsymbol{w}, T(\boldsymbol{v}) \rangle$ that $T^{**} = T$.

2.2 Positive Definite Matrices

Definition 2.6. *A hermitian matrix $\boldsymbol{A}$ is said to be positive definite if $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} > 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}$.*
A hermitian matrix $\boldsymbol{A}$ is said to be positive semidefinite if $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} \geq 0$, for all $\boldsymbol{x} \neq \boldsymbol{0}$.

Theorem 5. *Let $\boldsymbol{A}$ be a hermitian matrix. Then the following statements are equivalent.*
(a) $\boldsymbol{A}$ is positive definite.
(b) The eigenvalues of $\boldsymbol{A}$ are positive.
(c) $\boldsymbol{A} = \boldsymbol{H}^*\boldsymbol{H}$ for some invertible matrix $\boldsymbol{H}$.

Proof. (a) $\Rightarrow$ (b). Let λ be an eigenvalue of $\boldsymbol{A}$ with associated eigenvector $\boldsymbol{x}$ of unit length. Since $\boldsymbol{x}$ is an eigenvector, it is nonzero. As $\boldsymbol{A}$ is positive definite, we have $\lambda = \lambda|\boldsymbol{x}|^2 = \lambda\boldsymbol{x}^*\boldsymbol{x} = \boldsymbol{x}^*\lambda\boldsymbol{x} = \boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} > 0$.
(b) $\Rightarrow$ (c). Since $\boldsymbol{A}$ is hermitian, all its eigenvalues are real, and there exists an invertible matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{D}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$. By (b), $\lambda_i > 0$ for all $i = 1, \ldots, n$. Let $\boldsymbol{K} = \boldsymbol{U}\boldsymbol{D}(\lambda_1^{-\frac{1}{2}}, \ldots, \lambda_n^{-\frac{1}{2}})$. Then $\boldsymbol{K}$ is invertible and $\boldsymbol{K}^* = \boldsymbol{D}(\lambda_1^{-\frac{1}{2}}, \ldots, \lambda_n^{-\frac{1}{2}})\boldsymbol{U}^*$. Thus $\boldsymbol{K}^*\boldsymbol{A}\boldsymbol{K} = \boldsymbol{I}_n$. That is $\boldsymbol{A} = (\boldsymbol{K}^*)^{-1}\boldsymbol{I}_n\boldsymbol{K}^{-1} = (\boldsymbol{K}^*)^{-1}\boldsymbol{K}^{-1} = (\boldsymbol{K}^{-1})^*\boldsymbol{K}^{-1}$. Letting $\boldsymbol{H} = \boldsymbol{K}^{-1}$, we have $\boldsymbol{A} = \boldsymbol{H}^*\boldsymbol{H}$.
(c) $\Rightarrow$ (a). Suppose $\boldsymbol{A} = \boldsymbol{H}^*\boldsymbol{H}$. Then for any $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^*\boldsymbol{H}^*\boldsymbol{H}\boldsymbol{x} = (\boldsymbol{H}\boldsymbol{x})^*(\boldsymbol{H}\boldsymbol{x}) = |\boldsymbol{H}\boldsymbol{x}|^2 > 0$, since $\boldsymbol{H}$ is invertible.

Proposition 2.6. *Let $\boldsymbol{A}$ be a $n \times n$ positive semidefinite hermitian matrix such that $\boldsymbol{A}^2$ is a diagonal matrix. Then $\boldsymbol{A}$ is a diagonal matrix.*

Proof. Since $\boldsymbol{A}$ is positive semidefinite and hermitian, its eigenvalues are real and nonnegative. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of $\boldsymbol{A}$ and n_i the multiplicity of λ_i. $\boldsymbol{A}$ is hermitian implies that $\boldsymbol{A}$ is unitarily diagonalizable. That is there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \lambda_1\boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k\boldsymbol{I}_{n_k}$. It follows that $\boldsymbol{U}^*\boldsymbol{A}^2\boldsymbol{U} = \lambda_1^2\boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k^2\boldsymbol{I}_{n_k}$. Let f be a polynomial such that $f(\lambda_i^2) = \lambda_i$ for all $i = 1, \ldots, k$. The polynomial f can be constructed by Lagrange interpolation formula as the λ_i^2's are distinct. Then $\boldsymbol{U}^*f(\boldsymbol{A}^2)\boldsymbol{U} = f(\boldsymbol{U}^*\boldsymbol{A}^2\boldsymbol{U}) = f(\lambda_1^2\boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k^2\boldsymbol{I}_{n_k}) = f(\lambda_1^2)\boldsymbol{I}_{n_1} \oplus \cdots \oplus f(\lambda_k^2)\boldsymbol{I}_{n_k} = \lambda_1\boldsymbol{I}_{n_1} \oplus \cdots \oplus \lambda_k\boldsymbol{I}_{n_k} = \boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}$. Thus $f(\boldsymbol{A}^2) = \boldsymbol{A}$. Since $\boldsymbol{A}^2$ is a diagonal matrix, it follows that $\boldsymbol{A} = f(\boldsymbol{A}^2)$ is also a diagonal matrix. Consequently, if $\boldsymbol{A}^2 = \boldsymbol{D}(\alpha_1, \ldots, \alpha_n)$, then $\alpha_i \geq 0$ for all i and $\boldsymbol{A} = \boldsymbol{D}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n})$.

Remark. Thus a positive semidefinite diagonal matrix has a unique positive semidefinite hermitian square root and this root is also diagonal. Note that the result may not be true if $\boldsymbol{A}$ is not positive semidefinite. For example consider $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $\boldsymbol{A}^2 = \boldsymbol{I}_2$ certainly has many square roots.

Proposition 2.7. *Let $\boldsymbol{B}$ and $\boldsymbol{C}$ be $n \times n$ positive semidefinite hermitian matrices. If $\boldsymbol{B}^2 = \boldsymbol{C}^2$, then $\boldsymbol{B} = \boldsymbol{C}$.*

Proof. Let $\boldsymbol{U}$ be a unitary matrix such that $\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U}$ is a diagonal matrix $\boldsymbol{D}$. Then $\boldsymbol{D}^2 = (\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U})^2 = \boldsymbol{U}^*\boldsymbol{B}^2\boldsymbol{U} = \boldsymbol{U}^*\boldsymbol{C}^2\boldsymbol{U} = (\boldsymbol{U}^*\boldsymbol{C}\boldsymbol{U})^2$. Since the positive semidefinite hermitian square root of $\boldsymbol{D}^2$ is unique, we have $\boldsymbol{D} = \boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U} = \boldsymbol{U}^*\boldsymbol{C}\boldsymbol{U}$. Therefore, $\boldsymbol{B} = \boldsymbol{C}$.

2nd Proof (*by Tang Wai Shing*). The matrix $\boldsymbol{C}-\boldsymbol{B}$ is hermitian. We will show that all eigenvalues of $\boldsymbol{C}-\boldsymbol{B}$ are zero. Consider $(\boldsymbol{C}-\boldsymbol{B})\boldsymbol{x}=\lambda\boldsymbol{x}$, where $\lambda\in\mathbb{R}, \boldsymbol{x}\in\mathbb{C}^n, \boldsymbol{x}^*\boldsymbol{x}=1$. We have $\boldsymbol{x}^*(\boldsymbol{C}\boldsymbol{x}-\boldsymbol{B}\boldsymbol{x})=\lambda\boldsymbol{x}^*\boldsymbol{x}=\lambda$. We claim that $\lambda(\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}+\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x})=0$.

$\lambda\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}=\boldsymbol{x}^*\boldsymbol{B}(\lambda\boldsymbol{x})=\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{C}\boldsymbol{x}-\boldsymbol{x}^*\boldsymbol{B}^2\boldsymbol{x}=\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{C}\boldsymbol{x}-\boldsymbol{x}^*\boldsymbol{C}^2\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{B}-\boldsymbol{C})\boldsymbol{C}\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{B}-\boldsymbol{C})^*\boldsymbol{C}\boldsymbol{x}=-\lambda\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}$. Therefore, $\lambda(\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}+\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x})=0$.

Note that $\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}\geq 0$ and $\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}\geq 0$. If $\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}+\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}>0$, then $\lambda=0$.
If $\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}+\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}=0$, then $\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}=\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}=0$, and thus $\lambda=\boldsymbol{x}^*\boldsymbol{C}\boldsymbol{x}-\boldsymbol{x}^*\boldsymbol{B}\boldsymbol{x}=0$. Consequently, all eigenvalues of $\boldsymbol{C}-\boldsymbol{B}$ are zero and so $\boldsymbol{C}-\boldsymbol{B}=\boldsymbol{O}$.

Proposition 2.8. *Let $\boldsymbol{A}$ be an $n\times n$ positive semidefinite hermitian matrix. Then there exists a unique positive semidefinite hermitian $\boldsymbol{R}$ such that $\boldsymbol{R}^2=\boldsymbol{A}$. That is $\boldsymbol{A}$ has a unique positive semidefinite hermitian square root.*

Proof. As $\boldsymbol{A}$ is hermitian, there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}=\boldsymbol{D}(\lambda_1,\ldots,\lambda_n)$, where $\lambda_1,\ldots,\lambda_n$ are the eigenvalues of $\boldsymbol{A}$. Since $\boldsymbol{A}$ is positive semidefinite, $\lambda_i\geq 0$ for all $i=1,\ldots,n$. Let $\boldsymbol{R}=\boldsymbol{U}\boldsymbol{D}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n})\boldsymbol{U}^*$. Then $\boldsymbol{R}^2=\boldsymbol{A}$. That $\boldsymbol{R}$ is unique follows from proposition 2.7.

Proposition 2.9. *Let $\boldsymbol{A}$ be an $n\times n$ nonsingular matrix. There exist unique positive definite hermitian matrices $\boldsymbol{R}$ and $\boldsymbol{S}$ and a unique unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{A}=\boldsymbol{R}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{S}$. This is called the polar representation of $\boldsymbol{A}$.*

Proof. First $\boldsymbol{A}\boldsymbol{A}^*$ is positive definite hermitian, it has a unique positive definite hermitian square root $\boldsymbol{R}$. That is $\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{R}^2$. Let $\boldsymbol{U}=\boldsymbol{R}^{-1}\boldsymbol{A}$. Then $\boldsymbol{A}=\boldsymbol{R}\boldsymbol{U}$. Next we check that $\boldsymbol{U}$ is unitary: $\boldsymbol{U}\boldsymbol{U}^*=(\boldsymbol{R}^{-1}\boldsymbol{A})(\boldsymbol{R}^{-1}\boldsymbol{A})^*=\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{A}^*(\boldsymbol{R}^{-1})^*=\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{A}^*(\boldsymbol{R}^*)^{-1}=\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{R}^{-1}=\boldsymbol{I}_n$ as $\boldsymbol{R}$ is hermitian. Similarly $\boldsymbol{U}^*\boldsymbol{U}=\boldsymbol{I}_n$. Since $\boldsymbol{R}$ is unique, $\boldsymbol{U}=\boldsymbol{R}^{-1}\boldsymbol{A}$ is uniquely determined. Next we define $\boldsymbol{S}=\boldsymbol{U}^*\boldsymbol{R}\boldsymbol{U}$. Then $\boldsymbol{S}$ is positive definite hermitian, and $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}$.

Example 2.2. Let $\boldsymbol{A}=\begin{pmatrix}-1 & 8\\ 4 & 4\end{pmatrix}$. Following proposition 2.9, we find that

$$\boldsymbol{R}=\begin{pmatrix}\frac{101}{13} & \frac{28}{13}\\ \frac{28}{13} & \frac{68}{13}\end{pmatrix},\quad \boldsymbol{U}=\begin{pmatrix}\frac{-5}{13} & \frac{12}{13}\\ \frac{12}{13} & \frac{5}{13}\end{pmatrix},\quad \boldsymbol{S}=\begin{pmatrix}\frac{53}{13} & \frac{8}{13}\\ \frac{8}{13} & \frac{116}{13}\end{pmatrix}.$$

Then $\boldsymbol{A}=\boldsymbol{R}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{S}$.

Example 2.3. Find the polar representation of the matrix $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$.

Following proposition 2.9, we find that $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}=\begin{pmatrix}\frac{3}{\sqrt{5}} & \frac{1}{\sqrt{5}}\\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{pmatrix}\begin{pmatrix}\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}}\\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\end{pmatrix}$.

Chapter 3

Trace and Commutators

The commutator of two matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{M}_{n\times n}(\mathbb{K})$ is defined to be $[\boldsymbol{X}, \boldsymbol{Y}] \equiv \boldsymbol{XY} - \boldsymbol{YX}$. Note that $\mathrm{tr}([\boldsymbol{X}, \boldsymbol{Y}]) = 0$.

Proposition 3.1. *Let $\alpha \in \mathbb{K}$. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$, where $n \geq 2$ and $\boldsymbol{A}$ is not a scalar multiple of the identity. Then there exists an invertible matrix $\boldsymbol{P}$ such that the first diagonal entry of $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is α.*

Proof. Since $\boldsymbol{A}$ is not a scalar multiple of the identity, $\boldsymbol{A}$ is not the zero matrix in particular.
Claim. There exists a non-zero vector $\boldsymbol{x}$ such that $\boldsymbol{x}$ and $\boldsymbol{Ax}$ are linearly independent.
Proof. Suppose, on the contrary, that for any non-zero vector $\boldsymbol{x}$, $\boldsymbol{x}$ and $\boldsymbol{Ax}$ are linearly dependent. That is for any non-zero vector $\boldsymbol{x}$, we have $\boldsymbol{Ax} = \lambda_x \boldsymbol{x}$. Let $\boldsymbol{u}, \boldsymbol{v}$ be two linearly independent vectors. Thus $\boldsymbol{Au} = \lambda_u \boldsymbol{u}$, $\boldsymbol{Av} = \lambda_v \boldsymbol{v}$ and $\boldsymbol{A}(\boldsymbol{u}+\boldsymbol{v}) = \lambda_{u+v}(\boldsymbol{u}+\boldsymbol{v})$. It follows that $\lambda_{u+v}(\boldsymbol{u}+\boldsymbol{v}) = \lambda_u \boldsymbol{u} + \lambda_v \boldsymbol{v}$. That is $(\lambda_{u+v} - \lambda_u)\boldsymbol{u} + (\lambda_{u+v} - \lambda_v)\boldsymbol{v} = \boldsymbol{0}$. Since $\boldsymbol{u}, \boldsymbol{v}$ are linearly independent, we have $\lambda_{u+v} = \lambda_u = \lambda_v$. Applying this to a basis of $\mathbb{K}^n$ shows that $\boldsymbol{A}$ is a scalar multiple of the identity, contradicting the hypothesis. This proves the claim.
Now let $\boldsymbol{x}$ be a non-zero vector such that $\boldsymbol{x}$ and $\boldsymbol{Ax}$ are linearly independent. Then there exists a non-zero vector $\boldsymbol{y}$ such that $\boldsymbol{y}^\top \boldsymbol{x} = 1$ and $\boldsymbol{y}^\top \boldsymbol{Ax} = \alpha$. This is done by solving the linear system of equations: $\boldsymbol{y}^\top \boldsymbol{x} = 1, \boldsymbol{y}^\top \boldsymbol{Ax} = \alpha$ in the unknown vector $\boldsymbol{y}$.
Choose a basis $\mathcal{B} = \{\boldsymbol{x}, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ of $\mathbb{K}^n$, where $\boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are in the subspace annihilated by $\boldsymbol{y}$. That is $\boldsymbol{y}^\top \boldsymbol{e}_i = 0, i = 2, \ldots, n$. (Note that $\boldsymbol{x}$ is not in this subspace as $\boldsymbol{y}^\top \boldsymbol{x} = 1$. Thus such a basis $\mathcal{B}$ can be chosen.)
Let $\boldsymbol{P} = (\boldsymbol{x}, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$ be the matrix formed by the column vectors $\boldsymbol{x}, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$. Then the first row of $\boldsymbol{P}^{-1}$ is $\boldsymbol{y}^\top$, as $\boldsymbol{y}^\top \boldsymbol{x} = 1$, and $\boldsymbol{y}^\top \boldsymbol{e}_i = 0$, for $i = 2, \ldots, n$. We see that the first diagonal entry of $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is $\boldsymbol{y}^\top \boldsymbol{Ax} = \alpha$. Thus the conclusion holds.

Remark. $\boldsymbol{P}$ can be chosen to be unitary if α lies in the numerical range of $\boldsymbol{A}$. The numerical range of $\boldsymbol{A}$ is defined to be $W(\boldsymbol{A}) = \{\boldsymbol{x}^* \boldsymbol{Ax} \in \mathbb{C} \mid \|\boldsymbol{x}\| = 1\}$ and is known to be a compact convex subset of $\mathbb{C}$. Note that $W(\boldsymbol{A})$ contains all the eigenvalues of $\boldsymbol{A}$. For instance, if $\mathrm{tr}(\boldsymbol{A}) = 0$, then $0 \in W(\boldsymbol{A})$.

Proposition 3.2. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ be such that $\mathrm{tr}(\boldsymbol{A}) = 0$. Then $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B}$ whose diagonal entries are all 0.*

Proof. We use induction on n. The case $n = 1$ is trivially true. Let $n \geq 2$.

Suppose any $(n-1) \times (n-1)$ complex matrix with trace 0 is similar to a matrix whose diagonal entries are all 0.

Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ such that $\mathrm{tr}(\boldsymbol{A}) = 0$. If $\boldsymbol{A}$ is a scalar matrix, then $\mathrm{tr}(\boldsymbol{A}) = 0$ implies that $\boldsymbol{A}$ is just the zero matrix, and so the conclusion holds. Now we consider $\boldsymbol{A}$ is not a scalar matrix. By 3.1, there exists an invertible matrix $\boldsymbol{P}$ such that the first diagonal element of $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is zero. Thus

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} 0 & \boldsymbol{u}^\top \\ \boldsymbol{v} & \boldsymbol{X} \end{pmatrix},$$

where $\boldsymbol{u}, \boldsymbol{v}$ are $(n-1)\times 1$ matrices, and $\boldsymbol{X}$ is a $(n-1)\times(n-1)$ matrix. Also $\mathrm{tr}(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}) = 0$.
By induction hypothesis applied to $\boldsymbol{X}$, there exists an invertible $(n-1)\times(n-1)$ matrix $\boldsymbol{R}$ such that the diagonal entries of $\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{R}$ are all 0.
Let

$$\boldsymbol{Q} = \begin{pmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{R} \end{pmatrix}.$$

Then

$$(\boldsymbol{P}\boldsymbol{Q})^{-1}\boldsymbol{A}(\boldsymbol{P}\boldsymbol{Q}) = \begin{pmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{R}^{-1} \end{pmatrix}\begin{pmatrix} 0 & \boldsymbol{u}^\top \\ \boldsymbol{v} & \boldsymbol{X} \end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{R} \end{pmatrix} = \begin{pmatrix} 0 & \boldsymbol{u}^\top\boldsymbol{R} \\ \boldsymbol{R}^{-1}\boldsymbol{v} & \boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{R} \end{pmatrix}$$

has all the diagonal entries equal to 0.

Proposition 3.3. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ be such that all of its diagonal entries are 0. Then $\boldsymbol{A}$ is a commutator.*

Proof. Let $\boldsymbol{A} = (a_{ij})$. Take any set of distinct numbers $x_1, \ldots, x_n$. Let $\boldsymbol{X}$ be the diagonal matrix $\boldsymbol{D}(x_1, \ldots, x_n)$. Let $\boldsymbol{Y} = (y_{ij})$ be the matrix defined by

$$y_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{a_{ij}}{x_i - x_j} & \text{if } i \neq j \end{cases}.$$

One can easily check that $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}$.

Summarizing, we have the following result.

Theorem 6. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$. Then the following statements are equivalent.*
(1) $\mathrm{tr}(\boldsymbol{A}) = 0$.
(2) $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B}$ whose diagonal entries are all 0.
(3) $\boldsymbol{A}$ is a commutator.

Remark. It can be proved directly that a traceless matrix is a sum of commutators using elemenatry matrices.

Proposition 3.4. *Let D be the set of all commutators in $\mathcal{M}_{n\times n}(\mathbb{K})$. $\mathrm{tr} : \mathcal{M}_{n\times n}(\mathbb{K}) \to \mathbb{K}$ induces an isomorhism $\mathcal{M}_{n\times n}(\mathbb{K})/D \cong \mathbb{K}$.*

Proof. This follows from theorem 6.

Proposition 3.5. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$. Then $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B}$ whose diagonal entries are all equal.*

Proof. This follows by applying 3.2 to the matrix $\boldsymbol{A} - \frac{\mathrm{tr}(\boldsymbol{A})}{n}\boldsymbol{I}$ which has trace zero.

Proposition 3.6. *A sum of commutators is a commutator.*

Proof. Since the trace of a sum of commutators is zero, it is also a commutator by theorem 6.

Proposition 3.7. *Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{K})$ which is not a nonzero scalar multiple of the identity. Then $\boldsymbol{A}$ is similar to a matrix $\boldsymbol{B} = (b_{ij})$ such that $b_{11} = \operatorname{tr}(\boldsymbol{A})$ and $b_{ii} = 0$ for all $i = 2, \ldots, n$.*

Proof. This follows from 3.1 by taking $\alpha = \operatorname{tr}(\boldsymbol{A})$, and then apply 3.2 to the resulting lower $(n-1)\times(n-1)$ submatrix which has trace zero.

Example 3.1. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix}$. Find a matrix $\boldsymbol{P}$ so that the diagonal entries of $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ are all zero.

Solution. Take $\boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. Then $\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$. Then $\boldsymbol{x}$ and $\boldsymbol{A}\boldsymbol{x}$ are linearly independent. Let $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. Let's solve the system $\boldsymbol{y}^\top\boldsymbol{x} = 1, \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{x} = 0$. That is $y_1 + y_2 = 1, y_1 + 2y_2 = 0$.

We obtain $y_1 = 2, y_2 = -1$ and y_3 is arbitrary. So let's take $y_3 = 0$ so that $\boldsymbol{y} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

Next we take $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Then $\boldsymbol{e}_1, \boldsymbol{e}_2$ are in the subspace annihilated by $\boldsymbol{y}$.

Take $\boldsymbol{P} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ with $\boldsymbol{P}^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$.

Using exercise 2.7, we can further conjugate the lower 2×2 matrix to $\begin{pmatrix} 0 & -9 \\ -1 & 0 \end{pmatrix}$ by means of the matrix $\begin{pmatrix} 1 & -3 \\ 1 & 3 \end{pmatrix}$. Combining the two conjugations, we have

$$\begin{pmatrix} 1 & 1 & -3 \\ 1 & 2 & -6 \\ 0 & 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & -3 \\ 1 & 2 & -6 \\ 0 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 6 \\ \frac{1}{2} & 0 & -9 \\ -\frac{1}{6} & -1 & 0 \end{pmatrix}.$$

Chapter 4

Conjugacy Classes in $SL(2,\mathbb{R})$

4.1 Conjugacy Classes

Let $\boldsymbol{X} \in SL(2,\mathbb{R})$. The characteristics polynomial of $\boldsymbol{X}$ is $x^2 - \mathrm{tr}(\boldsymbol{X})x + 1 = 0$. Thus the eigenvalues of $\boldsymbol{X}$ are $\frac{1}{2}\left(\mathrm{tr}(\boldsymbol{X}) \pm \sqrt{\mathrm{tr}(\boldsymbol{X})^2 - 4}\right)$.

Case 1. $|\mathrm{tr}(\boldsymbol{X})| < 2$.

The eigenvalues are complex, $\alpha \pm i\beta$, where $\beta \neq 0$ and $\alpha^2 + \beta^2 = 1$. Let $\boldsymbol{v} \in \mathbb{C}^2$ be an eigenvector of $\boldsymbol{X}$ corresponding to the eigenvalue $\alpha + i\beta$. Then $\overline{\boldsymbol{v}} \in \mathbb{C}^2$ is an eigenvector of $\boldsymbol{X}$ corresponding to the eigenvalue $\alpha - i\beta$. Since $\beta \neq 0$, the eigenvalues $\alpha + i\beta$ and $\alpha - i\beta$ are distinct so that $\boldsymbol{v}$ and $\overline{\boldsymbol{v}}$ are linearly independent over $\mathbb{C}$. This implies $\Re(\boldsymbol{v}) = \frac{1}{2}(\boldsymbol{v} + \overline{\boldsymbol{v}})$ and $\Im(\boldsymbol{v}) = \frac{1}{2i}(\boldsymbol{v} - \overline{\boldsymbol{v}})$ are linearly independent over $\mathbb{C}$. Hence $\Re(\boldsymbol{v})$ and $\Im(\boldsymbol{v})$ are linearly independent over $\mathbb{R}$.

Let $\boldsymbol{v}_1 = \Re(\boldsymbol{v})$ and $\boldsymbol{v}_2 = \Im(\boldsymbol{v})$. Then $\boldsymbol{X}(\boldsymbol{v}_1 + i\boldsymbol{v}_2) = (\alpha + i\beta)(\boldsymbol{v}_1 + i\boldsymbol{v}_2) \Leftrightarrow \boldsymbol{X}\boldsymbol{v}_1 + i\boldsymbol{X}\boldsymbol{v}_2 = (\alpha\boldsymbol{v}_1 - \beta\boldsymbol{v}_2) + i(\beta\boldsymbol{v}_1 + \alpha\boldsymbol{v}_2)$. Thus $\boldsymbol{X}\boldsymbol{v}_1 = \alpha\boldsymbol{v}_1 - \beta\boldsymbol{v}_2$ and $\boldsymbol{X}\boldsymbol{v}_2 = \beta\boldsymbol{v}_1 + \alpha\boldsymbol{v}_2$. Therefore, with respect to the ordered basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$, the matrix representing $\boldsymbol{X}$ is $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$. Whereas, with respect to the ordered basis $\{\boldsymbol{v}_2, \boldsymbol{v}_1\}$, the matrix representing $\boldsymbol{X}$ is $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$.

We pick the ordered basis either $\{\boldsymbol{v}_2, \boldsymbol{v}_1\}$ or $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ which has positive determinant d. Using that ordered basis, we may scale the matrix of change of basis by a factor of $\sqrt{d}$, so that $\boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ or $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$, and $\det(\boldsymbol{P}) = 1$.

Let $\alpha = \cos\theta, \beta = \sin\theta, \theta \in (0,\pi) \cup (\pi, 2\pi)$. Then $\boldsymbol{X}$ is conjugate in $SL(2,\mathbb{R})$ to either

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Note that $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ and $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ are not conjuage in $SL(2,\mathbb{R})$, where $\theta \in (0,\pi) \cup (\pi, 2\pi)$.

To see this, assume $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}^{-1} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} p & q \\ r & s \end{pmatrix}, ps - qr = 1.$

That is $\begin{pmatrix} p & q \\ r & s \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} p & q \\ r & s \end{pmatrix}$. Equating entries and simplifying, we obtain $(q-r)\sin\theta = 0, (p+s)\sin\theta = 0$. Since $\theta \in (0,\pi)\cup(\pi,2\pi)$, we have $\sin\theta \neq 0$. Thus $q = r, p = -s$ so that $1 = ps - qr = -p^2 - q^2$, a contradiction.

Case 2. $|\mathrm{tr}(\boldsymbol{X})| > 2$.

The eigenvalues are real, nonzero and unequal. Let the eigenvalues be α and $\frac{1}{\alpha}$, $\alpha \neq 0, \pm 1$. Then $\boldsymbol{X}$ is conjugate to $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ or $\begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \alpha \end{pmatrix}$. These two matrices are conjugate in $SL(2,R)$.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{-1}\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \alpha \end{pmatrix}.$$

Therefore $\boldsymbol{X}$ is conjuagate to $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ with $|\alpha| > 1$.

Case 3. $\mathrm{tr}(\boldsymbol{X}) = 2$.

The eigenvalues are 1 with multiplicity 2.

(3i) $\boldsymbol{X}$ has 2 linearly independent eigenvectors. Then $\boldsymbol{X} = \boldsymbol{I}$.

(3ii) $\boldsymbol{X}$ does not have 2 linearly independent eigenvectors. Let $\boldsymbol{v}_1 \in \mathbb{R}^2$ be an eigenvector of $\boldsymbol{X}$. Choose another vector $\boldsymbol{v}_2$ so that $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a basis of $\mathbb{R}^2$. Then $\boldsymbol{X}\boldsymbol{v}_1 = \boldsymbol{v}_1$ and $\boldsymbol{X}\boldsymbol{v}_2 = \alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2$. We have $\mathrm{tr}(\boldsymbol{X}) = 2$ implies $\beta = 1$, so that $\boldsymbol{X}\boldsymbol{v}_2 = \alpha\boldsymbol{v}_1 + \boldsymbol{v}_2$. Note that $\alpha \neq 0$ as $\boldsymbol{v}_2$ is not an eigenvector. We pick the ordered basis either $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ or $\{\boldsymbol{v}_1, -\boldsymbol{v}_2\}$ which has positive determinant d. Using that ordered basis, we may scale the matrix of change of basis by a factor of $\sqrt{d}$, so that $\boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix}$, and $\det(\boldsymbol{P}) = 1$.

Thus $\boldsymbol{X}$ is conjugate to either $\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix}$. Without loss of generality, take $\alpha > 0$.

Then, we have $\begin{pmatrix} \sqrt{\alpha} & 0 \\ 0 & \frac{1}{\sqrt{\alpha}} \end{pmatrix}^{-1}\begin{pmatrix} 1 & \pm\alpha \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \sqrt{\alpha} & 0 \\ 0 & \frac{1}{\sqrt{\alpha}} \end{pmatrix} = \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}$.

Also one can check that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ are not conjugate in $SL(2,\mathbb{R})$.

Consequently, $\boldsymbol{X}$ is conjugate to either $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.

Case 4. $\mathrm{tr}(\boldsymbol{X}) = -2$.

The eigenvalues are -1 with multiplicity 2.

(4i) $\boldsymbol{X}$ has 2 linearly independent eigenvectors. Then $\boldsymbol{X} = -\boldsymbol{I}$.

(4ii) $\boldsymbol{X}$ does not have 2 linearly independent eigenvectors. Let $\boldsymbol{v}_1 \in \mathbb{R}^2$ be an eigenvector of $\boldsymbol{X}$. Choose another vector $\boldsymbol{v}_2$ so that $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a basis of $\mathbb{R}^2$. Then $\boldsymbol{X}\boldsymbol{v}_1 = -\boldsymbol{v}_1$ and $\boldsymbol{X}\boldsymbol{v}_2 = \alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2$. We have $\mathrm{tr}(\boldsymbol{X}) = -2$ implies $\beta = -1$, so that $\boldsymbol{X}\boldsymbol{v}_2 = \alpha\boldsymbol{v}_1 - \boldsymbol{v}_2$. Note that $\alpha \neq 0$ as $\boldsymbol{v}_2$ is not an eigenvector. We pick the ordered basis either $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ or $\{\boldsymbol{v}_1, -\boldsymbol{v}_2\}$ which has positive determinant d. Using that ordered basis, we may scale the

matrix of change of basis by a factor of $\sqrt{d}$, so that $\boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P} = \begin{pmatrix} -1 & \alpha \\ 0 & -1 \end{pmatrix}$ or $\begin{pmatrix} -1 & -\alpha \\ 0 & -1 \end{pmatrix}$, and $\det(\boldsymbol{P}) = 1$.

Thus $\boldsymbol{X}$ is conjugate to either $\begin{pmatrix} -1 & \alpha \\ 0 & -1 \end{pmatrix}$ or $\begin{pmatrix} -1 & -\alpha \\ 0 & -1 \end{pmatrix}$. Without loss of generality, take $\alpha > 0$.

Then, we have $\begin{pmatrix} \sqrt{\alpha} & 0 \\ 0 & \frac{1}{\sqrt{\alpha}} \end{pmatrix}^{-1} \begin{pmatrix} -1 & \pm\alpha \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & 0 \\ 0 & \frac{1}{\sqrt{\alpha}} \end{pmatrix} = \begin{pmatrix} -1 & \pm 1 \\ 0 & -1 \end{pmatrix}$.

Also one can check that $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ and $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$ are not conjugate in $SL(2, \mathbb{R})$.

Consequently, $\boldsymbol{X}$ is conjugate to either $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ or $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$.

Theorem 7. *Let $\boldsymbol{X} \in SL(2, \mathbb{R})$. Then $\boldsymbol{X}$ is conjugate in $SL(2, \mathbb{R})$ to exactly one of the following matrices.*

$\boldsymbol{I}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$, $-\boldsymbol{I}$, $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ *where* $|\alpha| > 1$, $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ *where* $\theta \in (0, \pi) \cup (\pi, 2\pi)$.

Example 4.1. *By exercise 5.6,* $\begin{pmatrix} -1 & \pm 1 \\ 0 & -1 \end{pmatrix}$ *does not have a square root in* $SL(2, \mathbb{R})$. *In fact it has no square root in* $SL(2, \mathbb{C})$. *Direct calculation shows that* $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$, $\alpha < -1$ *does not have a square root in* $SL(2, \mathbb{R})$.

The rest of the matrices in theroem 7 have square roots. For instance, $-\boldsymbol{I} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2$.

4.2 Trace Identities

Proposition 4.1. *Let $\boldsymbol{A}, \boldsymbol{B} \in SL(2, \mathbb{R})$. Then* $\mathrm{tr}(\boldsymbol{AB}) + \mathrm{tr}(\boldsymbol{AB}^{-1}) = \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})$.

Proof. By the Cayley-Hamilton theorem, $\boldsymbol{B}^2 - \mathrm{tr}(\boldsymbol{B})\boldsymbol{B} + \boldsymbol{I} = \boldsymbol{O}$. Multiplying both sides by $\boldsymbol{B}^{-1}$, we have $\boldsymbol{B} - \mathrm{tr}(\boldsymbol{B})\boldsymbol{I} + \boldsymbol{B}^{-1} = \boldsymbol{O}$. That is $\boldsymbol{B} + \boldsymbol{B}^{-1} = \mathrm{tr}(\boldsymbol{B})\boldsymbol{I}$.
Multiplying both sides by $\boldsymbol{A}$, we have $\boldsymbol{AB} + \boldsymbol{AB}^{-1} = \mathrm{tr}(\boldsymbol{B})\boldsymbol{A}$. Taking the trace, we get $\mathrm{tr}(\boldsymbol{AB}) + \mathrm{tr}(\boldsymbol{AB}^{-1}) = \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})$.

Remark. Letting $\boldsymbol{A} = \boldsymbol{B}$, we obtain the identity $\mathrm{tr}(\boldsymbol{A}^2) + 2 = \mathrm{tr}(\boldsymbol{A})^2$.

Theorem 8. [The Fricke Identity] *Let $\boldsymbol{A}, \boldsymbol{B} \in SL(2, \mathbb{R})$. Then*

$$\mathrm{tr}(\boldsymbol{A})^2 + \mathrm{tr}(\boldsymbol{B})^2 + \mathrm{tr}(\boldsymbol{AB})^2 = \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})\mathrm{tr}(\boldsymbol{AB}) + \mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) + 2.$$

Proof. Apply proposition 4.1 to $\boldsymbol{ABA}^{-1}$ and $\boldsymbol{B}$, we have

$$\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = \mathrm{tr}(\boldsymbol{B})^2.$$

Apply proposition 4.1 to $\boldsymbol{AB}$ and $\boldsymbol{A}^{-1}\boldsymbol{B}$, we have

$$\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{A}^2) = \mathrm{tr}(\boldsymbol{AB})\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{B}).$$

Since $\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{B}) = -\mathrm{tr}(\boldsymbol{AB}) + \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})$, we have

$$\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{A}^2) = \mathrm{tr}(\boldsymbol{AB})\left(-\mathrm{tr}(\boldsymbol{AB}) + \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})\right).$$

That is

$$\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{A}^2) = -\mathrm{tr}(\boldsymbol{AB})^2 + \mathrm{tr}(\boldsymbol{AB})\mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}).$$

Using the first equation, we get

$$-\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) + \mathrm{tr}(\boldsymbol{B})^2 + \mathrm{tr}(\boldsymbol{A}^2) = -\mathrm{tr}(\boldsymbol{AB})^2 + \mathrm{tr}(\boldsymbol{AB})\mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}).$$

Using $\mathrm{tr}(\boldsymbol{A}^2) = \mathrm{tr}(\boldsymbol{A})^2 - 2$, we obtain the final result.

Let $d \in \mathbb{R}$. Consider the Markov equation $x^2 + y^2 + z^2 - xyz = d$. Treating this as a quadratic equation in z, we have $z = \frac{1}{2}\left(xy \pm \sqrt{x^2y^2 - 4(x^2 + y^2 - d)}\right)$.
When x, y are large enough, the discriminant is positive so that z is real.
Let's choose $x, y > 2$ and the root $z = \frac{1}{2}\left(xy - \sqrt{x^2y^2 - 4(x^2 + y^2 - d)}\right)$. Then $xy - z = \frac{1}{2}\left(xy + \sqrt{x^2y^2 - 4(x^2 + y^2 - d)}\right) > \frac{1}{2}xy > 2$. Therefore, there exists a point (x, y, z) in the surface S satisfying $xy - z > 2$.

Proposition 4.2. *Let $x, y, z \in \mathbb{R}$ such that $xy - z > 2$. There exist $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $\mathrm{tr}(\boldsymbol{A}) = x, \mathrm{tr}(\boldsymbol{B}) = y$, $\mathrm{tr}(\boldsymbol{AB}) = z$.*

Proof. The quadratic equation $X^2 - (xy - z)X + 1 = 0$ has 2 real roots as the discriminant $(xy - z)^2 - 4 > 0$. Let the roots be γ and $\frac{1}{\gamma}$. Note that $\gamma \neq 0$. Then $\gamma + \frac{1}{\gamma} - (xy - z) = 0$.
Let $\boldsymbol{A} = \begin{pmatrix} 0 & \gamma \\ -\frac{1}{\gamma} & x \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & y \end{pmatrix}$.
Then $\mathrm{tr}(\boldsymbol{A}) = x, \mathrm{tr}(\boldsymbol{B}) = y$, $\mathrm{tr}(\boldsymbol{AB}) = xy - \gamma - \frac{1}{\gamma} = z$.

Example 4.2. Let's find (x, y, z) satisfying $x^2 + y^2 + z^2 - xyz = 2 + \alpha + \frac{1}{\alpha}$, where $|\alpha| > 1$, such that $xy - z > 2$.

(i) For $\alpha > 1$, take $x, y = 0, z = -\sqrt{\alpha} - \frac{1}{\sqrt{\alpha}}$. Then $xy - z = \sqrt{\alpha} + \frac{1}{\sqrt{\alpha}} > 2$.

(ii) For $\alpha < -1$, let $\beta = -\alpha$ so that $\beta > 1$ and let $c = \sqrt{\beta} + \frac{1}{\sqrt{\beta}} > 2$. Take $x = y = \sqrt{2c + 4}, z = c + 2$. Then
$x^2 + y^2 + z^2 - xyz = 2(2c+4) + (c+2)^2 - (2c+4)(c+2) = 4 - c^2 = 4 - (\sqrt{\beta} + \frac{1}{\sqrt{\beta}})^2 = 2 - \beta - \frac{1}{\beta} = 2 + \alpha + \frac{1}{\alpha}$. Also $xy - z = 2c + 4 - c - 2 = c + 2 > 4 > 2$.

Example 4.3. Let's find $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = \alpha + \frac{1}{\alpha}$, where $|\alpha| > 1$.

(i) For $\alpha > 1$, let $\boldsymbol{A} = \begin{pmatrix} 0 & \sqrt{\alpha} \\ -\frac{1}{\sqrt{\alpha}} & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.
Then $\mathrm{tr}(\boldsymbol{A}) = 0, \mathrm{tr}(\boldsymbol{B}) = 0$ and $\mathrm{tr}(\boldsymbol{AB}) = -\sqrt{\alpha} - \frac{1}{\sqrt{\alpha}}$. By the Fricke Identity, we have $\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = \alpha + \frac{1}{\alpha}$.

(ii) For $\alpha < -1$, we take the solution of the Markov equation in example 4.2(ii). Let $c = \sqrt{-\alpha} + \frac{1}{\sqrt{-\alpha}} > 2$. Take $x = y = \sqrt{2c+4}, z = c+2$. Then $xy - z = c + 2 > 2$. By proposition 4.2, let $\gamma \in \mathbb{R}$ satisfying $\gamma + \frac{1}{\gamma} - (c+2) = 0$.

Let $\boldsymbol{A} = \begin{pmatrix} 0 & \gamma \\ -\frac{1}{\gamma} & \sqrt{2c+4} \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & \sqrt{2c+4} \end{pmatrix}$. Then $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B}) = \sqrt{2c+4}$ and $\mathrm{tr}(\boldsymbol{AB}) = c + 2$. By the Fricke Identity, we have $\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = \alpha + \frac{1}{\alpha}$.

Example 4.4. Let's find $\boldsymbol{A}, \boldsymbol{B} \in SL(2, \mathbb{R})$ such that $\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = 2\cos\theta$.

The equation $x^2 + y^2 + z^2 - xyz = 2 + 2\cos\theta$ has a solution $x = 3, y = 3, z = \frac{1}{2}(9 - \sqrt{17 + 8\cos\theta})$.
Note $xy - z = 9 - \frac{1}{2}(9 - \sqrt{17 + 8\cos\theta}) = \frac{1}{2}(9 + \sqrt{17 + 8\cos\theta}) \geq \frac{1}{2}(9 + \sqrt{17 - 8}) = 4 > 2$.
By proposition 4.2, let $\gamma \in \mathbb{R}$ satisfying $\gamma + \frac{1}{\gamma} - \frac{1}{2}(9 + \sqrt{17 + 8\cos\theta}) = 0$.

Let $\boldsymbol{A} = \begin{pmatrix} 0 & \gamma \\ -\frac{1}{\gamma} & 3 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & 3 \end{pmatrix}$. Then $\mathrm{tr}(\boldsymbol{AB}) = \frac{1}{2}(9 - \sqrt{17 + 8\cos\theta})$. By the Fricke Identity, we have $\mathrm{tr}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}) = 2\cos\theta$.

Theorem 9. *Let $\boldsymbol{X} \in SL(2, \mathbb{R})$. If $\boldsymbol{X} \neq -\boldsymbol{I}$, then $\boldsymbol{X}$ is a single multiplicative commutator. $-\boldsymbol{I}$ is a product of 2 multiplicative commutators.*

Proof. It suffices to show each matrix except $-\boldsymbol{I}$ in theorem 7 is a single multiplicative commutator.
(i) If $\mathrm{tr}(\boldsymbol{X}) = 2$, then $\boldsymbol{X}$ is conjugate in $SL(2, \mathbb{R})$ to the either $\boldsymbol{I}$, $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.
Certainly, $\boldsymbol{I}$ is a single multiplicative commutator. Let's show that the other 2 matrices can also be expressed as a single multiplicative commutator in $SL(2, \mathbb{R})$.

Take the solution $(3, 3, 2)$ of the Markov equation $x^2 + y^2 + z^2 - xyz = 4$. Let γ satisfy $\gamma + \frac{1}{\gamma} = 3 - 2 = 7$. That is $\gamma = \frac{1}{2}(7 \pm 3\sqrt{5})$, so there are 2 choices of γ. If we take $\gamma = \frac{1}{2}(7 - 3\sqrt{5})$ and apply proposition 4.2, we get the matrices $\boldsymbol{A} = \begin{pmatrix} 0 & \frac{7-3\sqrt{5}}{2} \\ -\frac{7+3\sqrt{5}}{2} & 3 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & 3 \end{pmatrix}$. Then $\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1} = \begin{pmatrix} 1 + 3\sqrt{5} & \frac{15}{2} - \frac{9\sqrt{5}}{2} \\ \frac{15}{2} + \frac{9\sqrt{5}}{2} & 1 - 3\sqrt{5} \end{pmatrix}$ whose Jordan form is $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Thus there exists $\boldsymbol{J} = \begin{pmatrix} 3 - \sqrt{5} & -\frac{1}{3} + \frac{\sqrt{5}}{5} \\ 2 & 0 \end{pmatrix}$ such that $\boldsymbol{J}^{-1}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1})\boldsymbol{J} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Though $\det\boldsymbol{J} < 0$, the matrices $\boldsymbol{J}^{-1}\boldsymbol{AJ}$ and $\boldsymbol{J}^{-1}\boldsymbol{BJ}$ still have determinant 1.

Then $\boldsymbol{J}^{-1}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1})\boldsymbol{J}$ still expresses $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ as a single multiplicative commutator in $SL(2,\mathbb{R})$.

$$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{5+3\sqrt{5}}{30}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{5-3\sqrt{5}}{30}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{5+3\sqrt{5}}{30}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}^{-1}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{5-3\sqrt{5}}{30}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}^{-1}.$$

However, if we prefer $\boldsymbol{J}$ to have positive determinant, we may change the second column of $\boldsymbol{J}$ to its negative. The resulting matrix has positive determinant. Denote that matrix by $\boldsymbol{P}$. [This corresponds to use the new ordered basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2'\}$ where $\boldsymbol{v}_2' = -\boldsymbol{v}_2$. Thus if we denote $\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}$ by $\boldsymbol{X}$, then $\boldsymbol{X}\boldsymbol{v}_1 = \boldsymbol{v}_1, \boldsymbol{X}\boldsymbol{v}_2' = -\boldsymbol{X}\boldsymbol{v}_2 = -(\boldsymbol{v}_1 + \boldsymbol{v}_2) = -\boldsymbol{v}_1 + \boldsymbol{v}_2'$]. Then $\boldsymbol{P}^{-1}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1})\boldsymbol{P} = \begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix}$. More explicitly, if we work out $\boldsymbol{P}^{-1}\boldsymbol{AP}$ and $\boldsymbol{P}^{-1}\boldsymbol{BP}$ we get

$$\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}\frac{3-\sqrt{5}}{2} & \frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & -\frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3-\sqrt{5}}{2} & \frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}^{-1}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & -\frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}^{-1}.$$

Conjugating both sides by $\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$, we get back

$$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{1}{6}-\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{1}{6}-\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{1}{6}-\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}^{-1}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{1}{6}-\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}^{-1}.$$

Alternatively, if we use the root $\gamma = \frac{1}{2}(7+3\sqrt{5})$ and apply proposition 4.2, we get

$$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}\begin{pmatrix}\frac{3+\sqrt{5}}{2} & \frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3-\sqrt{5}}{2}\end{pmatrix}^{-1}\begin{pmatrix}\frac{3-\sqrt{5}}{2} & -\frac{1}{6}+\frac{\sqrt{5}}{10}\\ 0 & \frac{3+\sqrt{5}}{2}\end{pmatrix}^{-1}.$$

(ii) If $\mathrm{tr}(\boldsymbol{X}) = -2$, then $\boldsymbol{X}$ is conjugate in $SL(2,\mathbb{R})$ to the either $-\boldsymbol{I}$, $\begin{pmatrix}-1 & 1\\ 0 & -1\end{pmatrix}$ or $\begin{pmatrix}-1 & -1\\ 0 & -1\end{pmatrix}$. By exercise 5.9 and 5.10, $-\boldsymbol{I}$ cannot be expressed as a single multiplicative commutator in $SL(2,\mathbb{R})$. Instead, it can be expressed as a product of two multiplicative commutators in $SL(2,\mathbb{R})$. Let's show that the other 2 matrices can also be expressed as a single multiplicative commutator in $SL(2,\mathbb{R})$.
Take the solution $(3,3,3)$ of the Markov equation $x^2+y^2+z^2-xyz=0$. Let γ satisfy $\gamma + \frac{1}{\gamma} = 3\times 3 - 3 = 6$. That is $\gamma = \frac{1}{2}(6 \pm \sqrt{32}) = 3 \pm 2\sqrt{2}$. Take $\gamma = 3+2\sqrt{2}$. Applying proposition 4.2, we get the matrices $\boldsymbol{A} = \begin{pmatrix}0 & 3+2\sqrt{2}\\ -3+2\sqrt{2} & 3\end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix}0 & 1\\ -1 & 3\end{pmatrix}$. Then $\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1} = \begin{pmatrix}-1-6\sqrt{2} & 6+6\sqrt{2}\\ 12-12\sqrt{2} & -1+6\sqrt{2}\end{pmatrix}$. In this case, we are certain that $\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1} \neq -\boldsymbol{I}$ as $-\boldsymbol{I}$ cannot be a single multiplicative commutator. Therefore the Jordan form of $\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1}$ must be $\begin{pmatrix}-1 & 1\\ 0 & -1\end{pmatrix}$. Let $\boldsymbol{P} = \begin{pmatrix}-6\sqrt{2} & 1\\ 12-12\sqrt{2} & 0\end{pmatrix}$. Then $\boldsymbol{P}^{-1}(\boldsymbol{ABA}^{-1}\boldsymbol{B}^{-1})\boldsymbol{P} = \begin{pmatrix}-1 & 1\\ 0 & -1\end{pmatrix}$. More explicitly, if we work out $\boldsymbol{P}^{-1}\boldsymbol{AP}$ and

$\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$ we get

$$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2+\frac{\sqrt{2}}{2} & \frac{-1+\sqrt{2}}{12} \\ -6 & 1-\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2-\frac{\sqrt{2}}{2} & \frac{1+\sqrt{2}}{12} \\ 6 & 1+\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2+\frac{\sqrt{2}}{2} & \frac{-1+\sqrt{2}}{12} \\ -6 & 1-\frac{\sqrt{2}}{2} \end{pmatrix}^{-1} \begin{pmatrix} 2-\frac{\sqrt{2}}{2} & \frac{1+\sqrt{2}}{12} \\ 6 & 1+\frac{\sqrt{2}}{2} \end{pmatrix}^{-1}.$$

Conjugating both sides by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, we have

$$\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2+\frac{\sqrt{2}}{2} & \frac{1-\sqrt{2}}{12} \\ 6 & 1-\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2-\frac{\sqrt{2}}{2} & -\frac{1+\sqrt{2}}{12} \\ -6 & 1+\frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2+\frac{\sqrt{2}}{2} & \frac{1-\sqrt{2}}{12} \\ 6 & 1-\frac{\sqrt{2}}{2} \end{pmatrix}^{-1} \begin{pmatrix} 2-\frac{\sqrt{2}}{2} & -\frac{1+\sqrt{2}}{12} \\ -6 & 1+\frac{\sqrt{2}}{2} \end{pmatrix}^{-1}.$$

(iii) If $\mathrm{tr}(\boldsymbol{X}) > 2$, then $\boldsymbol{X}$ is conjugate in $SL(2,\mathbb{R})$ to $\begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}$ with $|\alpha| > 1$. By example 4.3, there exist $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}) = \mathrm{tr}(\boldsymbol{X}) = \alpha + \frac{1}{\alpha}$. In this case, the conjugacy class of $\boldsymbol{X}$ is unique. Thus $\boldsymbol{X} = \boldsymbol{P}^{-1}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1})\boldsymbol{P}$ for some $\boldsymbol{P} \in SL(2,\mathbb{R})$. Therefore $\boldsymbol{X}$ can be expressed as a single multiplicative commutator in $SL(2,\mathbb{R})$.

(iv) If $\mathrm{tr}(\boldsymbol{X}) < 2$, then $\boldsymbol{X}$ is conjugate in $SL(2,\mathbb{R})$ to $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ where $\theta \in (0,\pi) \cup (\pi, 2\pi)$. By example 4.4, there exist $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}) = \mathrm{tr}(\boldsymbol{X}) = 2\cos\theta$. In this case, the conjugacy class of $\boldsymbol{X}$ is unique. Therefore $\boldsymbol{X} = \boldsymbol{P}^{-1}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1})\boldsymbol{P}$ for some $\boldsymbol{P} \in SL(2,\mathbb{R})$. Therefore $\boldsymbol{X}$ can be expressed as a single multiplicative commutator in $SL(2,\mathbb{R})$.

Remark. In proposition 4.2,

$$\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1} = \begin{pmatrix} \gamma^2 - y(\gamma x - y) & \gamma x - y \\ \gamma x - x^2 y + \frac{xy^2}{\gamma} - \frac{y}{\gamma^2} & x^2 - \frac{xy}{\gamma} + \frac{1}{\gamma^2} \end{pmatrix}.$$

If it is the identity matrix, then $\gamma^2 = 1$ and $y = \gamma x$. That is $\gamma = \pm 1$ and $y = \pm x$. To ensure we do not get $\boldsymbol{I}$, we simply pick the solution (x, y, z) of the Markov equation or γ that do not satisfy these conditions.

Exercise 1

1.1. What is the dimension of the vector space $\mathbb{C}$ over $\mathbb{R}$?

1.2. Let V be a vector space over $\mathbb{R}$. Let V_1 and V_2 be subspaces of V such that $V_1 \subsetneq V_2$ and $V_2 \subsetneq V_1$. Is $V_1 \cup V_2$ a subspace of V?

1.3. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Is it true that $\boldsymbol{AB} = \boldsymbol{O}_n$ implies $\boldsymbol{BA} = \boldsymbol{O}_n$?

1.4. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Is it true that $\boldsymbol{AB} = \boldsymbol{I}_n$ implies $\boldsymbol{BA} = \boldsymbol{I}_n$?

1.5. Let $\boldsymbol{A}$ be a 2×2 matrix over $\mathbb{C}$. The trace of $\boldsymbol{A}$, denoted by $\mathrm{tr}(\boldsymbol{A})$, is the sum of the diagonal entries of $\boldsymbol{A}$. Suppose $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}^2) = 0$. Is it true that $\boldsymbol{A} = \boldsymbol{O}_2$?

1.6. Prove that every 2×2 matrix of determinant 1 is the product of 3 elementary matrices.

1.7. Show that matrices $\boldsymbol{A} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}$ are similar but not unitarily similar over $\mathbb{C}$.

1.8. Let $\boldsymbol{A}$ be an $n \times n$ positive definite hermitian matrix which is unitary. Show that $\boldsymbol{A} = \boldsymbol{I}_n$.

1.9. Let $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & a \\ 0 & b & 0 \\ c & 0 & 0 \end{pmatrix}$ be in $\mathcal{M}_{n\times n}(\mathbb{C})$. Prove that $\boldsymbol{A}$ is diagonalizable if and only if either $ac \neq 0$ or $a = c = 0$.

1.10. For any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and any permutation σ of $1, 2, \ldots, n$, define $\sigma(\boldsymbol{x}) = \begin{pmatrix} x_{\sigma(1)} \\ \vdots \\ x_{\sigma(n)} \end{pmatrix}$. Fix a vector $\boldsymbol{x} \in \mathbb{R}^n$. Let V be the span of $\{\sigma(\boldsymbol{x}) \mid \sigma \text{ is a permutation of } 1, 2, \ldots, n\}$. What are the possibilities for the dimension of V?

Exercise 2

2.1. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\boldsymbol{AB} = \boldsymbol{I}_n$. Show that the system of linear equations $\boldsymbol{BX} - x\boldsymbol{I}_n = \boldsymbol{0}_n$ has a nontrivial solution in the unknowns $\boldsymbol{X}$ and x. Hence show that $\boldsymbol{BA} = \boldsymbol{I}_n$.

2.2. Let $\boldsymbol{A}, \boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3, \boldsymbol{B}_4$ be 2×2 real matrices and $\boldsymbol{A} \neq \boldsymbol{O}_2$ such that

$$\det(\boldsymbol{A} + \boldsymbol{B}_i) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}_i), \quad \text{for } i = 1, 2, 3, 4.$$

Prove that there exist real numbers c_1, c_2, c_3, c_4, not all zero, such that

$$c_1\boldsymbol{B}_1 + c_2\boldsymbol{B}_2 + c_3\boldsymbol{B}_3 + c_4\boldsymbol{B}_4 = \boldsymbol{O}_2.$$

2.3. Let $f_1, f_2, \ldots, f_n$ be linearly independent functions defined on $\mathbb{R}$. Prove that some $n-1$ of their derivatives $f_1', f_2', \ldots, f_n'$ are linearly independent.

2.4. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same characteristic polynomial.

2.5. Let $\boldsymbol{A}$ be an $m \times n$ matrix over $\mathbb{C}$. Prove that $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ have the same set of nonzero eigenvalues.

2.6. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ are similar.

2.7. Let $\boldsymbol{A} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where $a \neq b$. Prove that $\boldsymbol{A}$ is similar to the matrix $\begin{pmatrix} a+b & ab \\ -1 & 0 \end{pmatrix}$.

2.8. Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

(a) Suppose $a_{12} \neq 0$. Let $\boldsymbol{P} = \begin{pmatrix} 0 & -a_{12} \\ 1 & a_{11} \end{pmatrix}$. Show that $\boldsymbol{P}$ is invertible and

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}.$$

(b) Suppose $a_{21} \neq 0$. Find an invertible matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{Q}^{-1}\boldsymbol{AQ} = \begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}.$$

2.9. Let $\boldsymbol{A} = \begin{pmatrix} 9 & -40 \\ 40 & 9 \end{pmatrix}$. Find a positive definite hermitian matrice $\boldsymbol{R}$ and a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{A} = \boldsymbol{RU}$.

2.10. Let $\boldsymbol{U} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be an upper triangular unitary matrix. Prove that $\boldsymbol{U}$ is diagonal.

Exercise 3

3.1. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $f(x)$ be a polynomial with complex coefficients such that $f(\boldsymbol{A}\boldsymbol{A}^*) = \boldsymbol{O}$. Prove that $f(\boldsymbol{A}^*\boldsymbol{A}) = \boldsymbol{O}$.

3.2. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a normal matrix and V an eigenspace of $\boldsymbol{A}^*$. Prove that V and $V^\perp$ are subspaces of $\mathbb{C}^n$ invariant under $\boldsymbol{A}$.

3.3. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\boldsymbol{A} \neq \boldsymbol{O}_n$. Prove that there exists a 1-dimensional or a 2-dimensional subspace of $\mathbb{R}^n$ invariant under $\boldsymbol{A}$.

3.4. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $\boldsymbol{A}^2 = \boldsymbol{A}$. Prove that $\mathrm{rank}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$.

3.5. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be such that $\boldsymbol{A}^2 = \boldsymbol{A}$, $\boldsymbol{B}^2 = \boldsymbol{B}$ and $\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{A}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A} = \boldsymbol{O}_n$. What can be said about $\boldsymbol{A}$ and $\boldsymbol{B}$?

3.6. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Prove that $\boldsymbol{A}$ is symmetrically similar to $\boldsymbol{A}^\top$ over $\mathbb{R}$ if and only if $\boldsymbol{A}$ is a product of two real symmetric matrices and one of them is invertible.

3.7. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Prove that $\boldsymbol{A}$ is diagonalizable over $\mathbb{R}$ with all its eigenvalues positive if and only if $\boldsymbol{A}$ is a product of two real positive definite symmetric matrices.

3.8. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a normal matrix. Prove that the null space of $\boldsymbol{A}$ is the same as the null space of $\boldsymbol{A}^*$.

3.9. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}$ is a normal matrix. Prove that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{O}_n$ implies $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{O}_n$.

3.10. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}$. Prove that $\boldsymbol{A}$ is normal.

Exercise 4

4.1. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}$ is a normal matrix. Prove that $\boldsymbol{AB} = \boldsymbol{BA}$ implies $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{BA}^*$.

4.2. Let $\boldsymbol{A}, \boldsymbol{B}$ be positive definite hermitian matrices. Prove that $\det(\boldsymbol{I} + \boldsymbol{AB}) > 1$.

4.3. Let $\boldsymbol{A} = \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 3 & 7 \\ 7 & 18 \end{pmatrix}$ and $\boldsymbol{C} = \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{3}{5} \end{pmatrix}$. Show that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are positive definite hermitian matrices. Is it true that $\det(\boldsymbol{I} + \boldsymbol{ABC}) > 1$?

4.4. Show that the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ cannot be expressed as a product of two singular symmetric matrices.

4.5. Let $\boldsymbol{Q}$ and $\boldsymbol{M}$ be symmetric matrices with $\boldsymbol{Q}$ positive definite. Show there is a matrix $\boldsymbol{A}$ such that $\boldsymbol{AQ} + \boldsymbol{QA} = \boldsymbol{M}$.

4.6. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be Hermitian matrices such that $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{B}^2 = \boldsymbol{B}$. Prove that if $\boldsymbol{AB}$ is normal, then $\boldsymbol{AB} = \boldsymbol{BA}$.

4.7. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ such that $\boldsymbol{A}^2 = \boldsymbol{A}^*$. Prove that $\operatorname{rank}(\boldsymbol{A} + \boldsymbol{A}^*) = \operatorname{rank}(\boldsymbol{A})$ and $\boldsymbol{I} + \boldsymbol{A}$ is non-singular. [AMM problem 1793]

4.8. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 matrices. Prove that $\det(\boldsymbol{A} + \boldsymbol{B}) + \operatorname{tr}(\boldsymbol{AB}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}) + \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$.

4.9. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 nonzero matrices such that $\boldsymbol{AB}+\boldsymbol{BA} = \boldsymbol{O}$ and $\det(\boldsymbol{A}+\boldsymbol{B}) = 0$. Prove that $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{B}) = 0$.

4.10. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 matrices such that $\boldsymbol{A}^2 + \boldsymbol{B}^2 = 2\boldsymbol{AB}$. Prove that $\boldsymbol{AB} = \boldsymbol{BA}$ and $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{B})$. [Romanian Mathematical Competitions 2011]

Exercise 5

5.1. Show that the matrix $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ is a product of two symmetric matrices.

5.2. Prove that any matrix is a product of two symmetric matrices and one of them is non-singular.

5.3. Prove that any matrix $\boldsymbol{A}$ is similar to its transpose $\boldsymbol{A}^\top$.

5.4. Let $\boldsymbol{X} \in SL(2,\mathbb{C})$. Prove that $\boldsymbol{X}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}$ for some $\boldsymbol{P} \in SL(2,\mathbb{C})$.

5.5. Let $\boldsymbol{A} \in SL(2,\mathbb{C})$. Prove that $\operatorname{tr}(\boldsymbol{A}^2) = (\operatorname{tr}(\boldsymbol{A}))^2 - 2$.

5.6. Let $\boldsymbol{A} \in SL(2,\mathbb{C})$. Prove that if $\operatorname{tr}(\boldsymbol{A}^2) = -2$, then $\boldsymbol{A}^2 = -\boldsymbol{I}$.

5.7. Let $\boldsymbol{X} \in SL(2,\mathbb{C})$. Prove that if $\boldsymbol{X}$ has a square root $\boldsymbol{A} \in SL(2,\mathbb{C})$, then $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{B} \in SL(2,\mathbb{C})$.

5.8. Let $\boldsymbol{X} \in SL(2,\mathbb{C})$. Prove that $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{C})$.

5.9. Do there exist $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $-\boldsymbol{I} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$?

5.10. Let $\boldsymbol{P} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 7 & 4 \end{pmatrix}$ and $\boldsymbol{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 3 & 1 \\ -2 & 0 \end{pmatrix}$. Show that $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ and $-\boldsymbol{I} = \boldsymbol{P}^2$.

Solutions to Exercises

1.1. What is the dimension of the vector space $\mathbb{C}$ over $\mathbb{R}$?

Solution. The dimension of $\mathbb{C}$ over $\mathbb{R}$ is 2. The elements 1 and i as vectors in $\mathbb{C}$ are linearly independent over $\mathbb{R}$. To see this, let $c_1 \cdot 1 + c_2 \cdot i = 0$ in $\mathbb{C}$, where $c_1, c_2 \in \mathbb{R}$. Thus $c_1 = c_2 = 0$. Clearly every complex number can be expressed in the form $c_1 \cdot 1 + c_2 \cdot i$ for some $c_1, c_2 \in \mathbb{R}$. Therefore $\{1, i\}$ is a basis of $\mathbb{C}$ over $\mathbb{R}$.

1.2. Let V be a vector space over $\mathbb{R}$. Let V_1 and V_2 be subspaces of V such that $V_1 \subsetneq V_2$ and $V_2 \subsetneq V_1$. Is $V_1 \cup V_2$ a subspace of V?

Solution. No, $V_1 \cup V_2$ is not a subspace of V. Let $\boldsymbol{v}_1 \in V_1 \setminus V_2$, and $\boldsymbol{v}_2 \in V_2 \setminus V_1$. Then $\boldsymbol{v}_1 + \boldsymbol{v}_2 \notin V_1$, otherwise $\boldsymbol{v}_2 = (\boldsymbol{v}_1 + \boldsymbol{v}_2) - \boldsymbol{v}_1 \in V_1$ which is not true. Similarly, $\boldsymbol{v}_1 + \boldsymbol{v}_2 \notin V_2$. Therefore, $\boldsymbol{v}_1 + \boldsymbol{v}_2 \notin V_1 \cup V_2$. That means $V_1 \cup V_2$ is not a subspace.

1.3. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Is it true that $\boldsymbol{AB} = \boldsymbol{O}_n$ implies $\boldsymbol{BA} = \boldsymbol{O}_n$?

Solution. No. For example let $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then $\boldsymbol{AB} = \boldsymbol{O}_n$, but $\boldsymbol{BA} = \boldsymbol{A} \neq \boldsymbol{O}_2$.

1.4. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Is it true that $\boldsymbol{AB} = \boldsymbol{I}_n$ implies $\boldsymbol{BA} = \boldsymbol{I}_n$?

Solution. Let $\boldsymbol{R}$ be the reduced row-echelon form of $\boldsymbol{A}$. Thus there exist elementary matrices $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_k$ such that $\boldsymbol{E}_1 \cdots \boldsymbol{E}_k \boldsymbol{A} = \boldsymbol{R}$. Suppose $\boldsymbol{AB} = \boldsymbol{I}$. Then $\boldsymbol{RB} = \boldsymbol{E}_1 \cdots \boldsymbol{E}_k \boldsymbol{AB} = \boldsymbol{E}_1 \cdots \boldsymbol{E}_k$. If $\boldsymbol{R}$ has a zero row, then $\boldsymbol{RB}$ has a zero row, and hence $\boldsymbol{E}_1 \cdots \boldsymbol{E}_k$ is singular which is a contradiction since an elementary matrix is invertible. Thus $\boldsymbol{R} = \boldsymbol{I}$ as $\boldsymbol{R}$ is in reduced row-echelon form. Consequently, $\boldsymbol{B} = \boldsymbol{E}_1 \cdots \boldsymbol{E}_k$. Therefore $\boldsymbol{BA} = \boldsymbol{E}_1 \cdots \boldsymbol{E}_k \boldsymbol{A} = \boldsymbol{R} = \boldsymbol{I}$.

1.5. Let $\boldsymbol{A}$ be a 2×2 matrix over $\mathbb{C}$. The trace of $\boldsymbol{A}$, denoted by $\operatorname{tr}(\boldsymbol{A})$, is the sum of the diagonal entries of $\boldsymbol{A}$. Suppose $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^2) = 0$. Is it true that $\boldsymbol{A} = \boldsymbol{O}_2$?

Solution. No. For example let $\boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then $\boldsymbol{A}$ satisfies the condition that $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^2) = 0$, but $\boldsymbol{A} \neq \boldsymbol{O}_2$. We can only conclude that $\boldsymbol{A}^2 = \boldsymbol{O}_2$ since 0 is the only eigenvalue of $\boldsymbol{A}$. Any nilpotent matrix $\boldsymbol{A}$ has the property that $\operatorname{tr}(\boldsymbol{A}^k) = 0$ for all $k \geq 1$.

1.6. Prove that every 2×2 matrix of determinant 1 is the product of 3 elementary matrices.

Solution. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix of determinat 1. Let's try to write

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}.$$

This is equivalent to

$$a = 1 + xy,\ b = z(1 + xy) + x,\ c = y,\ d = yz + 1.$$

If $c \neq 0$, we can solve for x, y, z using the first, third and fouth equations so that $y = c, x = (a-1)/c, z = (d-1)/c$. These values of x, y, z also satisfy the second

equation, since $z(1+xy)+x = c^{-1}(d-1)(1+a-1)+c^{-1}(a-1) = c^{-1}(ad-1) = b$, where we have used $ad - bc = 1$.

If $c = 0$, the matrix will be of the form

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & a^{-1} \end{pmatrix}.$$

1.7. Show that matrices $\boldsymbol{A} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}$ are similar but not unitarily similar over $\mathbb{C}$.

Solution. Note that $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of $\begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}$ with eigenvalue 1. Supplement it by another vector say $\begin{pmatrix} c1 \\ 0 \end{pmatrix}$ to form a basis of $\mathbb{R}^2$. We get the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Then we can verify $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$. Thus $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar.

Suppose there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{B}$. Taking conjugate transpose, we obtain $\boldsymbol{U}^*\boldsymbol{A}^*\boldsymbol{U} = \boldsymbol{B}^*$. Multiplying these two together, we get $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{U} = \boldsymbol{B}\boldsymbol{B}^*$. Taking the trace on both sides, we get $\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^*) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{B}^*)$. Direct computation gives $\boldsymbol{A}\boldsymbol{A}^* = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$ and $\boldsymbol{B}\boldsymbol{B}^* = \begin{pmatrix} 5 & 8 \\ 8 & 13 \end{pmatrix}$. Thus $\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^*) = 6$ but $\operatorname{tr}(\boldsymbol{B}\boldsymbol{B}^*) = 18$. As they are not equal, $\boldsymbol{A}$ and $\boldsymbol{B}$ cannot be unitarily similar over $\mathbb{C}$.

1.8. Let $\boldsymbol{A}$ be an $n \times n$ positive definite hermitian matrix which is unitary. Show that $\boldsymbol{A} = \boldsymbol{I}_n$.

Solution. Since $\boldsymbol{A}$ hermitian, its eigenvalues are all real. Since $\boldsymbol{A}$ positive definite, its eigenvalues are positive. Since $\boldsymbol{A}$ unitary, its eigenvalues all have absolute value 1. So, the eigenvalues for $\boldsymbol{A}$ all must equal 1. Finally, since $\boldsymbol{A}$ is hermitian (and thus normal), we see that $\boldsymbol{A}$ is diagonalizable.

Therefore, $\boldsymbol{A}$ is similar to $\boldsymbol{I}_n$. This implies that $\boldsymbol{A} = \boldsymbol{I}_n$. (Since $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{I}_n\boldsymbol{P}^{-1}$ for some $\boldsymbol{P}$, this immediately shows that $\boldsymbol{A} = \boldsymbol{I}_n$.)

1.9. Let $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & a \\ 0 & b & 0 \\ c & 0 & 0 \end{pmatrix}$ be in $\mathcal{M}_{n\times n}(\mathbb{C})$. Prove that $\boldsymbol{A}$ is diagonalizable if and only if either $ac \neq 0$ or $a = c = 0$.

Solution. Consider the ordered basis $\mathcal{B} = \{\boldsymbol{e}_3, \boldsymbol{e}_1, \boldsymbol{e}_2\}$, where $\boldsymbol{e}_i$ is the ith canonical basis vector. We have $\boldsymbol{A}\boldsymbol{e}_3 = a\boldsymbol{e}_1, \boldsymbol{A}\boldsymbol{e}_1 = c\boldsymbol{e}_3, \boldsymbol{A}\boldsymbol{e}_2 = b\boldsymbol{e}_2$.

Using this basis $\mathcal{B}$, we see that $\boldsymbol{A}$ is similar to the matrix $\begin{pmatrix} 0 & c & 0 \\ a & 0 & 0 \\ 0 & 0 & b \end{pmatrix}$. It suffices to determine when the left upper 2×2 submatrix of this matrix is diagonalizable.

Let $\boldsymbol{B} = \begin{pmatrix} 0 & c \\ a & 0 \end{pmatrix}$. $\chi_{\boldsymbol{B}}(x) = x^2 - ac$ has a repeated root if and only if $a = 0$ or $c = 0$. Thus $\chi_{\boldsymbol{B}}(x) = x^2 - ac$ factors into distinct linear factors if and only if $ac \neq 0$. If $\boldsymbol{B}$ is

the zero matrix, then $\boldsymbol{B}$ is certainly diagonalizable and $a = c = 0$. If $\boldsymbol{B}$ is not the zero matrix, then not both a and c are zero. Thus $\boldsymbol{B}$ is diagonalizable if and only if $ac \neq 0$.

1.10. For any $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and any permutation σ of $1, 2, \ldots, n$, define $\sigma(\boldsymbol{x}) = \begin{pmatrix} x_{\sigma(1)} \\ \vdots \\ x_{\sigma(n)} \end{pmatrix}$. Fix a vector $\boldsymbol{x} \in \mathbb{R}^n$. Let V be the span of $\{\sigma(\boldsymbol{x}) \mid \sigma \text{ is a permutation of } 1, 2, \ldots, n\}$. What are the possibilities for the dimension of V?

Solution. The possibilities for the dimensions of V are $0, 1, n-1$ and n. First consider the case when all the coordinates of $\boldsymbol{x}$ are equal. Let $\boldsymbol{x} = \begin{pmatrix} x \\ \vdots \\ x \end{pmatrix}$. For any permutation σ, $\sigma(\boldsymbol{x}) = \boldsymbol{x}$. Thus the span of $\{\boldsymbol{x}\}$ has dimension 0 if $\boldsymbol{x}$ is the zero vector, and 1 if $\boldsymbol{x}$ is not the zero vector. Suppose not all coordinates of $\boldsymbol{x}$ are equal. Let s and t with $s \neq t$ be among the coordinates of $\boldsymbol{x}$. Then we can find permutations σ_1 and σ_2 such that $\sigma_1(\boldsymbol{x}) = \begin{pmatrix} s \\ t \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$ and $\sigma_2(\boldsymbol{x}) = \begin{pmatrix} t \\ s \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$. Therefore, $\frac{1}{t-s}(\sigma_1(\boldsymbol{x}) - \sigma_2(\boldsymbol{x})) = \begin{pmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ is in V. That is $\boldsymbol{e}_2 - \boldsymbol{e}_1 \in V$, where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are the canonical basis vectors of $\mathbb{R}^n$. Similarly, $\boldsymbol{e}_3 - \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n - \boldsymbol{e}_1 \in V$. It is easy to see that the vectors $\boldsymbol{e}_2 - \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n - \boldsymbol{e}_1$ are linearly independent so that $\dim(V) \geq n - 1$.

Lastly we can write

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n = (x_1 + x_2 + \cdots + x_n)\boldsymbol{e}_1 + x_2(\boldsymbol{e}_2 - \boldsymbol{e}_1) + \cdots + x_n(\boldsymbol{e}_n - \boldsymbol{e}_1).$$

This shows that if $x_1 + x_2 + \cdots + x_n = 0$, then $\boldsymbol{x}$ is in the span of $\boldsymbol{e}_2 - \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n - \boldsymbol{e}_1$. Similarly, each $\sigma(\boldsymbol{x})$ will be in this span, so V equals to this span and $\dim(V) = n - 1$.

On the other hand, if $x_1 + x_2 + \cdots + x_n \neq 0$, then the above equation shows that $\boldsymbol{e}_1 \in V$. Consequently, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \in V$ and $V = \mathbb{R}^n$ so that $\dim(V) = n$.

2.1. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\boldsymbol{AB} = \boldsymbol{I}_n$. Show that the system of linear equations $\boldsymbol{BX} - x\boldsymbol{I}_n = \boldsymbol{0}_n$ has a nontrivial solution in the unknowns $\boldsymbol{X}$ and x. Hence show that $\boldsymbol{BA} = \boldsymbol{I}_n$.

Solution. Consider $\boldsymbol{BX} - x\boldsymbol{I}_n = \boldsymbol{0}_n$ as a homogeneous system of n^2 equations with $n^2 + 1$ unknowns, namely the n^2 entries of $\boldsymbol{X}$ and x. Since the number of unknowns is more than the number of equations, it has a 'nontrivial' solution $\boldsymbol{X} = \boldsymbol{S}$ and $x = s$. If $s = 0$, then $\boldsymbol{BS} = \boldsymbol{0}_n$. Then $\boldsymbol{S} = \boldsymbol{I}_n\boldsymbol{S} = \boldsymbol{ABS} = \boldsymbol{A0}_n = \boldsymbol{0}_n$, a contradiction. Therefore $s \neq 0$. Then $\boldsymbol{B}(\boldsymbol{S}s^{-1}) = \boldsymbol{I}_n$. Thus $\boldsymbol{C} = \boldsymbol{S}s^{-1}$ is a right inverse of $\boldsymbol{B}$. Furthermore, $\boldsymbol{C} = \boldsymbol{I}_n\boldsymbol{C} = \boldsymbol{ABC} = \boldsymbol{AI}_n = \boldsymbol{A}$. Therefore $\boldsymbol{BA} = \boldsymbol{I}_n$.

2.2. Let $\boldsymbol{A}, \boldsymbol{B}_1, \boldsymbol{B}_2, \boldsymbol{B}_3, \boldsymbol{B}_4$ be 2×2 real matrices and $\boldsymbol{A} \neq \boldsymbol{O}_2$ such that

$$\det(\boldsymbol{A} + \boldsymbol{B}_i) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}_i), \quad \text{for } i = 1, 2, 3, 4.$$

Prove that there exist real numbers c_1, c_2, c_3, c_4, not all zero, such that

$$c_1\boldsymbol{B}_1 + c_2\boldsymbol{B}_2 + c_3\boldsymbol{B}_3 + c_4\boldsymbol{B}_4 = \boldsymbol{O}_2.$$

Solution. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\boldsymbol{B}_i = \begin{pmatrix} x_i & y_i \\ z_i & w_i \end{pmatrix}$, $i = 1, 2, 3, 4$. If $\det(\boldsymbol{A} + \boldsymbol{B}_i) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}_i)$, then by direct simplification, we obtain $dx_i - cy_i - bz_i + aw_i = 0$, $i = 1, 2, 3, 4$. Because $\boldsymbol{A} \neq \boldsymbol{O}_2$, the solution space to $dx - cy - bz + aw = 0$ is a 3-dimensional subspace of $\mathbb{R}^4$. Since any 4 vectors in $\mathbb{R}^4$ are linearly dependent, there must exist c_1, c_2, c_3, c_4, not all zero, such that

$$c_1\boldsymbol{B}_1 + c_2\boldsymbol{B}_2 + c_3\boldsymbol{B}_3 + c_4\boldsymbol{B}_4 = \boldsymbol{O}_2.$$

Here we identify $\mathcal{M}_{2\times 2}(\mathbb{R})$ with $\mathbb{R}^4$.

2.3. Let $f_1, f_2, \ldots, f_n$ be linearly independent functions defined on $\mathbb{R}$. Prove that some $n-1$ of their derivatives $f_1', f_2', \ldots, f_n'$ are linearly independent.

Solution. The linear transformation taking f to f' has a 1-dimensional null space consisting of all the constant functions. Thus the image of any n-dimensional space of differentiable functions must have dimension at least $n-1$.

2.4. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same characteristic polynomial.

Solution. Suppose $\boldsymbol{A}$ is invertible. Then $\det(x\boldsymbol{I} - \boldsymbol{AB}) = \det(\boldsymbol{A})\det(x\boldsymbol{A}^{-1} - \boldsymbol{B}) = \det(x\boldsymbol{A}^{-1} - \boldsymbol{B})\det(\boldsymbol{A}) = \det(x\boldsymbol{I} - \boldsymbol{BA})$. If $\boldsymbol{A}$ is not invertible, use continuity argument.

The equation $\det(t\boldsymbol{I} + \boldsymbol{A}) = 0$ has finitely many solutions for $t \in \mathbb{R}$, there exists some $\epsilon > 0$ such that $\det(t\boldsymbol{I} + \boldsymbol{A}) \neq 0$, for all $0 < |t| < \epsilon$, i.e. $t\boldsymbol{I} + \boldsymbol{A}$ is invertible for all $0 < |t| < \epsilon$. Thus, for any t with $0 < |t| < \epsilon$, we have $\det(x\boldsymbol{I} - (t\boldsymbol{I} + \boldsymbol{A})\boldsymbol{B}) = \det(xI - \boldsymbol{B}(t\boldsymbol{I} + \boldsymbol{A}))$. Now letting $t \to 0$ we get $\det(x\boldsymbol{I} - \boldsymbol{AB}) = \det(x\boldsymbol{I} - \boldsymbol{BA})$.

Remark. For example, $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ have the same characteristic polynomial. Since both $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ are hermitian, $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ are unitarily similar.

In general, $\boldsymbol{AB}$ is not necessarily similar to $\boldsymbol{BA}$. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$. Then $\boldsymbol{AB}$ is the zero matrix while $\boldsymbol{BA}$ is a non-zero matrix and so they are not similar. Of course if $\boldsymbol{A}$ or $\boldsymbol{B}$ is invertible, then $\boldsymbol{AB}$ is similar to $\boldsymbol{BA}$.

2.5. Let $\boldsymbol{A}$ be an $m \times n$ matrix over $\mathbb{C}$. Prove that $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ have the same set of nonzero eigenvalues.

Solution. Suppose $m \leq n$. Augment $n - m$ zero rows to bottom of $\boldsymbol{A}$ to get an $n \times n$ matrix $\boldsymbol{P}$. By the previous result, $\boldsymbol{PP}^*$ and $\boldsymbol{P}^*\boldsymbol{P}$ have the same characteristic polynomial. Note that $\boldsymbol{PP}^* = \boldsymbol{AA}^* \oplus \boldsymbol{O}_{n-m,n-m}$ and $\boldsymbol{P}^*\boldsymbol{P} = \boldsymbol{A}^*\boldsymbol{A}$. Thus $\chi_{\boldsymbol{PP}^*}(x) = x^{n-m}\chi_{\boldsymbol{AA}^*}(x)$ and $\chi_{\boldsymbol{P}^*\boldsymbol{P}}(x) = \chi_{\boldsymbol{A}^*\boldsymbol{A}}(x)$. Hence $x^{n-m}\chi_{\boldsymbol{AA}^*}(x) = \chi_{\boldsymbol{A}^*\boldsymbol{A}}(x)$ so that $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ have the same set of nonzero eigenvalues.

2.6. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Then $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ are similar.

Solution. Let $\boldsymbol{J}$ be a square matrix whose entries on the skew diagonal are all 1, and the rest of the entries are 0. Then $\boldsymbol{J}^2 = \boldsymbol{I}$. Thus $\boldsymbol{J}^{-1} = \boldsymbol{J}$. If $\boldsymbol{A}$ is a Jordan block, then $\boldsymbol{A}^\top = \boldsymbol{J}\boldsymbol{A}\boldsymbol{J}$. For example,

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

In general, the result follows from the Jordan decomposition of $\boldsymbol{A}$. Note that $\boldsymbol{J}$ is symmetric. It follows that $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ are *symmetrically* similar. If $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$, then it is also true that $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ are similar over $\mathbb{R}$. See [3] page 76.

Note that in general $\boldsymbol{A}$ and $\boldsymbol{A}^*$ are not similar.

2.7. Let $\boldsymbol{A} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where $a \neq b$. Prove that $\boldsymbol{A}$ is similar to the matrix $\begin{pmatrix} a+b & ab \\ -1 & 0 \end{pmatrix}$.

Solution. Let $\boldsymbol{P} = \begin{pmatrix} 1 & b \\ 1 & a \end{pmatrix}$. We have $\boldsymbol{P}^{-1} = \frac{1}{a-b}\begin{pmatrix} a & -b \\ -1 & 1 \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} a+b & ab \\ -1 & 0 \end{pmatrix}$.

2.8. Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

(a) Suppose $a_{12} \neq 0$. Let $\boldsymbol{P} = \begin{pmatrix} 0 & -a_{12} \\ 1 & a_{11} \end{pmatrix}$. Show that $\boldsymbol{P}$ is invertible and

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}.$$

(b) Suppose $a_{21} \neq 0$. Find an invertible matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}.$$

Solution. (a) The result follows by direct computation.

(b) Let $\boldsymbol{Q} = \begin{pmatrix} 1 & a_{22} \\ 0 & -a_{21} \end{pmatrix}$. As $a_{21} \neq 0$, the matrix $\boldsymbol{Q}$ is invertible. Then

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}.$$

Remarks. It follows from exercises 2.**??** and 2.8 that any 2×2 non-scalar matrix $\boldsymbol{A}$ is similar to the matrix $\begin{pmatrix} \operatorname{tr}(\boldsymbol{A}) & \det(\boldsymbol{A}) \\ -1 & 0 \end{pmatrix}$. Thus if $\boldsymbol{A}$ and $\boldsymbol{B}$ are two 2×2 non-scalar matrices, then $\boldsymbol{A}$ is similar to $\boldsymbol{B}$ if and only if $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{B})$ and $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$.

2.9. Let $\boldsymbol{A} = \begin{pmatrix} 9 & -40 \\ 40 & 9 \end{pmatrix}$. Find a positive definite hermitian matrice $\boldsymbol{R}$ and a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{A} = \boldsymbol{RU}$.

Solution. By proposition 2.9, we find that $\boldsymbol{R} = \begin{pmatrix} 41 & 0 \\ 0 & 41 \end{pmatrix}$ and $\boldsymbol{U} = \begin{pmatrix} \frac{9}{41} & -\frac{40}{41} \\ \frac{40}{41} & \frac{9}{41} \end{pmatrix}$.

In fact the 2 columns of $\boldsymbol{A}$ are orthogonal. The length of each column is $\sqrt{9^2 + 40^2} = 41$. Thus $\frac{1}{41}\boldsymbol{A}$ is unitary. From this see that $\boldsymbol{A} = \boldsymbol{RU}$.

2.10. Let $\boldsymbol{U} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be an upper triangular unitary matrix. Prove that $\boldsymbol{U}$ is diagonal.

Solution. Let $\boldsymbol{U} = (u_{ij})$. Since $\boldsymbol{U}$ is upper triangular, $u_{ij} = 0$ for $i > j$. It follows that $\det(\boldsymbol{U}) = u_{11} \cdots u_{nn} \neq 0$ as $\boldsymbol{U}$ is unitary. Thus $u_{ii} \neq 0$ and hence $\overline{u_{ii}} \neq 0$ for all $i = 1, \ldots, n$. We have $\boldsymbol{U}^* = (q_{ij})$, where $q_{ij} = \overline{u}_{ji}$. Since $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}_n$, we have $\sum_{k=1}^n q_{ik}u_{kj} = \delta_{ij}$. That is

$$\sum_{k=1}^{n} \overline{u}_{ki}u_{kj} = \delta_{ij} \tag{1}$$

Letting $i = 1$ in (4), we have $\sum_{k=1}^n \overline{u}_{k1}u_{kj} = \delta_{1j}$. That is $\overline{u}_{11}u_{1j} = \delta_{1j}$. Since $\overline{u}_{11} \neq 0$, we have $u_{1j} = 0$ for all $j \geq 2$.

Letting $i = 2$ in (4), we have $\sum_{k=1}^n \overline{u}_{k2}u_{kj} = \delta_{2j}$. That is $\overline{u}_{12}u_{1j} + \overline{u}_{22}u_{2j} = \delta_{2j}$. By the previous result, we know $u_{1j} = 0$ for all $j \geq 2$. Thus we have $\overline{u}_{22}u_{2j} = \delta_{2j}$ for $j \geq 2$. Since $\overline{u}_{22} \neq 0$, we have $u_{2j} = 0$ for all $j \geq 3$. Continuing in this way or using induction, we conclude that $u_{ij} = 0$ for all $j > i$. Therefore, $\boldsymbol{U}$ is a diagonal matrix.

3.1. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $f(x)$ be a polynomial with complex coefficients such that $f(\boldsymbol{AA}^*) = \boldsymbol{O}$. Prove that $f(\boldsymbol{A}^*\boldsymbol{A}) = \boldsymbol{O}$.

Solution. This follows from the fact that $\boldsymbol{AA}^*$ and $\boldsymbol{A}^*\boldsymbol{A}$ are unitarily similar. We may also use the polar decomposition of $\boldsymbol{A}$. Let $\boldsymbol{A} = \boldsymbol{RU}$, with $\boldsymbol{R}$ positive semi-definite, and $\boldsymbol{U}$ unitary. Thus $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{U}^*\boldsymbol{R}^2\boldsymbol{U} = \boldsymbol{U}^*(\boldsymbol{AA}^*)\boldsymbol{U}$, and so $f(\boldsymbol{A}^*\boldsymbol{A}) = \boldsymbol{U}^*f(\boldsymbol{AA}^*)\boldsymbol{U} = \boldsymbol{O}$.

3.2. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a normal matrix and V an eigenspace of $\boldsymbol{A}^*$. Prove that V and $V^\perp$ are subspaces of $\mathbb{C}^n$ invariant under $\boldsymbol{A}$.

Solution. Let V be an eigenspace of $\boldsymbol{A}^*$ associated with the eigenvalue λ. That is $V = \{\boldsymbol{x} \in \mathbb{C}^n \mid \boldsymbol{A}^*\boldsymbol{x} = \lambda\boldsymbol{x}\}$. Let $\boldsymbol{x} \in V$. Since $\boldsymbol{A}$ is normal, $\boldsymbol{A}^*(\boldsymbol{Ax}) = \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{x}) = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda(\boldsymbol{Ax})$. This shows that $\boldsymbol{Ax} \in V$. That is $\boldsymbol{A}[V] \subset V$. Next let $\boldsymbol{y} \in V^\perp$. Thus $\boldsymbol{y}^*\boldsymbol{x} = 0$ for all $\boldsymbol{x} \in V$. It follows that $(\boldsymbol{Ay})^*\boldsymbol{x} = \boldsymbol{y}^*\boldsymbol{A}^*\boldsymbol{x} = \boldsymbol{y}^*(\lambda\boldsymbol{x}) = \lambda\boldsymbol{y}^*\boldsymbol{x} = 0$ for all $\boldsymbol{x} \in V$. This shows that $\boldsymbol{Ay} \in V^\perp$. That is $\boldsymbol{A}[V^\perp] \subset V^\perp$. Consequently, both V and $V^\perp$ are invariant under $\boldsymbol{A}$.

3.3. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\boldsymbol{A} \neq \boldsymbol{O}_n$. Prove that there exists a 1-dimensional or a 2-dimensional subspace of $\mathbb{R}^n$ invariant under $\boldsymbol{A}$.

Solution. Let $m_{\boldsymbol{A}}(x)$ be the minimal polynomial of $\boldsymbol{A}$. Since $\boldsymbol{A}$ is a real matrix, $m_{\boldsymbol{A}}(x)$ is a polynomial with real coefficients. A polynomial with real coefficients always factors into a product of linear and (irreducible) quadratic factors. Suppose $(x-\lambda)$ is a linear factor of $m_{\boldsymbol{A}}(x)$. Then λ is an eigenvalue of $\boldsymbol{A}$. Let $\boldsymbol{x}$ be an associated eigenvector

of $\boldsymbol{A}$. Then the span of $\boldsymbol{x}$ is a 1-dimensional subspace of $\mathbb{R}^n$ invariant under $\boldsymbol{A}$. Suppose $m_{\boldsymbol{A}}(x)$ has a quadratic factor x^2-ax+b. If $(\boldsymbol{A}^2+a\boldsymbol{A}+b\boldsymbol{I})(\boldsymbol{x}) \neq \boldsymbol{0}$ for all $\boldsymbol{x} \neq \boldsymbol{0}$, then $\boldsymbol{A}^2+a\boldsymbol{A}+b\boldsymbol{I}$ is injective, and hence non-singular. Thus $m_{\boldsymbol{A}}(\boldsymbol{A})(\boldsymbol{A}^2+a\boldsymbol{A}+b\boldsymbol{I})^{-1} = \boldsymbol{O}_n$. That means $\boldsymbol{A}$ satisfies the polynomial $m_{\boldsymbol{A}}(x)/(x^2+ax+b)$ which is of degree less than $m_{\boldsymbol{A}}(x)$, contradicting the minimality of the degree of $m_{\boldsymbol{A}}(x)$. Therefore, there exists $\boldsymbol{x} \neq \boldsymbol{0}$ such that $(\boldsymbol{A}^2+a\boldsymbol{A}+b\boldsymbol{I})(\boldsymbol{x}) = \boldsymbol{0}$. It follows that the span of $\boldsymbol{x}$ and $\boldsymbol{A}(\boldsymbol{x})$ is an invariant subspace of $\boldsymbol{A}$ of dimension 1 or 2.

3.4. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $\boldsymbol{A}^2 = \boldsymbol{A}$. Prove that $\mathrm{rank}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$.

Solution. Since $\boldsymbol{A}^2 = \boldsymbol{A}$, the eigenvalues of $\boldsymbol{A}$ are 0 or 1. Thus $\mathrm{rank}(\boldsymbol{A}) =$ multiplicity of the eigenvalue $1 = \mathrm{tr}(\boldsymbol{A})$.

3.5. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be such that $\boldsymbol{A}^2 = \boldsymbol{A}$, $\boldsymbol{B}^2 = \boldsymbol{B}$ and $\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{A}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A} = \boldsymbol{O}_n$. What can be said about $\boldsymbol{A}$ and $\boldsymbol{B}$?

Solution. Both $\boldsymbol{A}$ and $\boldsymbol{B}$ are zero matrices. This is proved as follow. Since $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{B}^2 = \boldsymbol{B}$, we have $(\boldsymbol{A}+\boldsymbol{B})^2 = \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{A}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{A} = \boldsymbol{O}_n$. Thus $\boldsymbol{A}+\boldsymbol{B}$ is nilpotent so that its eigenvalues are zero. Therefore $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = 0$. Thus $0 = \mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B}) = \mathrm{rank}(\boldsymbol{A}) + \mathrm{rank}(\boldsymbol{B})$ as $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{B}^2 = \boldsymbol{B}$. It follows that $\mathrm{rank}(\boldsymbol{A}) = 0$ and $\mathrm{rank}(\boldsymbol{B}) = 0$, so $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{O}_n$.

3.6. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Prove that $\boldsymbol{A}$ is symmetrically similar to $\boldsymbol{A}^\top$ over $\mathbb{R}$ if and only if $\boldsymbol{A}$ is a product of two real symmetric matrices and one of them is invertible.

Solution. Suppose $\boldsymbol{A}^\top = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, where $\boldsymbol{P}$ is symmetric. Then $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{A}^\top$ is symmetric as $(\boldsymbol{A}\boldsymbol{P})^\top = (\boldsymbol{P}\boldsymbol{A}^\top)^\top = \boldsymbol{A}\boldsymbol{P}$ since $\boldsymbol{P}$ is symmetric. Thus $\boldsymbol{A} = (\boldsymbol{A}\boldsymbol{P})\boldsymbol{P}^{-1}$ is a product of two symmetric matrices $\boldsymbol{A}\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$, where $\boldsymbol{P}^{-1}$ is invertible. Alternatively, we can also have $\boldsymbol{A} = \boldsymbol{P}(\boldsymbol{P}^{-1}\boldsymbol{A})$, where $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}\boldsymbol{A}$ are symmetric and $\boldsymbol{P}$ is invertible.

Conversely, suppose $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{Q}$ where $\boldsymbol{P}, \boldsymbol{Q}$ are symmetric and $\boldsymbol{P}$ is invertible. Then $\boldsymbol{A}^\top = (\boldsymbol{P}\boldsymbol{Q})^\top = \boldsymbol{Q}\boldsymbol{P} = \boldsymbol{P}^{-1}(\boldsymbol{P}\boldsymbol{Q})\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

Using rational canonical form, it can be proved that a square real matrix is always symmetrically similar to its transpose. See [3] page 76. Thus any square real matrix is a product of two real symmetric matrices.

3.7. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Prove that $\boldsymbol{A}$ is diagonalizable over $\mathbb{R}$ with all its eigenvalues positive if and only if $\boldsymbol{A}$ is a product of two real positive definite symmetric matrices.

Solution. Suppose $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{D}\boldsymbol{P}$, where $\boldsymbol{D} = \boldsymbol{D}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix and $\lambda_i > 0$ for $i = 1, \ldots, n$. Then $\boldsymbol{A} = \left(\boldsymbol{P}^{-1}(\boldsymbol{P}^{-1})^\top\right)\left(\boldsymbol{P}^\top\boldsymbol{D}\boldsymbol{P}\right)$. Here $\boldsymbol{P}^{-1}(\boldsymbol{P}^{-1})^\top$ is positive definite symmetric and $\boldsymbol{P}^\top\boldsymbol{D}\boldsymbol{P}$ is also positive definite symmetric as $\lambda_i > 0$ for all i.

Conversely, suppose $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{N}$, where $\boldsymbol{M}$ and $\boldsymbol{N}$ are real positive definite symmetric matrices. Then $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{N}^{\frac{1}{2}}\boldsymbol{N}^{\frac{1}{2}} = \boldsymbol{N}^{-\frac{1}{2}}\left(\boldsymbol{N}^{\frac{1}{2}}\boldsymbol{M}\boldsymbol{N}^{\frac{1}{2}}\right)\boldsymbol{N}^{\frac{1}{2}}$. Thus $\boldsymbol{A}$ is similar to the symmetric matrix $\boldsymbol{N}^{\frac{1}{2}}\boldsymbol{M}\boldsymbol{N}^{\frac{1}{2}}$, and hence is diagonalizable. Since $\boldsymbol{M}$ is positive definite, the symmetric matrix $\boldsymbol{N}^{\frac{1}{2}}\boldsymbol{M}\boldsymbol{N}^{\frac{1}{2}}$ is also positive definite so that all its eigenvalues are positive. Therefore, all eigenvalues of $\boldsymbol{A}$ are positive.

3.8. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a normal matrix. Prove that the null space of $\boldsymbol{A}$ is the same as the null space of $\boldsymbol{A}^*$.

Solution. Let $\boldsymbol{x} \in \mathbb{C}^n$. We have $|\boldsymbol{A}\boldsymbol{x}|^2 = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle = \langle \boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x}\rangle = \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{A}^*\boldsymbol{x}\rangle = \langle \boldsymbol{A}^*\boldsymbol{x}, \boldsymbol{A}^*\boldsymbol{x}\rangle = |\boldsymbol{A}^*\boldsymbol{x}|^2$. Thus $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ if and only if $\boldsymbol{A}^*\boldsymbol{x} = \boldsymbol{0}$.

3.9. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}$ is a normal matrix. Prove that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{O}_n$ implies $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{O}_n$.

Solution. Since each column vector of $\boldsymbol{B}$ is in the null space of $\boldsymbol{A}$, it is also in the null space of $\boldsymbol{A}^*$ by the previous exercise. Alternatively we can use the inner product $\mathrm{tr}(\boldsymbol{X}^*\boldsymbol{Y})$ of matrices. Since $|\boldsymbol{A}^*\boldsymbol{B}|^2 = \mathrm{tr}((\boldsymbol{A}^*\boldsymbol{B})^*(\boldsymbol{A}^*\boldsymbol{B})) = \mathrm{tr}(\boldsymbol{B}^*\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{B}) = 0$, we have $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{O}_n$.

3.10. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}$. Prove that $\boldsymbol{A}$ is normal.

Solution. We use the inner product $\mathrm{tr}(\boldsymbol{X}^*\boldsymbol{Y})$ of matrices. Note that both $\boldsymbol{A}^*\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^*$ are hermitian matrices.

$|\boldsymbol{A}^*\boldsymbol{A}-\boldsymbol{A}\boldsymbol{A}^*|^2 = \mathrm{tr}((\boldsymbol{A}^*\boldsymbol{A}-\boldsymbol{A}\boldsymbol{A}^*)(\boldsymbol{A}^*\boldsymbol{A}-\boldsymbol{A}\boldsymbol{A}^*)) = \mathrm{tr}(\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A})-(\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A})\boldsymbol{A}^*-\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}+\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}^*) = \mathrm{tr}(\boldsymbol{A}^*(\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A})-(\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A})\boldsymbol{A}^*-\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}+\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}^*) = \mathrm{tr}(\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}) - \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}) - \mathrm{tr}(\boldsymbol{A}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}) = 0$.

Thus $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^*$, and $\boldsymbol{A}$ is normal. Similarly if $\boldsymbol{A}$ commutes with $\boldsymbol{A}\boldsymbol{A}^* - \boldsymbol{A}^*\boldsymbol{A}$, then $\boldsymbol{A}$ is normal.

4.1. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$. Suppose $\boldsymbol{A}$ is a normal matrix. Prove that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$ implies $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}^*$.

Solution. We use the inner product $\mathrm{tr}(\boldsymbol{X}^*\boldsymbol{Y})$. We have $(\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}^*)^*(\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}^*) = (\boldsymbol{B}^*\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B}^*)(\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}^*) = \boldsymbol{B}^*\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}^*\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^* + \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{A}^* = \boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{B} - \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}^*\boldsymbol{B}\boldsymbol{A}\boldsymbol{A}^* + \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{A}^* = \boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{B}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{B} - \boldsymbol{B}^*\boldsymbol{B}\boldsymbol{A}^*\boldsymbol{A} + \boldsymbol{A}\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{A}^*$. By the cyclic property of the trace, the trace of this last expression of matrices is zero. Thus $\boldsymbol{A}^*\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}^*$. In particular, if we take $\boldsymbol{B} = \boldsymbol{A}$, then $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^*$. Therefore, $\boldsymbol{A}$ is normal if and only if $\boldsymbol{A}^*$ commutes with every matrix that commutes with $\boldsymbol{A}$.

4.2. Let $\boldsymbol{A}, \boldsymbol{B}$ be positive definite hermitian matrices. Prove that $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) > 1$.

Solution. As $\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}^{-\frac{1}{2}}(\boldsymbol{I} + \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{\frac{1}{2}})\boldsymbol{B}^{\frac{1}{2}}$, we have $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I} + \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{\frac{1}{2}})$. The matrix $\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{\frac{1}{2}}$ is positive definite hermitian. Thus there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^*(\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}\boldsymbol{B}^{\frac{1}{2}})\boldsymbol{U} = \boldsymbol{D}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_i > 0$ for $i = 1, \ldots, n$. It follows that $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) = (1+\lambda_1)\cdots(1+\lambda_n) > 1$. Basically $\boldsymbol{A}\boldsymbol{B}$ is similar to a diagonalizable matrix with all eigenvalues positive.

4.3. Let $\boldsymbol{A} = \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 3 & 7 \\ 7 & 18 \end{pmatrix}$ and $\boldsymbol{C} = \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{3}{5} \end{pmatrix}$. Show that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are positive definite hermitian matrices. Is it true that $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) > 1$?

Solution. Direct computation gives $\boldsymbol{ABC} = \begin{pmatrix} -1 & 6 \\ 0 & -1 \end{pmatrix}$ and $\det(\boldsymbol{I} + \boldsymbol{ABC}) = 0$. By the previous exercise, the matrix $\begin{pmatrix} -1 & 6 \\ 0 & -1 \end{pmatrix}$ cannot be written as a product of two positive definite hermitian matrices.

4.4. Show that the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ cannot be expressed as a product of two singular symmetric matrices.

Solution. Let $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & a \\ a & y \end{pmatrix} \begin{pmatrix} p & b \\ b & q \end{pmatrix} = \begin{pmatrix} xp + ab & xb + aq \\ ap + yb & ab + yq \end{pmatrix}$. The solutions are $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & b \\ b & q \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & a \\ a & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{a} \end{pmatrix}$, where x and q are arbitrary, a and b are nonzero. In either form, one of the matrices is non-singular.

Note that for $a \neq 0$, we have $\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}$ is a product of two singular symmetric matrices.

4.5. Let $\boldsymbol{Q}$ and $\boldsymbol{M}$ be symmetric matrices with $\boldsymbol{Q}$ positive definite. Show there is a matrix $\boldsymbol{A}$ such that $\boldsymbol{AQ} + \boldsymbol{QA} = \boldsymbol{M}$.

Solution. We may assume $\boldsymbol{Q}$ is a diagonal matrix $\boldsymbol{D}(q_1, \ldots, q_n)$. Since $\boldsymbol{Q}$ is positive definite, we have $q_i > 0$ for all $i = 1, \ldots, n$. Let $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{M} = (m_{ij})$. Then the (i, j)-entry of $\boldsymbol{AQ} + \boldsymbol{QA}$ is $a_{ij}(q_i + q_j)$. Thus $a_{ij} = m_{i,j}/(q_i + q_j)$ and that determines the matrix $\boldsymbol{A}$.

4.6. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{C})$ be Hermitian matrices such that $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{B}^2 = \boldsymbol{B}$. Prove that if $\boldsymbol{AB}$ is normal, then $\boldsymbol{AB} = \boldsymbol{BA}$.

Solution. Since $\boldsymbol{A}$ and $\boldsymbol{B}$ are Hermitian, we have $\boldsymbol{A}^2 = \boldsymbol{A} = \boldsymbol{A}^*$ and $\boldsymbol{B}^2 = \boldsymbol{B} = \boldsymbol{B}^*$. As $\boldsymbol{AB}$ is normal, it is unitarily similar to a diagonal matrix. Thus there exist a unitary matrix $\boldsymbol{U}$ and a diagonal matrix $\boldsymbol{D}$ such that $\boldsymbol{AB} = \boldsymbol{U}^*\boldsymbol{DU}$. Now $(\boldsymbol{AB})^3 = \boldsymbol{AB}(\boldsymbol{ABA})\boldsymbol{B} = \boldsymbol{AB}(\boldsymbol{AB}(\boldsymbol{AB})^*)\boldsymbol{B} = \boldsymbol{AB}((\boldsymbol{AB})^*\boldsymbol{AB})\boldsymbol{B} = \boldsymbol{AB}(\boldsymbol{BAB})\boldsymbol{B} = (\boldsymbol{AB})^2$. Thus 0 and 1 are the only possible eigenvalues for $\boldsymbol{AB}$. It follows that $\boldsymbol{D}^2 = \boldsymbol{D}$ and $\boldsymbol{D}^* = \boldsymbol{D}$. Therefore $(\boldsymbol{AB})^* = \boldsymbol{U}^*\boldsymbol{D}^*\boldsymbol{U} = \boldsymbol{U}^*\boldsymbol{DU} = \boldsymbol{AB}$. That is $\boldsymbol{AB}$ is Hermitian. Consequently, $\boldsymbol{AB} = (\boldsymbol{AB})^* = \boldsymbol{B}^*\boldsymbol{A}^* = \boldsymbol{BA}$.

4.7. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ such that $\boldsymbol{A}^2 = \boldsymbol{A}^*$. Prove that $\operatorname{rank}(\boldsymbol{A} + \boldsymbol{A}^*) = \operatorname{rank}(\boldsymbol{A})$ and $\boldsymbol{I} + \boldsymbol{A}$ is non-singular. [AMM problem 1793]

Solution. If $(\boldsymbol{I} + \boldsymbol{A})\boldsymbol{u} = \boldsymbol{0}$, then $\boldsymbol{Au} = -\boldsymbol{u}$. Hence, after multiplying on the left by $\boldsymbol{A}^*$, we have $\boldsymbol{A}^*\boldsymbol{Au} = \boldsymbol{A}^2(-\boldsymbol{u}) = -\boldsymbol{u}$. Since all eigenvalues of $\boldsymbol{A}^*\boldsymbol{A}$ are nonnegative, we conclude that $\boldsymbol{u} = \boldsymbol{0}$ and therefore $\boldsymbol{I} + \boldsymbol{A}$ is non-singular.

Since $\boldsymbol{A} + \boldsymbol{A}^* = \boldsymbol{A} + \boldsymbol{A}^2 = (\boldsymbol{I} + \boldsymbol{A})\boldsymbol{A}$ and $\boldsymbol{I} + \boldsymbol{A}$ is non-singular, $\boldsymbol{A} + \boldsymbol{A}^*$ and $\boldsymbol{A}$ have the same rank.

4.8. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 matrices. Prove that $\det(\boldsymbol{A} + \boldsymbol{B}) + \operatorname{tr}(\boldsymbol{AB}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}) + \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$.

Solution. Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. Then

$\mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}) = (a_{11}+a_{22})(b_{11}+b_{22}) = a_{11}b_{11}+a_{22}b_{22}+a_{11}b_{22}+a_{22}b_{11}$,

$\mathrm{tr}(\boldsymbol{AB}) = a_{11}b_{11}+a_{12}b_{21}+a_{21}b_{12}+a_{22}b_{22}$,

$\mathrm{tr}(\boldsymbol{AB}) - \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}) = (a_{11}b_{22}+a_{22}b_{11}) - (a_{12}b_{21}+a_{21}b_{12})$.

Therefore $\det(\boldsymbol{A}+\boldsymbol{B}) = (a_{11}+b_{11})(a_{22}+b_{22}) - (a_{12}+b_{12})(a_{21}+b_{21}) = (a_{11}a_{22}+a_{11}b_{22}+b_{11}a_{22}+b_{11}b_{22}) - (a_{12}a_{21}+a_{12}b_{21}+b_{12}a_{21}+b_{12}b_{21}) = (a_{11}a_{22}-a_{12}a_{21}) + (b_{11}b_{22}-b_{12}b_{21}) + (a_{11}b_{22}+b_{11}a_{22}) - (a_{12}b_{21}+b_{12}a_{21}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{AB}) - \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B})$.

4.9. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 nonzero matrices such that $\boldsymbol{AB}+\boldsymbol{BA} = \boldsymbol{O}$ and $\det(\boldsymbol{A}+\boldsymbol{B}) = 0$. Prove that $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B}) = 0$.

Solution. $\boldsymbol{AB}+\boldsymbol{BA} = \boldsymbol{O} \Rightarrow (\boldsymbol{A}+\boldsymbol{B})^2 = \boldsymbol{A}^2+\boldsymbol{B}^2 = (\boldsymbol{A}-\boldsymbol{B})^2$. Thus $\det(\boldsymbol{A}+\boldsymbol{B}) = 0 \Rightarrow \det(\boldsymbol{A}-\boldsymbol{B}) = 0$. By Cayley-Hamilton Theorem applied to the matrices $\boldsymbol{A}+\boldsymbol{B}$ and $\boldsymbol{A}-\boldsymbol{B}$, we have,

$$(\boldsymbol{A}+\boldsymbol{B})^2 - \mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B}) + \det(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{I}_2 = \boldsymbol{O},$$

$$(\boldsymbol{A}-\boldsymbol{B})^2 - \mathrm{tr}(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B}) + \det(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{I}_2 = \boldsymbol{O}.$$

Since $(\boldsymbol{A}+\boldsymbol{B})^2 = (\boldsymbol{A}-\boldsymbol{B})^2$ and $\det(\boldsymbol{A}+\boldsymbol{B}) = \det(\boldsymbol{A}-\boldsymbol{B}) = 0$, we have

$$\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})$$

$$\Rightarrow (\mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{B}))(\boldsymbol{A}+\boldsymbol{B}) = (\mathrm{tr}(\boldsymbol{A})-\mathrm{tr}(\boldsymbol{B}))(\boldsymbol{A}-\boldsymbol{B}).$$

$$\Rightarrow \mathrm{tr}(\boldsymbol{A})\boldsymbol{B}+\mathrm{tr}(\boldsymbol{B})\boldsymbol{A} = \boldsymbol{O}.$$

If $\mathrm{tr}(\boldsymbol{A}) \neq 0$, then $\boldsymbol{B} = -(\mathrm{tr}(\boldsymbol{B})/\mathrm{tr}(\boldsymbol{A}))\boldsymbol{A}$. Thus $\mathrm{tr}(\boldsymbol{B}) = -(\mathrm{tr}(\boldsymbol{B})/\mathrm{tr}(\boldsymbol{A}))\mathrm{tr}(\boldsymbol{A}) = -\mathrm{tr}(\boldsymbol{B})$ so that $\mathrm{tr}(\boldsymbol{B}) = 0$. This implies $\boldsymbol{B} = -(\mathrm{tr}(\boldsymbol{B})/\mathrm{tr}(\boldsymbol{A}))\boldsymbol{A} = \boldsymbol{O}$, contradicting the assumption that $\boldsymbol{B}$ is not the zero matrix. Therefore $\mathrm{tr}(\boldsymbol{A}) = 0$. Similarly, $\mathrm{tr}(\boldsymbol{B}) = 0$.

Remark: Using

$$\det(\boldsymbol{A}+\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{AB}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}) + \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}),$$

$$\det(\boldsymbol{A}-\boldsymbol{B}) - \mathrm{tr}(\boldsymbol{AB}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B}) - \mathrm{tr}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}),$$

we also have $\det(\boldsymbol{A}) + \det(\boldsymbol{B}) = 0$.

4.10. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 2×2 matrices such that $\boldsymbol{A}^2+\boldsymbol{B}^2 = 2\boldsymbol{AB}$. Prove that $\boldsymbol{AB} = \boldsymbol{BA}$ and $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B})$. [Romanian Mathematical Competitions 2011]

Solution. $\boldsymbol{A}^2+\boldsymbol{B}^2 = 2\boldsymbol{AB} \Leftrightarrow (\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{AB}-\boldsymbol{BA} \Leftrightarrow \boldsymbol{A}(\boldsymbol{A}-\boldsymbol{B}) = (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{B}$.

Suppose $\boldsymbol{AB}-\boldsymbol{BA}$ is non-singular, then $\boldsymbol{A}-\boldsymbol{B}$ is also non-singular. Thus $\boldsymbol{A}-\boldsymbol{B} = \boldsymbol{A}-(\boldsymbol{A}-\boldsymbol{B})^{-1}(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{B} = \boldsymbol{A}-(\boldsymbol{A}-\boldsymbol{B})^{-1}\boldsymbol{A}(\boldsymbol{A}-\boldsymbol{B})$. This implies that $\boldsymbol{I}_2 = \boldsymbol{A}(\boldsymbol{A}-\boldsymbol{B})^{-1} - (\boldsymbol{A}-\boldsymbol{B})^{-1}\boldsymbol{A}$. Taking the trace on both sides gives a contradiction.

Therefore $\boldsymbol{AB}-\boldsymbol{BA}$ and $\boldsymbol{A}-\boldsymbol{B}$ are singular. Thus $\det(\boldsymbol{A}-\boldsymbol{B}) = 0$. By Cayley-Hamilton Theorem, we have $(\boldsymbol{A}-\boldsymbol{B})^2 - (\mathrm{tr}(\boldsymbol{A}-\boldsymbol{B}))(\boldsymbol{A}-\boldsymbol{B}) = \boldsymbol{O}$.

Thus $\boldsymbol{AB}-\boldsymbol{BA} = (\boldsymbol{A}-\boldsymbol{B})^2 = (\mathrm{tr}(\boldsymbol{A}-\boldsymbol{B}))(\boldsymbol{A}-\boldsymbol{B})$. Taking the trace on both sides, we have $0 = (\mathrm{tr}(\boldsymbol{A}-\boldsymbol{B}))^2$. This gives $\boldsymbol{AB}-\boldsymbol{BA} = \boldsymbol{O}$ and $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{B})$.

5.1. Show that the matrix $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ is a product of two symmetric matrices.

Solution. $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \lambda \\ 0 & \lambda & 1 \\ \lambda & 1 & 0 \end{pmatrix}$. Note that the first matrix is non-singular.

5.2. Prove that any matrix is a product of two symmetric matrices and one of them is non-singular.

Solution. By 5.1, any Jordan block is a product of two symmetric matrices. It follows that any matrix in Jordan form is a product of two symmetric matrices. Let $\boldsymbol{A}$ be a matrix and $\boldsymbol{J}$ is its Jordan form. Thus $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P}$, and $\boldsymbol{J} = \boldsymbol{X}\boldsymbol{Y}$, where $\boldsymbol{X}, \boldsymbol{Y}$ are symmetric matrices. Also as remarked in 5.1, $\boldsymbol{X}$ can be chosen to be non-singular. Therefore, $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{P} = (\boldsymbol{P}^{-1}\boldsymbol{X}(\boldsymbol{P}^{-1})^\top)(\boldsymbol{P}^\top\boldsymbol{Y}\boldsymbol{P})$, where both $\boldsymbol{P}^{-1}\boldsymbol{X}(\boldsymbol{P}^{-1})^\top$ and $\boldsymbol{P}^\top\boldsymbol{Y}\boldsymbol{P}$ are symmetric and $\boldsymbol{P}^{-1}\boldsymbol{X}(\boldsymbol{P}^{-1})^\top$ is non-singular.

5.3. Prove that any matrix $\boldsymbol{A}$ is similar to its transpose $\boldsymbol{A}^\top$.

Solution. This problem is solved in 2.6. We give a slightly different solution based on 5.2. By 5.2, there exist symmetric matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ with $\boldsymbol{M}$ invertible such that $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{N}$. Thus $\boldsymbol{A}^\top = \boldsymbol{N}^\top\boldsymbol{M}^\top = \boldsymbol{N}\boldsymbol{M} = \boldsymbol{M}^{-1}(\boldsymbol{M}\boldsymbol{N})\boldsymbol{M} = \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M}$.

5.4. Let $\boldsymbol{X} \in SL(2, \mathbb{C})$. Prove that $\boldsymbol{X}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}$ for some $\boldsymbol{P} \in SL(2, \mathbb{C})$.

Solution. Since $\boldsymbol{X} \in SL(2, \mathbb{C})$, we have $\operatorname{tr}(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X}^{-1})$ and $\det(\boldsymbol{X}) = \det(\boldsymbol{X}^{-1}) = 1$. By exercises 2.**??** and 2.8, if $\boldsymbol{X} \neq \pm\boldsymbol{I}$, then $\boldsymbol{X}$ and $\boldsymbol{X}^{-1}$ are similar. Certainly, this is true for $\boldsymbol{X} = \pm\boldsymbol{I}$ too since $\pm\boldsymbol{I}$ is its own inverse. Thus there exists $\boldsymbol{Q}$ such that $\boldsymbol{X}^{-1} = \boldsymbol{Q}^{-1}\boldsymbol{X}\boldsymbol{Q}$. Let $\boldsymbol{P} = \frac{1}{\sqrt{\det(\boldsymbol{Q})}}\boldsymbol{Q}$. Then $\boldsymbol{P} \in SL(2, \mathbb{C})$ and $\boldsymbol{X}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}$.

Note that this is not true in $SL(2, R)$ as $\det(\boldsymbol{Q})$ can be negative. For example, $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is not conjugate to its inverse in $SL(2, R)$.

5.5. Let $\boldsymbol{A} \in SL(2, \mathbb{C})$. Prove that $\operatorname{tr}(\boldsymbol{A}^2) = (\operatorname{tr}(\boldsymbol{A}))^2 - 2$.

Solution. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $ad - bc = 1$. Then $\boldsymbol{A}^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & d^2 + bc \end{pmatrix}$.

Thus $\operatorname{tr}(\boldsymbol{A}^2) = a^2 + bc + d^2 + bc = (a+d)^2 - 2ad + 2bc = (\operatorname{tr}(\boldsymbol{A}))^2 - 2$.

Alternatively, by the Cayley-Hamilton theorem, we have $\boldsymbol{A}^2 - \operatorname{tr}(\boldsymbol{A})\boldsymbol{A} + \boldsymbol{I} = \boldsymbol{O}$. The result follows by taking the trace of both sides.

5.6. Let $\boldsymbol{A} \in SL(2, \mathbb{C})$. Prove that if $\operatorname{tr}(\boldsymbol{A}^2) = -2$, then $\boldsymbol{A}^2 = -\boldsymbol{I}$.

Solution. By 5.5, $\operatorname{tr}(\boldsymbol{A}) = 0$. By the Cayley-Hamilton theorem, $\boldsymbol{A}^2 = -\boldsymbol{I}$.

Thus the matrix $\begin{pmatrix} -1 & a \\ 0 & -1 \end{pmatrix}$, $a \neq 0$ cannot have a square root in $SL(2, \mathbb{C})$.

5.7. Let $\boldsymbol{X} \in SL(2, \mathbb{C})$. Prove that if $\boldsymbol{X}$ has a square root $\boldsymbol{A} \in SL(2, \mathbb{C})$, then $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{B} \in SL(2, \mathbb{C})$.

Solution. Since $\boldsymbol{X}$ has a square root $\boldsymbol{A}$, we have $\boldsymbol{X} = \boldsymbol{A}^2$. By 5.4, $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$, for some $\boldsymbol{B} \in SL(2,\mathbb{C})$. Thus $\boldsymbol{X} = \boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$.

5.8. Let $\boldsymbol{X} \in SL(2,\mathbb{C})$. Prove that $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{C})$.

Solution. Since $\det \boldsymbol{X} = 1$, the 2 eigenvalues of $\boldsymbol{X}$ are either distinct, or both 1 or both -1. It suffices to show that the Jordan form of $\boldsymbol{X}$ can be expressed as a multiplicative commutator.

If the Jordan form of $\boldsymbol{X}$ is $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$, then $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a^{\frac{1}{2}} & 0 \\ 0 & a^{-\frac{1}{2}} \end{pmatrix}^2$.

If the Jordan form of $\boldsymbol{X}$ is $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix}^2$.

In these two cases, we apply 5.7 so get $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{C})$.

If the Jordan form of $\boldsymbol{X}$ is $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$, then $\boldsymbol{X}$ can be expressed as a multiplicative commutator:

$\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$, where $\boldsymbol{A} = \begin{pmatrix} \frac{3+\sqrt{5}}{2} & 0 \\ -6 & \frac{3-\sqrt{5}}{2} \end{pmatrix}$ $\boldsymbol{B} = \begin{pmatrix} \frac{5-\sqrt{5}}{2} & \frac{\sqrt{5}-1}{6} \\ 6 & \frac{1+\sqrt{5}}{2} \end{pmatrix}$.

5.9. Do there exist $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$ such that $-\boldsymbol{I} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$?

Solution. No. Suppose $-\boldsymbol{I} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ for some $\boldsymbol{A}, \boldsymbol{B} \in SL(2,\mathbb{R})$. Then $\boldsymbol{A}\boldsymbol{B} = -\boldsymbol{B}\boldsymbol{A} \Rightarrow \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B} = -\boldsymbol{A} \Rightarrow \mathrm{tr}(\boldsymbol{A}) = -\mathrm{tr}(\boldsymbol{A}) \Rightarrow \mathrm{tr}(\boldsymbol{A}) = 0$. Similarly, $\mathrm{tr}(\boldsymbol{B}) = 0$.

Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} u & v \\ w & -u \end{pmatrix}$, where $-a^2 - bc = 1, -u^2 - vw = 1$ and $a, b, c, u, v, w \in \mathbb{R}$. The condition $\boldsymbol{A}\boldsymbol{B} = -\boldsymbol{B}\boldsymbol{A}$ gives rise to only one equation $2au = -bw - cv$. Multiplying both sides by bv, and using the equations $-a^2 - bc = 1, -u^2 - vw = 1$, we have

$2abuv = -b^2vw - bcv^2 \Rightarrow 2abuv = b^2(1+u^2) + v^2(1+a^2) \Rightarrow b^2 + v^2 + (bu - av)^2 = 0$ $\Rightarrow b = v = 0 \Rightarrow -a^2 = 1$ and $-u^2 = 1$, a contradiction.

5.10. Let $\boldsymbol{P} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 7 & 4 \end{pmatrix}$ and $\boldsymbol{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 3 & 1 \\ -2 & 0 \end{pmatrix}$. Show that $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{-1}$ and $-\boldsymbol{I} = \boldsymbol{P}^2$.

Solution. Just direct computation. This shows that $-\boldsymbol{I}$ is a product of two multiplicative commutators.

Bibliography

[1] S. Lipschutz, *Schaum's Theory & Problems of Linear Algebra*, Schaum Publishing Co, McGraw-Hill (1968).

[2] K.M.Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice Hall, (1971).

[3] I.Kaplansky, *Linear Algebra and Geometry, A Second Course*, Dover, (2003).