Chapter 1

Matrix Calculus

1.1 Definitions and Notation

We assume that the reader is familiar with some basic terms in linear algebra such as vector spaces, linearly dependent vectors, matrix addition and matrix multiplication (Horn and Johnson [30], Laub [39]).

Throughout we consider matrices over the field of complex numbers $\mathbb{C}$ or real number $\mathbb{R}$. Let $z \in \mathbb{C}$ with $z = x + iy$ and $x, y \in \mathbb{R}$. Then $\bar{z} = x - iy$. In some cases we restrict the underlying field to the real numbers $\mathbb{R}$. The matrices are denoted by A, B, C, D, X, Y. The matrix elements (entries) of the matrix A are denoted by a_{jk}. For the column vectors we write $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$. The zero column vector is denoted by $\mathbf{0}$. Let A be a matrix. Then A^T denotes the transpose and $\bar{A}$ is the complex conjugate matrix. We call A^* the adjoint matrix, where $A^* := \bar{A}^T$. A special role is played by the $n \times n$ matrices, i.e. the square matrices. In this case we also say the matrix is of order n. I_n denotes the $n \times n$ unit matrix (also called identity matrix). The zero matrix is denoted by 0.

Let V be a vector space of finite dimension n, over the field $\mathbb{R}$ of real numbers, or the field $\mathbb{C}$ of complex numbers. If there is no need to distinguish between the two, we speak of the field $\mathbb{F}$ of *scalars*. A *basis* of V is a set $\{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \}$ of n linearly independent vectors of V, denoted by $(\mathbf{e}_j)_{j=1}^n$. Every vector $\mathbf{v} \in V$ then has the unique representation

$$\mathbf{v} = \sum_{j=1}^{n} v_j \mathbf{e}_j$$

the scalars v_j, which we will sometimes denote by $(\mathbf{v})_j$, being the *components* of the vector $\mathbf{v}$ relative to the basis $(\mathbf{e})_j$. As long as a basis is fixed unambiguously, it is thus always possible to identify V with $\mathbb{F}^n$. In matrix notation, the vector $\mathbf{v}$ will always be represented by the *column vector*

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

while $\mathbf{v}^T$ and $\mathbf{v}^*$ will denote the following *row vectors*

$$\mathbf{v}^T = (v_1, v_2, \ldots, v_n), \qquad \mathbf{v}^* = (\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n)$$

where $\bar{\alpha}$ is the complex conjugate of α. The row vector $\mathbf{v}^T$ is the *transpose* of the column vector $\mathbf{v}$, and the row vector $\mathbf{v}^*$ is the *conjugate transpose* of the column vector $\mathbf{v}$.

Definition 1.1. Let $\mathbb{C}^n$ be the familiar n dimensional vector space. Let $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$. Then the *scalar* (or *inner*) *product* is defined as

$$(\mathbf{u}, \mathbf{v}) := \sum_{j=1}^{n} \bar{u}_j v_j.$$

Obviously

$$(\mathbf{u}, \mathbf{v}) = \overline{(\mathbf{v}, \mathbf{u})}$$

and

$$(\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}) = (\mathbf{u}_1, \mathbf{v}) + (\mathbf{u}_2, \mathbf{v}).$$

Since $\mathbf{u}$ and $\mathbf{v}$ are considered as column vectors, the scalar product can be written in matrix notation as

$$(\mathbf{u}, \mathbf{v}) \equiv \mathbf{u}^* \mathbf{v}.$$

Definition 1.2. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ are called *orthogonal* if

$$(\mathbf{u}, \mathbf{v}) = 0.$$

Example 1.1. Let

$$\mathbf{u} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Then $(\mathbf{u}, \mathbf{v}) = 0$. ♣

The scalar product induces a *norm* of $\mathbf{u}$ defined by

$$\|\mathbf{u}\| := \sqrt{(\mathbf{u},\mathbf{u})}.$$

In section 1.14 a detailed discussion of norms is given.

Definition 1.3. A vector $\mathbf{u} \in \mathbb{C}^n$ is called *normalized* if $(\mathbf{u},\mathbf{u}) = 1$.

Example 1.2. The vectors

$$\mathbf{u} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

are normalized and form an orthonormal basis in the vector space $\mathbb{R}^2$. ♣

Let V and W be two vector spaces over the same field, equipped with bases $(\mathbf{e}_j)_{j=1}^n$ and $(\mathbf{f}_i)_{i=1}^m$, respectively. Relative to these bases, a linear transformation

$$\mathcal{A} : V \to W$$

is represented by the matrix having m rows and n columns

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The elements a_{ij} of the matrix A are defined uniquely by the relations

$$\mathcal{A}\mathbf{e}_j = \sum_{i=1}^{m} a_{ij}\mathbf{f}_i, \qquad j = 1, 2, \ldots, n.$$

Equivalently, the jth column vector

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

of the matrix A represents the vector $\mathcal{A}e_j$ relative to the basis $(\mathbf{f}_i)_{i=1}^m$. We call

$$(a_{i1}\, a_{i2}\, \ldots\, a_{in})$$

the ith row vector of the matrix A. A matrix with m rows and n columns is called a matrix of type (m,n), and the vector space over the field $\mathbb{F}$ consisting of matrices of type (m,n) with elements in $\mathbb{F}$ is denoted by $\mathcal{A}_{m,n}$.

A column vector is then a matrix of type $(m, 1)$ and a row vector a matrix of type $(1, n)$. A matrix is called real or complex according whether its elements are in the field $\mathbb{R}$ or the field $\mathbb{C}$. A matrix A with elements a_{ij} is written as

$$A = (a_{ij})$$

the first index i always designating the row and the second, j, the column.

Definition 1.4. A matrix with all its elements 0 is called the *zero matrix* or *null matrix*.

Definition 1.5. Given a matrix $A \in \mathcal{A}_{m,n}(\mathbb{C})$, the matrix $A^* \in \mathcal{A}_{n,m}(\mathbb{C})$ denotes the *adjoint* of the matrix A and is defined uniquely by the relations

$$(A\mathbf{u}, \mathbf{v})_m = (\mathbf{u}, A^*\mathbf{v})_n \quad \text{for every } \mathbf{u} \in \mathbb{C}^n, \quad \mathbf{v} \in \mathbb{C}^m$$

which imply that $(A^*)_{ij} = \bar{a}_{ji}$.

Definition 1.6. Given a matrix $A = \mathcal{A}_{m,n}(\mathbb{R})$, the matrix $A^T \in \mathcal{A}_{n,m}(\mathbb{R})$ denotes the *transpose* of a matrix A and is defined uniquely by the relations

$$(A\mathbf{u}, \mathbf{v})_m = (\mathbf{u}, A^T\mathbf{v})_n \text{ for every } \mathbf{u} \in \mathbb{R}^n, \quad \mathbf{v} \in \mathbb{R}^m$$

which imply that $(A^T)_{ij} = a_{ji}$.

To the composition of linear transformations there corresponds the multiplication of matrices.

Definition 1.7. If $A = (a_{ij})$ is a matrix of type (m, l) and $B = (b_{kj})$ of type (l, n), their *matrix product* AB is the matrix of type (m, n) defined by

$$(AB)_{ij} = \sum_{k=1}^{l} a_{ik} b_{kj}.$$

We have

$$(AB)^T = B^T A^T, \qquad (AB)^* = B^* A^*.$$

Note that $AB \neq BA$, in general, where A and B are $n \times n$ matrices.

Definition 1.8. A matrix of type (n, n) is said to be *square*, or a matrix of *order* n if it is desired to make explicit the integer n; it is convenient to speak of a matrix as rectangular if it is not necessarily square.

Definition 1.9. If $A = (a_{ij})$ is a square matrix, the elements a_{ii} are called *diagonal elements*, and the elements $a_{ij}, i \neq j$, are called *off-diagonal elements*.

Definition 1.10. The *identity matrix* (also called *unit matrix*) is the square matrix

$$I := (\delta_{ij}).$$

Definition 1.11. A square matrix A is *invertible* if there exists a matrix (which is unique, if it does exist), written as A^{-1} and called the *inverse* of the matrix A, which satisfies

$$AA^{-1} = A^{-1}A = I.$$

Otherwise, the matrix is said to be *singular*.

Recall that if A and B are invertible matrices then

$$(AB)^{-1} = B^{-1}A^{-1}, \qquad (A^T)^{-1} = (A^{-1})^T, \qquad (A^*)^{-1} = (A^{-1})^*.$$

Definition 1.12. A square matrix A is *symmetric* if A is real and $A = A^T$.

The sum of two symmetric matrices is again a symmetric matrix.

Definition 1.13. A square matrix A is *skew-symmetric* if A is real and $A = -A^T$.

Every square matrix A over $\mathbb{R}$ can be written as sum of a symmetric matrix S and a skew-symmetric matrix T, i.e. $A = T + S$. Thus

$$S = \frac{1}{2}(A + A^T), \qquad T = \frac{1}{2}(A - A^T)\,.$$

Definition 1.14. A square matrix A over $\mathbb{C}$ is *Hermitian* if $A = A^*$.

The sum of two Hermitian matrices is again a Hermitian matrix.

Definition 1.15. A square matrix A over $\mathbb{C}$ is *skew-Hermitian* if $A = -A^*$.

The sum of two skew-Hermitian matrices is again a skew-Hermitian matrix.

Definition 1.16. A Hermitian $n \times n$ matrix A is *positive semidefinite* if

$$\mathbf{u}^* A\mathbf{x} \geq 0$$

for all nonzero $\mathbf{u} \in \mathbb{C}^n$.

Example 1.3. The 2×2 matrix

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

is positive semidefinite. ♣

Let B be an arbitrary $m \times n$ matrix over $\mathbb{C}$. Then the $n \times n$ matrix B^*B is positive semidefinite.

Definition 1.17. A square matrix A is *orthogonal* if A is real and

$$AA^T = A^TA = I.$$

Thus for an orthogonal matrix A we have $A^{-1} = A^T$. The product of two orthogonal matrices is again an orthogonal matrix. The inverse of an orthogonal matrix is again an orthogonal matrix. The orthogonal matrices form a group under matrix multiplication.

Definition 1.18. A square matrix A is *unitary* if $AA^* = A^*A = I$.

The product of two unitary matrices is again a unitary matrix. The inverse of a unitary matrix is again a unitary matrix. The unitary matrices form a group under matrix multiplication.

Example 1.4. Consider the 2×2 matrix

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

The matrix σ_y is Hermitian and unitary. We have $\sigma_y^* = \sigma_y$ and $\sigma_y^* = \sigma_y^{-1}$. Furthermore $\sigma_y^2 = I_2$. The matrix is one of the Pauli spin matrices. ♣

Definition 1.19. A square matrix is *normal* if $AA^* = A^*A$.

Example 1.5. The matrix

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

is normal, whereas

$$B = \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix}$$

is not a normal matrix. Note that B^*B is normal. ♣

Normal matrices include diagonal, real symmetric, real skew-symmetric, orthogonal, Hermitian, skew-Hermitian, and unitary matrices.

Definition 1.20. A matrix $A = (a_{ij})$ is *diagonal* if $a_{ij} = 0$ for $i \neq j$ and is written as

$$A = \mathrm{diag}(a_{ii}) = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

The matrix product of two $n \times n$ diagonal matrices is again a diagonal matrix.

Definition 1.21. Let $A = (a_{ij})$ be an $m \times n$ matrix over a field $\mathbb{F}$. The columns of A generate a subspace of $\mathbb{F}^m$, whose dimension is called the *column rank* of A. The rows generate a subspace of $\mathbb{F}^n$ whose dimension is called the *row rank* of A. In other words: the column rank of A is the maximum number of linearly independent columns, and the row rank is the maximum number of linearly independent rows. The row rank and the column rank of A are equal to the same number r. Thus r is simply called the *rank* of the matrix A.

Example 1.6. The rank of the 2×3 matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

is $r(A) = 2$. The rank of the matrix product of two matrices cannot exceed the rank of either factors. ♣

Definition 1.22. The *kernel* (or null space) of an $m \times n$ matrix A is the subspace of vectors $\mathbf{x}$ in $\mathbb{C}^n$ for which $A\mathbf{x} = \mathbf{0}$. The dimension of this subspace is the *nullity* of A.

Example 1.7. Consider the matrix

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

Then from the linear equation $A\mathbf{x} = \mathbf{0}$ we obtain $x_1 + x_2 = 0$. The null space of A is the set of solutions to this equation, i.e. a line through the origin of $\mathbb{R}^2$. The nullity of A is equal to 1. ♣

Definition 1.23. Let A, B be $n \times n$ matrices. Then the *commutator* of A and B is defined by

$$[A, B] := AB - BA.$$

Obviously we have $[A, B] = -[B, A]$ and if C is another $n \times n$ matrix

$$[A, B + C] = [A, B] + [A, C].$$

Let A, B be $n \times n$ matrices. Then the *anticommutator* of A and B is defined by

$$[A, B]_+ := AB + BA.$$

Exercises. (1) Let A, B be $n \times n$ upper triangular matrices. Can we conclude that $AB = BA$?

(2) Let A be an arbitrary $n \times n$ matrix. Let B be a diagonal matrix. Is $AB = BA$?

(3) Let A be a normal matrix and U be a unitary matrix. Show that $U^* AU$ is a normal matrix.

(4) Show that the following operations, called elementary transformations, on a matrix do not change its rank:
(i) The interchange of the i-th and j-th rows.
(ii) The interchange of the i-th and j-th columns.

(5) Let A and B be two square matrices of the same order. Is it possible to have $AB + BA = 0$?

(6) Let $A_k,\ 1 \leq k \leq m$, be matrices of order n satisfying

$$\sum_{k=1}^{m} A_k = I.$$

Show that the following conditions are equivalent

$$\begin{aligned} &\text{(i)} \quad A_k = (A_k)^2, \quad 1 \leq k \leq m \\ &\text{(ii)} \quad A_k A_l = 0 \text{ for } k \neq l, \quad 1 \leq k, l \leq m \\ &\text{(iii)} \quad \sum_{k=1}^{m} r(A_k) = n \end{aligned}$$

where $r(A)$ denotes the rank of the matrix A.

(7) Prove that if A is of order $m \times n$, B is of order $n \times p$ and C is of order $p \times q$, then

$$A(BC) = (AB)C.$$

(8) Let S be an invertible $n \times n$ matrix. Let A be an arbitrary $n \times n$ matrix and $\widetilde{A} = SAS^{-1}$. Show that $\widetilde{A}^2 = SA^2S^{-1}$.

1.2 Matrix Operations

Let $\mathbb{F}$ be a field, for example the set of real numbers $\mathbb{R}$ or the set of complex numbers $\mathbb{C}$. Let m, n be two integers ≥ 1. An array A of numbers in $\mathbb{F}$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} = (a_{ij})$$

is called an $m \times n$ *matrix* with entry a_{ij} in the ith row and jth column. A *row vector* is a $1 \times n$ matrix. A *column vector* is an $n \times 1$ matrix. We have a *zero matrix*, in which $a_{ij} = 0$ for all i, j.

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. We define $A + B$ to be the $m \times n$ matrix whose entry in the i-th row and j-th column is $a_{ij} + b_{ij}$. The $m \times n$ matrices over a field $\mathbb{F}$ form a vector space.

Matrix multiplication is only defined between two matrices if the number of columns of the first matrix is the same as the number of rows of the second matrix. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the matrix product AB is an $m \times p$ matrix defined by

$$(AB)_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}$$

for each pair i and j, where $(AB)_{ij}$ denotes the (i, j)th entry in AB.

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices with entries in some field. Then their *Hadamard product* is the entry-wise product of A and B, that is the $m \times n$ matrix $A \bullet B$ whose (i, j)th entry is $a_{ij} b_{ij}$.

Example 1.8. Let

$$A = \begin{pmatrix} 1 & i \\ -1 & i \end{pmatrix}, \quad B = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$A + B = \begin{pmatrix} 3 & i - 1 \\ 0 & i \end{pmatrix}$$

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

$$AB = \begin{pmatrix} i + 2 & -1 \\ i - 2 & 1 \end{pmatrix}.$$

♣

1.3 Linear Equations

Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Let $b_1, \ldots, b_m$ be elements of the field $\mathbb{F}$. The system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \quad &\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is called a system of linear equations. We also write $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x}$ and $\mathbf{b}$ are considered as column vectors. The system is said to be homogeneous if all the numbers $b_1, \ldots, b_m$ are equal to 0. The number n is called the number of unknowns, and m is called the number of equations. The system of homogeneous equations also admits the trivial solution

$$x_1 = x_2 = \cdots = x_n = 0.$$

A system of homogeneous equations of m linear equations in n unknowns with $n > m$ admits a nontrivial solution. An under determined linear system is either inconsistent or has infinitely many solutions.

An important special case is $m = n$. Then for the system of linear equations $A\mathbf{x} = \mathbf{b}$ we investigate the cases A^{-1} exists and A^{-1} does not exist. If A^{-1} exists we can write the solution as

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

If $m > n$, then we have an overdetermined system and it can happen that no solution exists. One solves these problems in the least-square sense.

Example 1.9. Consider the system of linear equations

$$\begin{aligned} 3x_1 - x_2 &= -1 \\ x_2 - x_3 &= 0 \\ x_1 + x_2 + x_3 &= 1. \end{aligned}$$

These equations have the matrix representation

$$\begin{pmatrix} 3 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

with solution $x_1 = -\frac{1}{7}$, $x_2 = \frac{4}{7}$ and $x_3 = \frac{4}{7}$. ♣

1.4 Trace and Determinant

In this section we introduce the trace and determinant of a $n \times n$ matrix and summarize their properties.

Definition 1.24. The *trace* of a square matrix $A = (a_{jk})$ of order n is defined as the sum of its diagonal elements

$$\operatorname{tr}(A) := \sum_{j=1}^{n} a_{jj}.$$

Example 1.10. Let

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

Then $\operatorname{tr}(A) = 0$. ♣

The properties of the trace are as follows. Let $a, b \in \mathbb{C}$ and let A, B and C be three $n \times n$ matrices. Then

$$\begin{aligned} \operatorname{tr}(aA + bB) &= a\operatorname{tr}(A) + b\operatorname{tr}(B) \\ \operatorname{tr}(A^T) &= \operatorname{tr}(A) \\ \operatorname{tr}(AB) &= \operatorname{tr}(BA) \\ \operatorname{tr}(A) &= \operatorname{tr}(S^{-1}AS) \qquad S \text{ nonsingular } n \times n \text{ matrix} \\ \operatorname{tr}(A^*A) &= \operatorname{tr}(AA^*) \\ \operatorname{tr}(ABC) &= \operatorname{tr}(CAB) = \operatorname{tr}(BCA). \end{aligned}$$

Thus the trace is a linear functional. From the third property we find that

$$\operatorname{tr}([A, B]) = 0$$

where $[\,,\,]$ denotes the commutator. The last property is called the *cyclic invariance* of the trace. Notice, however, that

$$\operatorname{tr}(ABC) \neq \operatorname{tr}(BAC)$$

in general. An example is given by the following matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

We have $\operatorname{tr}(ABC) = 1$ but $\operatorname{tr}(BAC) = 0$.

If λ_j, $j = 1, 2, \ldots, n$ are the eigenvalues of the $n \times n$ matrix A (see section 1.5), then

$$\operatorname{tr}(A) = \sum_{j=1}^{n} \lambda_j, \qquad \operatorname{tr}(A^2) = \sum_{j=1}^{n} \lambda_j^2 .$$

More generally, if p designates a polynomial of degree r

$$p(x) = \sum_{j=0}^{r} a_j x^j$$

then

$$\operatorname{tr}(p(A)) = \sum_{k=1}^{n} p(\lambda_k).$$

Moreover we find

$$\operatorname{tr}(AA^*) = \operatorname{tr}(A^*A) = \sum_{j,k=1}^{n} |a_{jk}|^2 \geq 0.$$

Thus $\sqrt{\operatorname{tr}(AA^*)}$ is a norm of A.

Example 1.11. Let

$$A = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} .$$

Then $AA^* = I_2$, where I_2 is the 2×2 identity matrix. Thus $\|A\| = \sqrt{2}$. ♣

Let $\mathbf{x}$, $\mathbf{y}$ be column vectors in $\mathbb{R}^n$. Then

$$\mathbf{x}^T \mathbf{y} = \operatorname{tr}(\mathbf{x}\mathbf{y}^T) = \operatorname{tr}(\mathbf{y}\mathbf{x}^T) .$$

Let A an $n \times n$ matrix over $\mathbb{R}$. Then we have

$$\mathbf{x}^T A \mathbf{y} = \operatorname{tr}(A\mathbf{y}\mathbf{x}^T) .$$

Next we introduce the definition of the *determinant* of an $n \times n$ matrix. Then we give the properties of the determinant.

Definition 1.25. The *determinant* of an $n \times n$ matrix A is a scalar quantity denoted by $\det(A)$ and is given by

$$\det(A) := \sum_{j_1, j_2, \ldots, j_n} p(j_1, j_2, \ldots, j_n) a_{1j_1} a_{2j_2} \ldots a_{nj_n}$$

where $p(j_1, j_2, \ldots, j_n)$ is a permutation equal to ± 1 and the summation extends over $n!$ permutations $j_1, j_2, \ldots, j_n$ of the integers $1, 2, \ldots, n$. For an $n \times n$ matrix there exist $n!$ permutations. Therefore

$$p(j_1, j_2, \ldots, j_n) = \operatorname{sign} \prod_{1 \leq s < r \leq n} (j_r - j_s).$$

Example 1.12. For a matrix of order (3,3) we find

$$p(1,2,3) = 1, \quad p(1,3,2) = -1, \; p(3,1,2) = 1$$
$$p(3,2,1) = -1, \; p(2,3,1) = 1, \quad p(2,1,3) = -1.$$

Then the determinant for a 3×3 matrix is given by

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32}$$
$$-a_{13}a_{22}a_{31} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33}.$$

♣

Definition 1.26. We call a square matrix A a *nonsingular* matrix if

$$\det(A) \neq 0$$

whereas if $\det(A) = 0$ the matrix A is called a *singular* matrix.

If $\det(A) \neq 0$, then A^{-1} exists. Conversely, if A^{-1} exists, then $\det(A) \neq 0$.

Example 1.13. The matrix

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

is nonsingular since its determinant is -1, and the matrix

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

is singular since its determinant is 0. ♣

Next we list some properties of determinants.

1. Let A be an $n \times n$ matrix and A^T the transpose. Then

$$\det(A) = \det(A^T).$$

Example 1.14.

$$\det \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ -1 & 0 & 2 \end{pmatrix} = 5.$$

♣

Remark. Let

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

Then

$$A^T = \begin{pmatrix} 0 & 1 \\ -i & 0 \end{pmatrix}, \qquad A^* \equiv \bar{A}^T = \begin{pmatrix} 0 & 1 \\ i & 0 \end{pmatrix}.$$

Obviously

$$\det(A) \neq \det(A^*).$$

2. Let A be an $n \times n$ matrix and $\alpha \in \mathbb{R}$. Then

$$\det(\alpha A) = \alpha^n \det(A).$$

3. Let A be an $n \times n$ matrix. If two adjacent columns are equal, i.e. $A_j = A_{j+1}$ for some $j = 1, 2, \ldots, n-1$, then $\det(A) = 0$.

4. If any vector in A is a zero vector then $\det(A) = 0$.

5. Let A be an $n \times n$ matrix. Let j be some integer, $1 \leq j < n$. If the j-th and $(j+1)$-th columns are interchanged, then the determinant changes by a sign.

6. Let $A_1, \ldots, A_n$ be the column vectors of an $n \times n$ matrix A. If they are linearly dependent, then $\det(A) = 0$.

7. Let A and B be $n \times n$ matrices. Then

$$\det(AB) = \det(A)\det(B).$$

8. Let A be an $n \times n$ diagonal matrix. Then

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

9. $(d/dt)\det(A(t))$ = sum of the determinants where each of them is obtained by differentiating the rows of A with respect to t one at a time, then taking its determinant.

Proof. Since

$$\det(A(t)) = \sum_{j_1,\ldots,j_n} p(j_1,\ldots,j_n) a_{1j_1}(t)\cdots a_{nj_n}(t)$$

we find

$$\frac{d}{dt}\det(A(t)) = \sum_{j_1,\ldots,j_n} p(j_1,\ldots,j_n)\frac{da_{1j_1}(t)}{dt}a_{2j_2}(t)\cdots a_{nj_n}(t) + \\ + \sum_{j_1,\ldots,j_n} p(j_1,\ldots,j_n)a_{1j_1}(t)\frac{da_{2j_2}(t)}{dt}\cdots a_{nj_n}(t) + \cdots \\ + \sum_{j_1,\ldots,j_n} p(j_1,\ldots,j_n)a_{1j_1}(t)\cdots a_{n-1j_{n-1}}(t)\frac{da_{nj_n}(t)}{dt}.$$

□

Example 1.15. We have

$$\frac{d}{dt}\det\begin{pmatrix} e^t & \cos t \\ 1 & \sin t^2 \end{pmatrix} = \det\begin{pmatrix} e^t & -\sin t \\ 1 & \sin t^2 \end{pmatrix} + \det\begin{pmatrix} e^t & \cos t \\ 0 & 2t\cos t^2 \end{pmatrix}.$$

♣

10. Let A be an invertible $n \times n$ symmetric matrix over $\mathbb{R}$. Then

$$\mathbf{v}^T A^{-1}\mathbf{v} = \frac{\det(A + \mathbf{v}\mathbf{v}^T)}{\det(A)} - 1$$

for every vector $\mathbf{v} \in \mathbb{R}^n$.

Example 1.16. Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then $A^{-1} = A$ and therefore

$$\mathbf{v}^T A^{-1}\mathbf{v} = 2.$$

Since

$$\mathbf{v}\mathbf{v}^T = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and $\det(A) = -1$ we obtain

$$\frac{\det(A + \mathbf{v}\mathbf{v}^T)}{\det(A)} - 1 = 2.$$

♣

11. Let A be an $n \times n$ matrix. Then

$$\det(\exp(A)) \equiv \exp(\mathrm{tr}(A)).$$

12. The determinant of a diagonal matrix or triangular matrix is the product of its diagonal elements.

13. Let A be an $n \times n$ matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A (see section 1.5). Then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n .$$

14. Let A be a Hermitian matrix. Then $\det(A)$ is a real number.

15. Let U be a unitary matrix. Then

$$\det(U) = e^{i\phi}$$

for some $\phi \in \mathbb{R}$. Thus $|\det(U)| = 1$.

16. Let A, B, C be $n \times n$ matrices and let 0 be the $n \times n$ zero matrix. Then

$$\det \begin{pmatrix} A & 0 \\ C & B \end{pmatrix} = \det(A)\det(B).$$

17. The determinant of the matrix

$$A_n := \begin{pmatrix} b_1 & a_2 & 0 & \ldots & 0 & 0 \\ -1 & b_2 & a_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & b_{n-1} & a_n \\ 0 & 0 & 0 & \ldots & -1 & b_n \end{pmatrix}, \qquad n = 1, 2, \ldots$$

satisfies the recursion relation

$$\det(A_n) = b_n \det(A_{n-1}) + a_n \det(A_{n-2}), \quad \det(A_0) = 1, \quad \det(A_1) = b_1$$

where $n = 2, 3, \ldots$.

18. Let A be a 2×2 matrix. Then

$$\det(I_2 + A) \equiv 1 + \operatorname{tr}(A) + \det(A).$$

19. Let A be an invertible $n \times n$ matrix, i.e. $\det(A) \neq 0$. Then the inverse of A can be calculated as

$$(A)^{-1}_{kj} = \frac{\partial}{\partial (A)_{jk}} \ln(\det(A)).$$

Exercises. (1) Let X and Y be $n \times n$ matrices over $\mathbb{R}$. Show that

$$(X, Y) := \operatorname{tr}(XY^T)$$

defines a scalar product, i.e. prove that $(X, X) \geq 0$, $(X, Y) = (Y, X)$, $(cX, Y) = c(X, Y)$ $(c \in \mathbb{R})$, $(X + Y, Z) = (X, Z) + (Y, Z)$.

(2) Let A and B be $n \times n$ matrices. Show that

$$\operatorname{tr}([A, B]) = 0$$

where $[\,,\,]$ denotes the commutator (i.e. $[A, B] := AB - BA$).

(3) Use (2) to show that the relation

$$[A, B] = \lambda I, \qquad \lambda \in \mathbb{C}$$

for finite dimensional matrices can only be satisfied if $\lambda = 0$. For certain infinite dimensional matrices A and B we can find a nonzero λ.

(4) Let A and B be $n \times n$ matrices. Suppose that AB is nonsingular. Show that A and B are nonsingular matrices.

(5) Let A and B be $n \times n$ matrices over $\mathbb{R}$. Assume that A is skew-symmetric, i.e. $A^T = -A$. Assume that n is odd. Show that $\det(A) = 0$.

(6) Let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

be a square matrix partitioned into blocks. Assuming the submatrix A_{11} to be invertible, show that

$$\det(A) = \det(A_{11}) \det(A_{22} - A_{21} A_{11}^{-1} A_{12}).$$

(7) A square matrix A for which $A^n = 0$, where n is a positive integer, is called *nilpotent*. Let A be a nilpotent matrix. Show that $\det(A) = 0$.

(8) Let A be an $n \times n$ skew-symmetric matrix over $\mathbb{R}$. Show that if n is odd then $\det(A) = 0$. Hint. Apply $\det(A) = (-1)^n \det(A)$.

(9) Let A be an $n \times n$ matrix with $A^2 = I_n$. Calculate $\det(A)$.

1.5 Eigenvalue Problem

The eigenvalue problem plays a central role in theoretical and mathematical physics (Steeb [60; 61]). We give a short introduction into the eigenvalue calculation for finite dimensional matrices. In section 2.6 we study the eigenvalue problem for Kronecker products of matrices.

Definition 1.27. A complex number λ is said to be an *eigenvalue* (or *characteristic value*) of an $n \times n$ matrix A, if there is at least one nonzero vector $\mathbf{u} \in \mathbb{C}^n$ satisfying the *eigenvalue equation*

$$A\mathbf{u} = \lambda\mathbf{u}, \qquad \mathbf{u} \neq \mathbf{0}.$$

Each nonzero vector $\mathbf{u} \in \mathbb{C}^n$ satisfying the eigenvalue equation is called an *eigenvector* (or *characteristic vector*) of A with eigenvalue λ.

The eigenvalue equation can be written as

$$(A - \lambda I)\mathbf{u} = \mathbf{0}$$

where I is the $n \times n$ unit matrix and $\mathbf{0}$ is the zero vector.

This system of n linear simultaneous equations in $\mathbf{u}$ has a nontrivial solution for the vector $\mathbf{u}$ only if the matrix $(A - \lambda I)$ is singular, i.e.

$$\det(A - \lambda I) = 0.$$

The expansion of the determinant gives a polynomial in λ of degree equal to n, which is called the characteristic polynomial of the matrix A. The n roots of the equation $\det(A - \lambda I) = 0$, called the *characteristic equation*, are the eigenvalues of A.

Definition 1.28. Let λ be an *eigenvalue* of an $n \times n$ matrix A. The vector $\mathbf{u}$ is a *generalized eigenvector* of A corresponding to λ if

$$(A - \lambda I)^n \mathbf{u} = \mathbf{0}.$$

The eigenvectors of a matrix are also generalized eigenvectors of the matrix.

Theorem 1.1. *Every $n \times n$ matrix A has at least one eigenvalue and corresponding eigenvector.*

Proof. We follow the proof in Axler [2]. Suppose $\mathbf{v} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$. Then $\{\mathbf{v}, A\mathbf{v}, \ldots, A^n\mathbf{v}\}$ must be a linearly dependent set of vectors, i.e. there exist $c_0, \ldots, c_n \in \mathbb{C}$ such that

$$c_0\mathbf{v} + c_1 A\mathbf{v} + \cdots + c_n A^n \mathbf{v} = \mathbf{0}.$$

Let $m \in \{0, 1, \ldots, n\}$ be the largest index satisfying $c_m \neq 0$. Consider the following polynomial in x and its factorization over $\mathbb{C}$:

$$c_0 + c_1 x + \cdots + c_m x^m = c_m (x - x_0)(x - x_1) \cdots (x - x_m)$$

for some $x_0, \ldots, x_m \in \mathbb{C}$ (i.e. the roots of the polynomial). Then

$$c_0 \mathbf{v} + c_1 A\mathbf{v} + \cdots + c_n A^n \mathbf{v} = c_m (A - x_0 I_n)(A - x_1 I_n) \cdots (A - x_m I_n)\mathbf{v} = \mathbf{0}.$$

It follows that there is a largest $j \in \{0, 1, \ldots, m\}$ satisfying

$$(A - x_j I_n)\left[(A - x_{j+1} I_n) \cdots (A - x_m I_n)\mathbf{v}\right] = \mathbf{0}.$$

This is a solution to the eigenvalue equation, the eigenvalue is x_j and the corresponding eigenvector is

$$(A - x_{j+1} I_n) \cdots (A - x_m I_n)\mathbf{v}.$$

$\square$

Definition 1.29. The *spectrum* of the matrix A is the subset

$$\mathrm{sp}(A) := \bigcup_{i=1}^{n} \{ \lambda_i(A) \}$$

of the complex plane. The *spectral radius* of the matrix A is the nonnegative number defined by

$$\varrho(A) := \max\{ |\lambda_j(A)| \, : \, 1 \leq j \leq n \}.$$

If $\lambda \in \mathrm{sp}(A)$, the vector subspace

$$\{ \mathbf{v} \in V \, : \, A\mathbf{v} = \lambda \mathbf{v} \}$$

(of dimension at least 1) is called the *eigenspace* corresponding to the eigenvalue λ.

Example 1.17. Let

$$A = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Then

$$\det(A - \lambda I_2) \equiv \lambda^2 - 1 = 0.$$

Therefore the eigenvalues are given by $\lambda_1 = 1$, $\lambda_2 = -1$. To find the eigenvector of the eigenvalue $\lambda_1 = 1$ we have to solve

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 1 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Therefore $u_2 = iu_1$ and the eigenvector of $\lambda_1 = 1$ is given by

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

For $\lambda_2 = -1$ we have

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = -1 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

and hence

$$\mathbf{u}_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

We see that $(\mathbf{u}_1, \mathbf{u}_2) \equiv \mathbf{u}_2^* \mathbf{u}_1 = 0$. Both eigenvectors are not normalized. ♣

A special role in theoretical physics is played by the Hermitian matrices. In this case we have the following theorem.

Theorem 1.2. *Let A be a Hermitian matrix, i.e. $A^* = A$, where $A^* \equiv \bar{A}^T$. The eigenvalues of A are real, and two eigenvectors corresponding to two different eigenvalues are mutually orthogonal.*

Proof. The eigenvalue equation is $A\mathbf{u} = \lambda\mathbf{u}$, where $\mathbf{u} \neq \mathbf{0}$. Now we have the identity

$$(A\mathbf{u})^*\mathbf{u} \equiv \mathbf{u}^* A^* \mathbf{u} \equiv \mathbf{u}^*(A^*\mathbf{u}) \equiv \mathbf{u}^*(A\mathbf{u})$$

since A is Hermitian, i.e. $A = A^*$. Inserting the eigenvalue equation into this equation yields

$$(\lambda\mathbf{u})^*\mathbf{u} = \mathbf{u}^*(\lambda\mathbf{u})B \quad \text{or} \quad \bar{\lambda}(\mathbf{u}^*\mathbf{u}) = \lambda(\mathbf{u}^*\mathbf{u}).$$

Since $\mathbf{u}^*\mathbf{u} \neq 0$, we have $\bar{\lambda} = \lambda$ and therefore λ must be real. Let

$$A\mathbf{u}_1 = \lambda_1\mathbf{u}_1, \qquad A\mathbf{u}_2 = \lambda_2\mathbf{u}_2.$$

Now

$$\lambda_1(\mathbf{u}_1, \mathbf{u}_2) = (\lambda_1\mathbf{u}_1, \mathbf{u}_2) = (A\mathbf{u}_1, \mathbf{u}_2) = (\mathbf{u}_1, A\mathbf{u}_2) = (\mathbf{u}_1, \lambda_2\mathbf{u}_2) = \lambda_2(\mathbf{u}_1, \mathbf{u}_2).$$

Since $\lambda_1 \neq \lambda_2$, we find that $(\mathbf{u}_1, \mathbf{u}_2) = 0$. □

Theorem 1.3. *The eigenvalues λ_j of a unitary matrix U satisfy $|\lambda_j| = 1$.*

Proof. Since U is a unitary matrix we have $U^* = U^{-1}$, where U^{-1} is the inverse of U. Let

$$U\mathbf{u} = \lambda\mathbf{u}$$

be the eigenvalue equation. It follows that

$$(U\mathbf{u})^* = (\lambda\mathbf{u})^* \quad \text{or} \quad \mathbf{u}^*U^* = \bar{\lambda}\mathbf{u}^* .$$

Thus we obtain

$$\mathbf{u}^*U^*U\mathbf{u} = \bar{\lambda}\lambda\mathbf{u}^*\mathbf{u} .$$

Owing to $U^*U = I$ we obtain

$$\mathbf{u}^*\mathbf{u} = \bar{\lambda}\lambda\mathbf{u}^*\mathbf{u}.$$

Since $\mathbf{u}^*\mathbf{u} \neq 0$ we have $\bar{\lambda}\lambda = 1$. Thus the eigenvalue λ can be written as

$$\lambda = \exp(i\alpha), \qquad \alpha \in \mathbb{R} .$$

Thus $|\lambda| = 1$. $\square$

Theorem 1.4. *If* $\mathbf{x}$ *is an eigenvalue of the* $n \times n$ *normal matrix* A *corresponding to the eigenvalue* λ*, then* $\mathbf{x}$ *is also an eigenvalue of* A^* *corresponding to the eigenvalue* $\overline{\lambda}$.

Proof. Since $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$ and $AA^* = A^*A$ we find

$$\begin{aligned}((A^* - \overline{\lambda}I_n)\mathbf{x}, (A^* - \overline{\lambda}I_n)\mathbf{x}) &= \left[(A^* - \overline{\lambda}I_n)\mathbf{x}\right]^* \left[(A^* - \overline{\lambda}I_n)\mathbf{x}\right] \\ &= \mathbf{x}^*(A - \lambda I_n)(A^* - \overline{\lambda}I_n)\mathbf{x} \\ &= \mathbf{x}^*(A^* - \overline{\lambda}I_n)(A - \lambda I_n)\mathbf{x} \\ &= \mathbf{x}^*(A^* - \overline{\lambda}I_n)\mathbf{0} = 0\end{aligned}$$

Consequently $(A^* - \overline{\lambda}I_n)\mathbf{x} = \mathbf{0}$. $\square$

Theorem 1.5. *The eigenvalues of a skew-Hermitian matrix* $(A^* = -A)$ *can only be 0 or (purely) imaginary.*

The proof is left as an exercise to the reader.

Example 1.18. The skew-Hermitian matrix

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

has the eigenvalues $\pm i$. ♣

Now consider the general case. Let λ be an eigenvalue of A with the corresponding eigenvector $\mathbf{x}$, and let μ be an eigenvalue of A^* with the corresponding eigenvector $\mathbf{y}$. Then

$$\mathbf{x}^* A^* \mathbf{y} = (\mathbf{x}^* A^*)\mathbf{y} = (A\mathbf{x})^* \mathbf{y} = (\lambda \mathbf{x})^* \mathbf{y} = \overline{\lambda} \mathbf{x}^* \mathbf{y}$$

and

$$\mathbf{x}^* A^* \mathbf{y} = \mathbf{x}^* (A^* \mathbf{y}) = \mu \mathbf{x}^* \mathbf{y}.$$

It follows that $\mathbf{x}^* \mathbf{y} = 0$ or $\overline{\lambda} = \mu$.

Example 1.19. Consider

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

Both eigenvalues of A are zero. We have

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

Both eigenvalues of A^* are also zero. The eigenspaces corresponding to the eigenvalue 0 of A and A^* are

$$\left\{ \begin{pmatrix} t \\ 0 \end{pmatrix} : t \in \mathbb{C}, \, t \neq 0 \right\}$$

and

$$\left\{ \begin{pmatrix} 0 \\ t \end{pmatrix} : t \in \mathbb{C}, \, t \neq 0 \right\}$$

respectively. Obviously both conditions above are true. The eigenvalues of A^*A and AA^* are given by 0 and 1. ♣

Exercises. (1) Let A be an $n \times n$ matrix over $\mathbb{C}$. Show that the eigenvectors of A corresponding to distinct eigenvalues are linearly independent.

(2) Show that

$$\mathrm{tr}(A) = \sum_{j=1}^{n} \lambda_j(A), \qquad \det(A) = \prod_{j=1}^{n} \lambda_j(A).$$

(3) Let U be a Hermitian and unitary matrix. What can be said about the eigenvalues of U?

(4) Let A be an invertible matrix whose elements, as well as those of A^{-1}, are all nonnegative. Show that there exists a permutation matrix P and a matrix $D = \mathrm{diag}\,(d_j)$, with d_j positive, such that $A = PD$ (the converse is obvious).

(5) Let A and B be two square matrices of the same order. Show that the matrices AB and BA have the same characteristic polynomial.

(6) Let $a, b, c \in \mathbb{R}$. Find the eigenvalues and eigenvectors of the symmetric 4×4 matrix

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

(7) Let $a_1, a_2, \dots, a_n \in \mathbb{R}$. Show that the eigenvalues of the matrix

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-3} & a_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_3 & a_4 & a_5 & \cdots & a_1 & a_2 \\ a_2 & a_3 & a_4 & \cdots & a_n & a_1 \end{pmatrix}$$

called a *circulant matrix*, are of the form

$$\lambda_{l+1} = a_1 + a_2\xi_l + a_3\xi_l^2 + \cdots + a_n\xi_l^{n-1}, \qquad l = 0, 1, \dots, n-1$$

where $\xi_l := e^{2i\pi l/n}$.

1.6 Cayley-Hamilton Theorem

The Cayley-Hamilton theorem states that the matrix A satisfies its own characteristic equation, i.e.

$$(A - \lambda_1 I_n) \cdots (A - \lambda_n I_n) = 0_{n\times n}$$

where $0_{n\times n}$ is the $n \times n$ zero matrix. Notice that the factors commute.

In this section we follow Axler [2].

Definition 1.30. An $n \times n$ matrix A is *upper triangular* with respect to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathbb{C}^n$, where $m \le n$, if

$$A\mathbf{v}_j \in \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}, \qquad j = 1, \ldots, m.$$

Theorem 1.6. *For every $n \times n$ matrix A there exists a basis V for $\mathbb{C}^n$ such that A is upper triangular with respect to V.*

Proof. Let $\mathbf{v}_1$ be an eigenvector of A. The proof that every matrix A is upper triangular is by induction. The case $n = 1$ is obvious. Consider the subspace

$$U = \{(A - x_j I_n)\mathbf{x} \, : \, \mathbf{x} \in \mathbb{C}^n\}.$$

For all $\mathbf{u} \in U$

$$A\mathbf{u} = (A - x_j I_n)\mathbf{u} + x_j \mathbf{u} \in U$$

since $(A - x_j I_n)\mathbf{u} \in U$ by definition. Since x_j is an eigenvalue of A we have $\det(A - x_j I_n) = 0$ so that $\dim(U) < n$. The induction hypothesis is that any square matrix is upper triangular with respect to a basis for a (sub)space with dimension less than n. Consequently A has a triangular representation on U. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_{\dim U}\}$ be a basis for U and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{C}^n$. We have for $k \in \{\dim U + 1, \ldots, n\}$

$$A\mathbf{v}_k = (A - x_j I_n)\mathbf{v}_k + x_j \mathbf{v}_k \in \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{\dim U}, \mathbf{v}_k\}$$

where

$$\mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{\dim U}, \mathbf{v}_k\} \subseteq \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

and $(A - x_j I_n)\mathbf{v}_k \in U$ by definition. It follows that A is upper triangular with respect to

$$V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}.$$

□

If A has a triangular representation with respect to some basis we can find a triangular representation with respect to an orthonormal basis by applying the Gram-Schmidt orthonormalization process (see section 1.17).

Theorem 1.7. *Every $n \times n$ matrix A satisfies its own characteristic equation*

$$(A - \lambda_1 I_n) \cdots (A - \lambda_n I_n) = 0_{n \times n}$$

where $0_{n\times n}$ is the $n \times n$ zero matrix, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

Proof. Let A be triangular with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{C}^n$ for $\mathbb{C}^n$. Thus $\mathbf{v}_1$ is an eigenvector of A corresponding to an eigenvalue, say λ_{j_1}. Consequently $(A - \lambda_{j_1} I_n)\mathbf{v}_1 = \mathbf{0}$. Now suppose

$$(A - \lambda_{j_1} I_n)(A - \lambda_{j_2} I_n) \cdots (A - \lambda_{j_k} I_n)\mathbf{v}_k = \left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) \mathbf{v}_k = \mathbf{0}$$

for $k = 1, 2, \ldots, r$. Now $A\mathbf{v}_{r+1} \in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r+1}\}$ so that $A\mathbf{v}_{r+1} = \mathbf{u} + \alpha \mathbf{v}_{r+1}$ for some $\mathbf{u} \in \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}, \alpha \in \mathbb{C}$. The supposition above (induction hypothesis) implies

$$\left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) \mathbf{u} = \mathbf{0}$$

so that

$$\left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) A\mathbf{v}_{r+1} = \alpha \left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) \mathbf{v}_{r+1}$$

which simplifies to

$$(A - \alpha I_n) \left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) \mathbf{v}_{r+1} = \mathbf{0}$$

since A commutes with $(A - cI_n)$ $(c \in \mathbb{C})$. Thus either

$$\left(\prod_{p=1}^{k} (A - \lambda_{j_k} I_n) \right) \mathbf{v}_{r+1} = \mathbf{0}$$

or α is an eigenvalue, say $\lambda_{j_{r+1}}$, of A. If the first case holds a re-ordering of the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ postpones the arbitrary choice of $\lambda_{j_{r+1}}$. In either case, we have shown by induction that

$$\left(\prod_{p=1}^{n} (A - \lambda_p I_n) \right) \mathbf{v}_k = \mathbf{0}, \qquad k = 1, 2, \ldots, n.$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis we must have

$$\prod_{p=1}^{n} (A - \lambda_p I_n) = (A - \lambda_1 I_n) \cdots (A - \lambda_n I_n) = 0_{n \times n}.$$

□

1.7 Projection Matrices

First we introduce the definition of a projection matrix and give some of its properties. Projection matrices (projection operators) play a central role in finite group theory in the decomposition of Hilbert spaces into invariant subspaces (Steeb [60; 61; 57]).

Definition 1.31. An $n \times n$ matrix Π is called a *projection matrix* if

$$\Pi = \Pi^*$$

and

$$\Pi^2 = \Pi.$$

The element $\Pi\mathbf{u}$ ($\mathbf{u} \in \mathbb{C}^n$) is called the *projection* of the element $\mathbf{u}$.

Example 1.20. Let $n = 2$ and

$$\Pi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Pi_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then $\Pi_1^* = \Pi_1$, $\Pi_1^2 = \Pi_1$, $\Pi_2^* = \Pi_2$ and $\Pi_2^2 = \Pi_2$. Furthermore $\Pi_1\Pi_2 = 0$ and

$$\Pi_1 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \qquad \Pi_2 \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ u_2 \end{pmatrix}.$$

♣

Theorem 1.8. *Let Π_1 and Π_2 be two $n \times n$ projection matrices. Assume that $\Pi_1\Pi_2 = 0$. Then*

$$(\Pi_1\mathbf{u}, \Pi_2\mathbf{u}) = 0.$$

Proof. We find

$$(\Pi_1\mathbf{u}, \Pi_2\mathbf{u}) = (\Pi_1\mathbf{u})^*(\Pi_1\mathbf{u}) = (\mathbf{u}^*\Pi_1^*)(\Pi_2\mathbf{u}) = \mathbf{u}^*(\Pi_1\Pi_2)\mathbf{u} = 0.$$

□

Theorem 1.9. *Let I_n be the $n \times n$ unit matrix and Π be a projection matrix. Then $I_n - \Pi$ is a projection matrix.*

Proof. Since

$$(I_n - \Pi)^* = I_n^* - \Pi^* = I_n - \Pi$$

and

$$(I_n - \Pi)^2 = (I_n - \Pi)(I_n - \Pi) = I_n - \Pi - \Pi + \Pi = I_n - \Pi$$

we find that $I_n - \Pi$ is a projection matrix. □

Theorem 1.10. *The eigenvalues* λ_j *of a projection matrix* Π *are given by* $\lambda_j \in \{0, 1\}$.

Proof. From the eigenvalue equation $\Pi\mathbf{u} = \lambda\mathbf{u}$ we find

$$\Pi(\Pi\mathbf{u}) = (\Pi\Pi)\mathbf{u} = \lambda\Pi\mathbf{u}.$$

Using the fact that $\Pi^2 = \Pi$ we obtain

$$\Pi\mathbf{u} = \lambda^2\mathbf{u}.$$

Thus $\lambda = \lambda^2$ since $\mathbf{u} \neq \mathbf{0}$ and hence $\lambda \in \{0, 1\}$. $\square$

Theorem 1.11. (Projection Theorem.) *Let* U *be a nonempty, convex, closed subset of the vector space* $\mathbb{C}^n$. *Given any element* $\mathbf{w} \in \mathbb{C}^n$, *there exists a unique element* $\Pi\mathbf{w}$ *such that*

$$\Pi\mathbf{w} \in U \quad \textit{and} \quad \|\mathbf{w} - \Pi\mathbf{w}\| = \inf_{\mathbf{v}\in U} \|\mathbf{w} - \mathbf{v}\|.$$

This element $\Pi\mathbf{w} \in U$ *satisfies*

$$(\Pi\mathbf{w} - \mathbf{w}, \mathbf{v} - \Pi\mathbf{w}) \geq 0 \quad \textit{for every } \mathbf{v} \in U$$

and, conversely, if any element $\mathbf{u}$ *satisfies*

$$\mathbf{u} \in U \quad \textit{and} \quad (\mathbf{u}, \mathbf{v} - \mathbf{u}) \geq 0 \quad \textit{for every } \mathbf{v} \in U$$

then $\mathbf{u} = \Pi\mathbf{w}$. *Furthermore*

$$\|\Pi\mathbf{u} - \Pi\mathbf{v}\| \leq \|\mathbf{u} - \mathbf{v}\|.$$

For the proof refer to Ciarlet [13].

Let $\mathbf{u}$ be a nonzero normalized column vector. Then $\mathbf{u}\mathbf{u}^*$ is a projection matrix, since $(\mathbf{u}\mathbf{u}^*)^* = \mathbf{u}\mathbf{u}^*$ and

$$(\mathbf{u}\mathbf{u}^*)(\mathbf{u}\mathbf{u}^*) = \mathbf{u}(\mathbf{u}^*\mathbf{u})\mathbf{u}^* = \mathbf{u}\mathbf{u}^*.$$

If $\mathbf{u}$ is the zero column vector then $\mathbf{u}\mathbf{u}^*$ is the square zero matrix which is also a projection matrix.

Exercises. (1) Show that the matrices

$$\Pi_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \Pi_2 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

are projection matrices and that $\Pi_1\Pi_2 = 0$.

(2) Is the sum of two $n \times n$ projection matrices an $n \times n$ projection matrix?

(3) Let A be an $n \times n$ matrix with $A^2 = A$. Show that $\det(A)$ is either equal to zero or equal to 1.

(4) Let

$$\Pi = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

and

$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Show that

$$\|\Pi\mathbf{u} - \Pi\mathbf{v}\| \leq \|\mathbf{u} - \mathbf{v}\|.$$

(5) Consider the matrices

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \qquad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Show that $[A, I_2] = 0$ and $[A, C] = 0$. Show that $I_2C = C$, $CI = C$, $CC = I$. A group theoretical reduction (Steeb [60; 61]) leads to the projection matrices

$$\Pi_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \Pi_2 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Apply the projection operators to the standard basis to find a new basis. Show that the matrix A takes the form

$$\widetilde{A} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

within the new basis. Notice that the new basis must be normalized before the matrix $\widetilde{A}$ can be calculated.

1.8 Fourier and Hadamard Matrices

Fourier and Hadamard matrices play an important role in spectral analysis (Davis [15], Elliott and Rao [18], Regalia and Mitra [47]). We give a short introduction to these types of matrices. In sections 2.8 and 3.16 we discuss the connection with the Kronecker product.

Let n be a fixed integer ≥ 1. We define

$$w := \exp\left(\frac{2\pi i}{n}\right) \equiv \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)$$

where $i = \sqrt{-1}$. w might be taken as any primitive n-th root of unity. It can easily be proved that

$$\begin{aligned} w^n &= 1 \\ w\bar{w} &= 1 \\ \bar{w} &= w^{-1} \\ \bar{w}^k &= w^{-k} = w^{n-k} \end{aligned}$$

and

$$1 + w + w^2 + \cdots + w^{n-1} = 0$$

where $\bar{w}$ is the complex conjugate of w.

Definition 1.32. By the *Fourier matrix* of order n, we mean the matrix $F(= F_n)$ where

$$F^* := \frac{1}{\sqrt{n}}(w^{(i-1)(j-1)}) \equiv \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \ldots & w^{(n-1)(n-1)} \end{pmatrix}$$

where F^* is the conjugate transpose of F.

The sequence w^k, $k = 0, 1, 2\ldots$, is periodic with period n. Consequently there are only n distinct elements in F. Therefore F^* can be written as

$$F^* = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & w^{n-1} & w^{n-2} & \cdots & w \end{pmatrix}.$$

The following theorem can easily be proved

Theorem 1.12. *F is unitary, i.e.*

$$FF^* = F^*F = I_n \Longleftrightarrow F^{-1} = F^* .$$

Proof. This is a result of the geometric series identity

$$\sum_{r=0}^{n-1} w^{r(j-k)} \equiv \frac{1-w^{n(j-k)}}{1-w^{j-k}} = \begin{cases} n \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases} .$$

□

A second application of the geometrical identity yields

$$F^{*2} \equiv F^*F^* \equiv \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 0 & \cdots & & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \cdots & & 0 \end{pmatrix} = F^2.$$

This means F^{*2} is an $n \times n$ permutation matrix.

Corollary 1.1.

$$F^{*4} = I_n, \qquad F^{*3} = F^{*4}(F^*)^{-1} = I_nF = F.$$

Corollary 1.2. *The eigenvalues of F are ± 1, $\pm i$, with appropriate multiplicities.*

The characteristic polynomials $f(\lambda)$ of $F^*(= F_n^*)$ are as follows

$$\begin{array}{lll} n \equiv 0 & \text{modulo } 4, & f(\lambda) = (\lambda-1)^2(\lambda-i)(\lambda+1)(\lambda^4-1)^{(n/4)-1} \\ n \equiv 1 & \text{modulo } 4, & f(\lambda) = (\lambda-1)(\lambda^4-1)^{(1/4)(n-1)} \\ n \equiv 2 & \text{modulo } 4, & f(\lambda) = (\lambda^2-1)(\lambda^4-1)^{(n/4)(n-2)} \\ n \equiv 3 & \text{modulo } 4, & f(\lambda) = (\lambda-i)(\lambda^2-1)(\lambda^4-1)^{(1/4)(n-3)}. \end{array}$$

Definition 1.33. Let

$$Z = (z_1, z_2, \ldots, z_n)^T$$

and

$$\hat{Z} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_n)^T$$

where $z_j \in \mathbb{C}$. The linear transformation

$$\hat{Z} = FZ$$

where F is the Fourier matrix is called the discrete *Fourier transform.*

Its inverse transformation exists since F^{-1} exists and is given by

$$Z = F^{-1}\hat{Z} \equiv F^*\hat{Z}.$$

Let

$$p(z) = a_0 + a_1 z + \cdots + a_{n-1}z^{n-1}$$

be a polynomial of degree $\leq n-1$. It will be determined uniquely by specifying its values $p(z_n)$ at n distinct points z_k, $k = 1, 2, \ldots, n$ in the complex plane $\mathbb{C}$. Select these points z_k as the n roots of unity $1, w, w^2, \ldots, w^{n-1}$. Then

$$\sqrt{n}F^* \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} p(1) \\ p(w) \\ \vdots \\ p(w^{n-1}) \end{pmatrix}$$

so that

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \frac{1}{\sqrt{n}} F \begin{pmatrix} p(1) \\ p(w) \\ \vdots \\ p(w^{n-1}) \end{pmatrix}.$$

These formulas for interpolation at the roots of unity can be given another form.

Definition 1.34. By a *Vandermonde matrix* $V(z_0, z_1, \ldots, z_{n-1})$ is meant a matrix of the form

$$V(z_0, z_1, \ldots, z_{n-1}) := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_0 & z_1 & \cdots & z_{n-1} \\ z_0^2 & z_1^2 & \cdots & z_{n-1}^2 \\ \vdots & \vdots & & \vdots \\ z_0^{n-1} & z_1^{n-1} & \cdots & z_{n-1}^{n-1} \end{pmatrix}.$$

It follows that

$$V(1, w, w^2, \ldots, w^{n-1}) = n^{1/2}F^*$$
$$V(1, \bar{w}, \bar{w}^2, \ldots, \bar{w}^{n-1}) = n^{1/2}\bar{F}^* = n^{1/2}F.$$

Furthermore

$$\begin{aligned} p(z) &= (1, z, \ldots, z^{n-1})(a_0, a_1, \ldots, a_{n-1})^T \\ &= (1, z, \ldots, z^{n-1})n^{-1/2}F(p(1), p(w), \ldots, p(w^{n-1}))^T \\ &= n^{-1/2}(1, z, \ldots, z^{n-1})V(1, \bar{w}, \ldots, \bar{w}^{n-1})(p(1), p(w), \ldots, p(w^{n-1}))^T. \end{aligned}$$

Let F'_{2^n} denote the Fourier matrices of order 2^n whose rows have been permuted according to the bit reversing permutation.

Definition 1.35. A sequence in natural order can be arranged in *bit-reversed order* as follows: For an integer expressed in binary notation, reverse the binary form and transform to decimal notation, which is then called bit-reversed notation.

Example 1.21. The number 6 can be written as

$$6 = 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0.$$

Therefore in binary $6 \to 110$. Reversing the binary digits yields 011. Since

$$3 = 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

we have $6 \to 3$. ♣

Since the sequence 0, 1 is the bit reversed order of 0, 1 and 0, 2, 1, 3 is the bit reversed order of 0, 1, 2, 3 we find that the matrices F'_2 and F'_4 are given by

$$F'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = F_2$$

$$F'_4 = \frac{1}{\sqrt{4}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \\ 1 & -i & -1 & i \end{pmatrix}.$$

Definition 1.36. By a *Hadamard matrix* of order n, H ($\equiv H_n$), is meant a matrix whose elements are either $+1$ or -1 and for which

$$HH^T = H^T H = nI_n$$

where I_n is the $n \times n$ unit matrix. Thus, $n^{-1/2}H$ is an orthogonal matrix.

Example 1.22. The 1×1, 2×2 and 4×4 Hadamard matrices are given by

$$
\begin{aligned}
H_1 &= (1) \\
H_2 &= \sqrt{2}F_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
H_{4,1} &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \\
H_{4,2} &= \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}.
\end{aligned}
$$

♣

Note that the columns (or rows) considered as vectors are orthogonal to each other. Sometimes the term Hadamard matrix is limited to the matrices of order 2^n. These matrices have the property

$$H_{2^n} = H_{2^n}^T$$

so that

$$H_{2^n}^2 = 2^n I.$$

A recursion relation to find H_{2^n} using the Kronecker product will be given in sections 2.8 and 3.16.

Definition 1.37. The *Walsh-Hadamard transform* is defined as

$$\hat{Z} = HZ$$

where H is an Hadamard matrix, where $Z = (z_1, z_2, \ldots, z_n)^T$. Since H^{-1} exists we have

$$Z = H^{-1}\hat{Z}\,.$$

For example the inverse of H_2 is given by

$$H_2^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Exercises. (1) Show that

$$F = F^T, \qquad F^* = (F^*)^T = \bar{F}, \qquad F = \bar{F}^*.$$

This means F and F^* are symmetric.

(2) Show that the sequence

$$0, 8, 4, 12, 2, 10, 6, 14, 1, 9, 5, 13, 3, 11, 7, 15$$

is the bit reversed order of

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15.$$

(3) Find the eigenvalues of

$$F_4^* = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega & \omega^2 \\ 1 & \omega^3 & \omega^2 & \omega \end{pmatrix}.$$

Derive the eigenvalues of F_4.

(4) The discrete Fourier transform in one dimension can also be written as

$$\hat{x}(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\exp(-ik2\pi n/N)$$

where $N \in \mathbb{N}$ and $k = 0, 1, 2, \ldots, N-1$. Show that

$$x(n) = \sum_{k=0}^{N-1} \hat{x}(k)\exp(ik2\pi n/N).$$

Let

$$x(n) = \cos(2\pi n/N)$$

where $N = 8$ and $n = 0, 1, 2, \ldots, N-1$. Find $\hat{x}(k)$.

(5) Find all 8×8 Hadamard matrices and their eigenvalues.

1.9 Transformation of Matrices

Let V be a vector space of finite dimension n and let $\mathcal{A} : V \to V$ be a linear transformation, represented by a (square) matrix $A = (a_{ij})$ relative to a basis $(\mathbf{e}_i)$. Relative to another basis $(\mathbf{f}_i)$, the same transformation is represented by the matrix

$$B = Q^{-1}AQ$$

where Q is the invertible matrix whose jth column vector consists of the components of the vector $\mathbf{f}_j$ in the basis $(\mathbf{e}_i)$. Since the same linear transformation $\mathcal{A}$ can in this way be represented by different matrices, depending on the basis that is chosen, the problem arises of finding a basis relative to which the matrix representing the transformation is as simple as possible. Equivalently, given a matrix A, that is to say, those which are of the form $Q^{-1}AQ$, with Q invertible, those which have a form that is 'as simple as possible'.

Definition 1.38. If there exists an invertible matrix Q such that the matrix $Q^{-1}AQ$ is diagonal, then the matrix A is said to be *diagonalizable*.

In this case, the diagonal elements of the matrix $Q^{-1}AQ$ are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the matrix A. The jth column vector of the matrix Q consists of the components (relative to the same basis as that used for the matrix A) of a normalized eigenvector corresponding to λ_j. In other words, a matrix is diagonalizable if and only if there exists a basis of eigenvectors.

Example 1.23. The 2×2 matrix

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

is diagonalizable with

$$Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

We find

$$Q^{-1}AQ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The 2×2 matrix

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

cannot be diagonalized. ♣

For nondiagonalizable matrices *Jordan's theorem* gives the simplest form among all similar matrices.

Definition 1.39. A matrix $A = (a_{ij})$ of order n is *upper triangular* if $a_{ij} = 0$ for $i > j$ and *lower triangular* if $a_{ij} = 0$ for $i < j$. If there is no need to distinguish between the two, the matrix is simply called *triangular*.

Theorem 1.13. *(1) Given a square matrix A, there exists a unitary matrix U such that the matrix $U^{-1}AU$ is triangular.*
(2) Given a normal matrix A, there exists a unitary matrix U such that the matrix $U^{-1}AU$ is diagonal.
(3) Given a symmetric matrix A, there exists an orthogonal matrix O such that the matrix $O^{-1}AO$ is diagonal.

The proof of this theorem follows from the proofs in section 1.11.

The matrices U satisfying the conditions of the statement are not unique (consider, for example, $A = I$). The diagonal elements of the triangular matrix $U^{-1}AU$ of (1), or of the diagonal matrix $U^{-1}AU$ of (2), or of the diagonal matrix of (3), are the eigenvalues of the matrix A. Consequently, they are real numbers if the matrix A is Hermitian or symmetric and complex numbers of modulus 1 if the matrix is unitary or orthogonal. It follows from (2) that every Hermitian or unitary matrix is diagonalizable by a unitary matrix. The preceding argument shows that if, O is an orthogonal matrix, there exists a unitary matrix U such that $D = U^*OU$ is diagonal (the diagonal elements of D having modulus equal to 1), but the matrix U is not, in general, real, that is to say, orthogonal.

Definition 1.40. The *singular values* of a square matrix A are the positive square roots of the eigenvalues of the Hermitian matrix A^*A (or A^TA, if the matrix A is real).

Example 1.24. Let

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

Then

$$A^*A = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 5 & 4 \\ 0 & 4 & 4 \end{pmatrix} .$$

Obviously A^*A is a positive semidefinite matrix. The eigenvalues of A^*A are 0, 4, 9. Thus the singular values are 0, 2, 3. ♣

They are always nonnegative, since from the relation $A^*A\mathbf{u} = \lambda\mathbf{u}$, $\mathbf{u} \neq \mathbf{0}$, it follows that $(A\mathbf{u})^*A\mathbf{u} = \lambda\mathbf{u}^*\mathbf{u}$.

The singular values are all strictly positive if and only if the matrix A is invertible. In fact, we have

$$A\mathbf{u} = \mathbf{0} \Rightarrow A^*A\mathbf{u} = \mathbf{0} \Rightarrow \mathbf{u}^*A^*A\mathbf{u} = (A\mathbf{u})^*A\mathbf{u} = \mathbf{0} \Rightarrow A\mathbf{u} = \mathbf{0}.$$

Definition 1.41. Two matrices A and B of type (m, n) are said to be *equivalent* if there exists an invertible matrix Q of order m and an invertible matrix R of order n such that

$$B = QAR.$$

This is a more general notion than that of the similarity of matrices. In fact, it can be shown that every square matrix is equivalent to a diagonal matrix.

Theorem 1.14. *If A is a real, square matrix, there exist two orthogonal matrices U and V such that*

$$U^TAV = diag(\mu_i)$$

and, if A is a complex, square matrix, there exist two unitary matrices U and V such that

$$U^*AV = diag(\mu_i).$$

In either case, the numbers $\mu_i \geq 0$ are the singular values of the matrix A.

The proof of this theorem follows from the proofs in section 1.11.

If A is an $n \times n$ matrix and U is an $n \times n$ unitary matrix, then $(m \in \mathbb{N})$

$$UA^mU^* = (UAU^*)^m$$

since $UU^* = I_n$.

Exercises. (1) Find the eigenvalues and normalized eigenvectors of the matrix

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

Then use the normalized eigenvectors to construct the matrix Q^{-1} such that $Q^{-1}AQ$ is a diagonal matrix.

(2) Consider the skew-symmetric matrix

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

Find the eigenvalues and the corresponding normalized eigenvectors. Can one find an invertible 3×3 matrix S such that SAS^{-1} is a diagonal matrix?

(3) Show that the matrix

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

is not diagonalizable.

(4) Let O be an orthogonal matrix. Show that there exists an orthogonal matrix Q such that $Q^{-1}OQ$ is given by

$$(1) \oplus \cdots \oplus (1) \oplus (-1) \oplus (-1) \oplus \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} \cos\theta_r & \sin\theta_r \\ -\sin\theta_r & \cos\theta_r \end{pmatrix}$$

where $\oplus$ denotes the direct sum.

(5) Let A be a real matrix of order n. Show that a necessary and sufficient condition for the existence of a unitary matrix U of the same order and of a real matrix B (of the same order) such that $U = A + iB$ (in other word, such that the matrix A is the 'real part' of the matrix U) is that all the singular values of the matrix A should be not greater than 1.

1.10 Permutation Matrices

In this section we introduce permutation matrices and discuss their properties. The connection with the Kronecker product is described in section 2.4. By a *permutation* $\boldsymbol{\sigma}$ of the set

$$N := \{\, 1,\ 2,\ \ldots,\ n \,\}$$

is meant a one-to-one mapping of N onto itself. Including the identity permutation there are $n!$ distinct permutations of N. We indicate a permutation by

$$\boldsymbol{\sigma}(1) = i_1, \quad \boldsymbol{\sigma}(2) = i_2, \quad \ldots, \quad \boldsymbol{\sigma}(n) = i_n$$

which is written as

$$\boldsymbol{\sigma} : \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}.$$

The inverse permutation is designated by $\boldsymbol{\sigma}^{-1}$. Thus

$$\boldsymbol{\sigma}^{-1}(i_k) = k.$$

Let $\mathbf{e}_{j,n}^T$ denote the unit (row) vector of n components which has a 1 in the j-th position and 0's elsewhere

$$\mathbf{e}_{j,n}^T := (0, \ldots, 0, 1, 0, \ldots, 0).$$

Definition 1.42. By a *permutation matrix* of order n is meant a matrix of the form

$$P = P_\sigma = \begin{pmatrix} \mathbf{e}_{i_1,n}^T \\ \mathbf{e}_{i_2,n}^T \\ \vdots \\ \mathbf{e}_{i_n,n}^T \end{pmatrix}.$$

The i-th row of P has a 1 in the $\boldsymbol{\sigma}(i)$-th column and 0's elsewhere. The j-th column of P has a 1 in the $\boldsymbol{\sigma}^{-1}(j)$-th row and 0's elsewhere. Thus each row and each column of P has precisely one 1 in it. We have

$$P_\sigma \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{\sigma(1)} \\ x_{\sigma(2)} \\ \vdots \\ x_{\sigma(n)} \end{pmatrix}.$$

Example 1.25. Let

$$\boldsymbol{\sigma} : \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

Then the permutation matrix is

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

♣

Example 1.26. The set of all 3×3 permutation matrices are given by the 6 matrices

$$\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

♣

It can easily be proved that

$$P_\sigma P_\tau = P_{\sigma\tau}$$

where the product of the permutations $\boldsymbol{\sigma}$, $\boldsymbol{\tau}$ is applied from left to right. Furthermore,

$$(P_\sigma)^* = P_{\sigma^{-1}}.$$

Hence

$$(P_\sigma)^* P_\sigma = P_{\sigma^{-1}} P_\sigma = P_I = I_n$$

where I_n is the $n \times n$ unit matrix. It follows that

$$(P_\sigma)^* = P_{\sigma^{-1}} = (P_\sigma)^{-1}.$$

Consequently, the permutation matrices form a group under matrix multiplication. We find that the permutation matrices are unitary, forming a finite subgroup of the unitary group (see section 1.18 for more details on group theory).

The determinant of a permutation matrix is either $+1$ or -1. The trace of an $n \times n$ permutation matrix is in the set $\{0, 1, \ldots, n-1, n\}$.

Exercises. (1) Show that the number of $n \times n$ permutation matrices is given by $n!$.

(2) Find all 4×4 permutation matrices.

(3) Show that the determinant of a permutation matrix is either $+1$ or -1.

(4) A 3×3 permutation matrix P has $\mathrm{tr}(P) = 1$ and $\det(P) = -1$. What can be said about the eigenvalues of P?

(5) Show that the eigenvalues λ_j of a permutation matrix are $\lambda_j \in \{\, 1\,,\, -1\,\}$.

(6) Show that the rank of an $n \times n$ permutation matrix is n.

(7) Consider the set of all $n \times n$ permutation matrices. How many of the elements are their own inverses, i.e. $P = P^{-1}$?

(8) Consider the 4×4 permutation matrix

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

Find all the eigenvalues and normalized eigenvectors. From the normalized eigenvectors construct an invertible matrix Q such that $Q^{-1}PQ$ is a diagonal matrix.

(9) Let P_1 and P_2 be two $n \times n$ permutation matrices. Is $[P_1, P_2] = 0$, where $[\,,\,]$ denotes the commutator. The commutator is defined by $[P_1, P_2] := P_1P_2 - P_2P_1$.

(10) Is it possible to find $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{v}\mathbf{v}^T$ is an $n \times n$ permutation matrix?

(11) Let I_n be the $n \times n$ identity matrix and 0_n be the $n \times n$ zero matrix. Is the $2n \times 2n$ matrix

$$\begin{pmatrix} 0_n & I_n \\ I_n & 0_n \end{pmatrix}$$

a permutation matrix?

1.11 Matrix Decompositions

Normal matrices have a *spectral decomposition*. Let A be upper triangular with respect to an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{C}^n$. Then

$$A\mathbf{v}_j = \sum_{k=1}^{j} a_{j,k}\mathbf{v}_k$$

for some $a_{1,1}, a_{2,1}, a_{2,2}, \ldots, a_{n,n} \in \mathbb{C}$. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is orthonormal we have $a_{j,k} = \mathbf{v}_k^* A\mathbf{v}_j$. It follows that

$$\overline{a}_{k,j} = \mathbf{v}_j^* A^*\mathbf{v}_k,$$

or equivalently

$$A^*\mathbf{v}_k = \sum_{j=k}^{n} \overline{a}_{k,j}\mathbf{v}_j.$$

We have

$$(A\mathbf{v}_j)^*(A\mathbf{v}_j) = \sum_{k=1}^{j} |a_{j,k}|^2$$

and

$$(A^*\mathbf{v}_j)^*(A^*\mathbf{v}_j) = \sum_{k=j}^{n} |a_{k,j}|^2.$$

However

$$\begin{aligned}(A^*\mathbf{v}_1)^*(A^*\mathbf{v}_1) = \mathbf{v}_1 AA^*\mathbf{v}_1 = \mathbf{v}_1 A^*A\mathbf{v}_1 &= (A\mathbf{v}_j)^*(A\mathbf{v}_j) \\ = \sum_{k=1}^{j} |a_{j,k}|^2 &= \sum_{k=j}^{n} |a_{k,j}|^2,\end{aligned}$$

i.e.

$$\sum_{k=1}^{j-1} |a_{j,k}|^2 = \sum_{k=j+1}^{n} |a_{k,j}|^2.$$

For $j = 1$ we find $a_{2,1} = a_{3,1} = \cdots = 0$, for $j = 2$ we find $a_{3,2} = a_{4,2} = \cdots = 0$, etc. Thus A is diagonal with respect to $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. As a consequence $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ form an orthonormal set of eigenvectors of A (and span $\mathbb{C}^n$). Thus we can write (*spectral theorem*)

$$A = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j\mathbf{v}_j^*$$

where $A\mathbf{v}_j = \lambda_j \mathbf{v}_j$. Since

$$\left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^*\right) \mathbf{e}_{j,n} = \mathbf{v}_j$$

it follows that

$$\left[\left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^*\right) \mathbf{e}_{j,n}\right]^* = \mathbf{e}_{j,n}^* \left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{l,n}^*\right)^* = \mathbf{v}_k^*$$

and consequently we write

$$A = \left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^*\right) \left(\sum_{j=1}^{n} \lambda_j \mathbf{e}_{j,n} \mathbf{e}_{j,n}^*\right) \left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^*\right)^* = VDV^*$$

where

$$V = \sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^*$$

is a unitary matrix. The columns are the orthonormal eigenvectors of A and

$$D = \sum_{k=1}^{n} \lambda_k \mathbf{e}_{k,n} \mathbf{e}_{k,n}^*$$

is a diagonal matrix of corresponding eigenvalues, i.e the eigenvalue λ_k in the k-th entry on the diagonal of D corresponds the the eigenvector $\mathbf{v}_k$ which is the k-th column of V. This decomposition is known as a *spectral decomposition* or *diagonalization* of A.

Example 1.27. Let

$$A = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

The eigenvalues are $\lambda_1 = 1$, $\lambda_2 = -1$ with the corresponding normalized eigenvectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

Then

$$\begin{aligned} A &\equiv \lambda_1 \mathbf{u}_1 \mathbf{u}_1^* + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^* \\ &\equiv \frac{1}{2} \begin{pmatrix} 1 \\ i \end{pmatrix} (1, -i) - \frac{1}{2} \begin{pmatrix} 1 \\ -i \end{pmatrix} (1, i) \\ &\equiv \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \\ &\equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \end{aligned}$$

♣

Example 1.28. Let

$$A = \begin{pmatrix} 5 & -2 & -4 \\ -2 & 2 & 2 \\ -4 & 2 & 5 \end{pmatrix}.$$

Then the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 10$. This means the eigenvalue $\lambda = 1$ is twofold. The eigenvectors are

$$\mathbf{u}_1 = \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}.$$

We find that

$$(\mathbf{u}_1, \mathbf{u}_3) = 0, \qquad (\mathbf{u}_2, \mathbf{u}_3) = 0, \qquad (\mathbf{u}_1, \mathbf{u}_2) = 1.$$

However, the two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent. Now we use the Gram-Schmidt algorithm to find orthogonal eigenvectors (see section 1.17). We choose

$$\mathbf{u}_1' = \mathbf{u}_1, \qquad \mathbf{u}_2' = \mathbf{u}_2 + \alpha \mathbf{u}_1$$

such that

$$\alpha := -\frac{(\mathbf{u}_1, \mathbf{u}_2)}{(\mathbf{u}_1, \mathbf{u}_1)} = -\frac{1}{5}.$$

Then

$$\mathbf{u}_2' = \frac{1}{5} \begin{pmatrix} -4 \\ 2 \\ -5 \end{pmatrix}.$$

The normalized eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2' = \frac{1}{3\sqrt{5}} \begin{pmatrix} -4 \\ 2 \\ -5 \end{pmatrix}, \qquad \mathbf{u}_3 = \frac{1}{3} \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}.$$

Consequently

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^* + \lambda_2 \mathbf{u}_2' \mathbf{u}_2'^* + \lambda_3 \mathbf{u}_3 \mathbf{u}_3^*.$$

From the normalized eigenvectors we obtain the orthogonal matrix

$$O = \begin{pmatrix} \frac{-1}{\sqrt{5}} & \frac{-4}{3\sqrt{5}} & \frac{2}{3} \\ \frac{-2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} & -\frac{1}{3} \\ 0 & \frac{-5}{3\sqrt{5}} & -\frac{2}{3} \end{pmatrix}.$$

Therefore

$$O^T A O = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

where $O^T = O^{-1}$ and the eigenvalues of A are 1, 1 and 10. ♣

If the normal matrix A has multiple eigenvalues with corresponding nonorthogonal eigenvectors, we proceed as follows. Let λ be an eigenvalue of multiplicity m. Then the eigenvalues with their corresponding eigenvectors can be ordered as

$$\lambda,\ \lambda, \ldots, \lambda,\ \lambda_{m+1}, \ldots, \lambda_n, \qquad \mathbf{u}_1,\ \mathbf{u}_2, \ldots, \mathbf{u}_m,\ \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n.$$

The vectors $\mathbf{u}_{m+1}, \ldots, \mathbf{u}_n$ are orthogonal to each other and to the rest. What is left is to find a new set of orthogonal vectors $\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_m'$ each being orthogonal to $\mathbf{u}_{m+1}, \ldots, \mathbf{u}_n$ together with each being an eigenvector of A. The procedure we use is the Gram-Schmidt in section 1.17.

Let

$$\mathbf{u}_1' = \mathbf{u}_1, \qquad \mathbf{u}_2' = \mathbf{u}_2 + \alpha \mathbf{u}_1 .$$

Then $\mathbf{u}_2'$ is an eigenvector of A, for it is a combination of eigenvectors corresponding to the same eigenvalue λ. Also $\mathbf{u}_2'$ is orthogonal to $\mathbf{u}_{m+1}, \ldots, \mathbf{u}_n$ since the latter are orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$. What remains is to make $\mathbf{u}_2'$ orthogonal to $\mathbf{u}_1'$ i.e. to $\mathbf{u}_1$. We obtain

$$\alpha = -\frac{(\mathbf{u}_1, \mathbf{u}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}.$$

Next we set

$$\mathbf{u}_3' = \mathbf{u}_3 + \alpha \mathbf{u}_1 + \beta \mathbf{u}_2$$

where α and β have to be determined. Using the same reasoning, we obtain the linear equation for α and β

$$\begin{pmatrix} (\mathbf{u}_1, \mathbf{u}_1) & (\mathbf{u}_1, \mathbf{u}_2) \\ (\mathbf{u}_2, \mathbf{u}_1) & (\mathbf{u}_2, \mathbf{u}_2) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = - \begin{pmatrix} (\mathbf{u}_1, \mathbf{u}_3) \\ (\mathbf{u}_2, \mathbf{u}_3) \end{pmatrix}.$$

The approach can be repeated until we obtain $\mathbf{u}'_m$. The Gramian matrix of the above equations is nonsingular, since the eigenvectors of a Hermitian matrix are linearly independent.

Next we consider the *singular value decomposition* (SVD). Let A be $m \times n$ over $\mathbb{C}$. The matrices A^*A and AA^* are Hermitian and therefore normal. The nonzero eigenvalues of A^*A and AA^* are identical, for if

$$(A^*A)\mathbf{x}_\lambda = \lambda \mathbf{x}_\lambda, \qquad \lambda \neq 0, \quad \mathbf{x}_\lambda \neq \mathbf{0}$$

for some eigenvector $\mathbf{x}_\lambda$, then

$$A(A^*A)\mathbf{x}_\lambda = \lambda A\mathbf{x}_\lambda \;\Rightarrow\; (AA^*)(A\mathbf{x}_\lambda) = \lambda(A\mathbf{x}_\lambda)$$

so, since $A\mathbf{x}_\lambda \neq \mathbf{0}$, λ is an eigenvalue of AA^*. Similarly the nonzero eigenvalues of AA^* are eigenvalues of A^*A.

Since A^*A is normal the spectral theorem provides

$$A^*A = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j \mathbf{v}_j^* = VDV^*$$

where $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are the orthonormal eigenvectors of A^*A and

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

are the corresponding eigenvalues. Here we choose a convenient ordering of the eigenvalues.

From

$$(A\mathbf{v}_k)^*(A\mathbf{v}_k) \geq 0 \;\Rightarrow\; \mathbf{v}_k^*(A^*A)\mathbf{v}_k = \lambda_k \mathbf{v}_k^* \mathbf{v}_k = \lambda_k \geq 0$$

we find that all of the eigenvalues are real and nonnegative. Define the *singular values* $\sigma_1 \geq \cdots \geq \sigma_r > 0$ by

$$\sigma_k := \sqrt{\lambda_k}, \quad \lambda_k \neq 0, \quad k = 1, \ldots, r.$$

Now we define the $m \times n$ matrix Σ by

$$(\Sigma)_{j,k} = \begin{cases} \sigma_j & j = k, j < r \\ 0 & \text{otherwise} \end{cases}$$

so that $A^*A = (\Sigma V^*)^*(\Sigma V)$. Let U be an arbitrary $m \times m$ unitary matrix, then $A^*A = (U\Sigma V^*)^*(U\Sigma V)$.

Let $\mathbf{v}_j$ be an eigenvector corresponding to the singular value σ_j (i.e. $A\mathbf{v}_j \neq \mathbf{0}$) and similarly for σ_k and $\mathbf{v}_k$. Then

$$(A\mathbf{v}_j)^*(A\mathbf{v}_k) = \mathbf{v}_j^*(A^*A)\mathbf{v}_k = \sigma_k^2\mathbf{v}_j^*\mathbf{v}_k = \sigma_k^2\delta_{j,k}.$$

Thus $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ are orthogonal. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be an orthonormal basis in $\mathbb{C}^m$ where

$$\mathbf{u}_j = \frac{1}{\sigma_j}A\mathbf{v}_j, \quad j = 1, \dots, r.$$

By construction, for $j, k = 1, \dots, r$

$$\begin{aligned}\mathbf{u}_j^*\mathbf{u}_k &= \frac{1}{\sigma_j\sigma_k}(A\mathbf{v}_j)^*(A\mathbf{v}_k) = \frac{1}{\sigma_j\sigma_k}\mathbf{v}_j^*(A^*A)\mathbf{v}_k = \frac{\sigma_k^2}{\sigma_j\sigma_k}\mathbf{v}_j^*\mathbf{v}_k \\ &= \frac{\sigma_k^2}{\sigma_j\sigma_k}\delta_{j,k} = \frac{\sigma_j^2}{\sigma_j\sigma_j}\delta_{j,k} = \delta_{j,k}\end{aligned}$$

and for $j = 1, \dots, r$ and $k = 1, \dots, m - r$

$$\mathbf{u}_{r+k}^*\mathbf{u}_j = 0 = \frac{1}{\sigma_j}\mathbf{u}_{r+k}^*A\mathbf{v}_j \;\Rightarrow\; \mathbf{u}_{r+k}^*A\mathbf{v}_j = 0.$$

We also have $A\mathbf{v}_{r+k} = 0$. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis we must have

$$A^*\mathbf{u}_{r+k} = \sum_{j=1}^{n}[\mathbf{v}_k^*(A^*\mathbf{u}_{r+k})]\mathbf{v}_k = \sum_{j=1}^{n}[\mathbf{u}_{r+k}^*A\mathbf{v}_k]^*\mathbf{v}_k = \mathbf{0}$$

so that $\mathbf{u}_{r+k}^*A = \mathbf{0}$. Choosing

$$U = \sum_{j=1}^{m}\mathbf{u}_j\mathbf{e}_{j,m}^*$$

(as an exercise, verify that U is unitary) we find

$$\begin{aligned}
U^*A &= \sum_{j=1}^{r} \mathbf{e}_{j,m} \frac{1}{\sigma_j} \mathbf{v}_j^* A^* A + \sum_{j=r+1}^{m} \mathbf{e}_{j,m} \mathbf{u}_j^* A \\
&= \sum_{j=1}^{r} \mathbf{e}_{j,m} \frac{1}{\sigma_j} (A^* A \mathbf{v}_j)^* = \sum_{j=1}^{r} \mathbf{e}_{j,m} \sigma_j \mathbf{v}_j^* \\
&= \sum_{j=1}^{r} \sigma_j \mathbf{e}_{j,m} \mathbf{e}_{j,n}^* \mathbf{e}_{j,n} \mathbf{v}_j^* \\
&= \left(\sum_{j=1}^{r} \sigma_j \mathbf{e}_{j,m} \mathbf{e}_{j,n}^* \right) \left(\sum_{k=1}^{n} \mathbf{e}_{k,n} \mathbf{v}_k^* \right) \\
&= \left(\sum_{j=1}^{r} \sigma_j \mathbf{e}_{j,m} \mathbf{e}_{j,n}^* \right) \left(\sum_{k=1}^{n} \mathbf{v}_k \mathbf{e}_{k,n}^* \right)^* = \Sigma V^*.
\end{aligned}$$

Thus $A = U\Sigma V^*$. This is a *singular value decomposition.*

Next we consider the *polar decomposition.* Let $m = n$ and

$$A = U\Sigma V^*$$

be a singular value decomposition of A. Here the order of the singular values in Σ does not matter. Then

$$H = V\Sigma V^*$$

is Hermitian and

$$\widetilde{U} = UV^*$$

is unitary. Consequently

$$A = (UV^*)(V\Sigma V^*) = \widetilde{U}H$$

is a polar decomposition of A (a product of a unitary and a Hermitian matrix).

Exercises. (1) Consider the symmetric 4×4 matrix

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

Find all eigenvalues and normalized eigenvectors. Reconstruct the matrix from the eigenvalues and normalized eigenvectors.

(2) Let A be a 4×4 symmetric matrix over $\mathbb{R}$ with eigenvalues $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$, $\lambda_4 = 3$ and the corresponding normalized eigenvectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,0,0,1)^T, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,0,0,-1)^T,$$

$$\mathbf{u}_3 = \frac{1}{\sqrt{2}}(0,1,1,0)^T, \qquad \mathbf{u}_4 = \frac{1}{\sqrt{2}}(0,1,-1,0)^T.$$

Find the matrix A.

(3) Let A be the skew-symmetric matrix

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

Find the eigenvalues λ_1 and λ_2 of A and the corresponding normalized eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Show that A is given by

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^* + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^*.$$

(4) Explain why the matrix

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

cannot be reconstructed from the eigenvalues and eigenvectors. Can the matrix

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

be reconstructed from the eigenvalues and eigenvectors?

1.12 Pseudo Inverse

A *1-inverse* of the $m \times n$ matrix A is an $n \times m$ matrix A^- such that $AA^-A = A$. If $m = n$ and A^{-1} exists we find

$$AA^-A = A \;\Rightarrow\; A^{-1}(AA^-A)A^{-1} = A^{-1}AA^{-1} \;\Rightarrow\; A^- = A^{-1}.$$

Consider

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

Then both A and the 2×2 identity matrix I_2 are 1-inverses of A. Consequently, the 1-inverse is in general not unique.

The *Moore-Penrose pseudo inverse* of the $m \times n$ matrix A is the 1-inverse A^- of A which additionally satisfies

$$A^-AA^- = A^- \qquad (AA^-)^* = AA^- \qquad (A^-A)^* = A^-A.$$

With these properties the matrices A^-A and AA^- are projection matrices.

Let $A = U\Sigma V^*$ be a singular value decomposition of A. Then

$$A^- = V\Sigma^-U^*$$

is a Moore-Penrose pseudo inverse of A, where

$$(\Sigma^-)_{jk} = \begin{cases} \frac{1}{(\Sigma)_{kj}} & (\Sigma)_{kj} \neq 0 \\ 0 & (\Sigma)_{kj} = 0 \end{cases}.$$

We verify this by first calculating $(j, k \in \{1, 2, \ldots, n\})$

$$(\Sigma^-\Sigma)_{jk} = \sum_{l=1}^{m} (\Sigma^-)_{jl}(\Sigma)_{lk} = \sum_{\substack{l=1 \\ (\Sigma)_{lj} \neq 0}}^{m} \frac{(\Sigma)_{lk}}{(\Sigma)_{lj}}.$$

Since $(\Sigma)_{lj} = 0$ when $l \neq j$ we find

$$(\Sigma^-\Sigma)_{jk} = \begin{cases} \frac{(\Sigma)_{jk}}{(\Sigma)_{jj}} & (\Sigma)_{jj} \neq 0 \\ 0 & (\Sigma)_{jj} = 0 \end{cases} = \begin{cases} \delta_{jk} & (\Sigma)_{jj} \neq 0 \\ 0 & (\Sigma)_{jj} = 0 \end{cases}.$$

Similarly $(j, k \in \{1, 2, \ldots, m\})$

$$(\Sigma\Sigma^-)_{jk} = \begin{cases} \delta_{jk} & (\Sigma)_{jj} \neq 0 \\ 0 & (\Sigma)_{jj} = 0 \end{cases}.$$

The matrix $\Sigma^-\Sigma$ is a diagonal $n \times n$ matrix while $\Sigma\Sigma^-$ is a diagonal $m \times m$ matrix. All the entries of these matrices are 0 or 1 so that

$$(\Sigma\Sigma^-)^* = \Sigma\Sigma^-, \qquad (\Sigma^-\Sigma)^* = \Sigma^-\Sigma.$$

Now

$$(\Sigma\Sigma^-\Sigma)_{jk} = \sum_{l=1}^{n}(\Sigma)_{jl}(\Sigma^-\Sigma)_{lk} = (\Sigma)_{jj}(\Sigma^-\Sigma)_{jk}$$
$$= \begin{cases} \delta_{jk}(\Sigma)_{jj} & (\Sigma)_{jj} \neq 0 \\ 0 & (\Sigma)_{jj} = 0 \end{cases} = (\Sigma)_{jk}$$

and

$$(\Sigma^-\Sigma\Sigma^-)_{jk} = \sum_{l=1}^{m}(\Sigma^-)_{jl}(\Sigma\Sigma^-)_{lk} = (\Sigma^-)_{jj}(\Sigma\Sigma^-)_{jk}$$
$$= \begin{cases} \delta_{jk}(\Sigma^-)_{jj} & (\Sigma)_{jj} \neq 0 \\ 0 & (\Sigma)_{jj} = 0 \end{cases} = (\Sigma^-)_{jk}$$

i.e.

$$\Sigma\Sigma^-\Sigma = \Sigma, \qquad \Sigma^-\Sigma\Sigma^- = \Sigma^-$$

so that Σ^- is the Moore-Penrose pseudo inverse of Σ. The remaining properties are easy to show

$$\begin{aligned} AA^-A &= (U\Sigma V^*)(V\Sigma^-U^*)(U\Sigma V^*) = U\Sigma\Sigma^-\Sigma V^* = U\Sigma V^* = A, \\ A^-AA^- &= (V\Sigma^-U^*)(U\Sigma V^*)(V\Sigma^-U^*) = V\Sigma^-\Sigma\Sigma^-U^* = V\Sigma^-U^* = A^-, \\ (AA^-)^* &= (U\Sigma V^*V\Sigma^-U^*)^* = I_m^* = I_m = AA^-, \\ (A^-A)^* &= (V\Sigma^-U^*U\Sigma V^*)^* = I_n^* = I_n = AA^-. \end{aligned}$$

Thus $A^- = V\Sigma^-U^*$ is a Moore-Penrose pseudo inverse of A.

Example 1.29. The matrix

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

has the singular value decomposition

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

so that

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

Analogously

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

♣

1.13 Vec Operator

Let A be an $m \times n$ matrix. A matrix operation is that of stacking the columns of a matrix one under the other to form a single column. This operation is called vec (Neudecker [43], Brewer [12], Graham [23], Searle [49]). Thus $\text{vec}(A)$ is a vector of order $m \times n$.

Example 1.30. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

Then

$$\text{vec}(A) = \begin{pmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{pmatrix}.$$

♣

Let A, B be $m \times n$ matrices. We can prove that

$$\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B).$$

It is also easy to see that

$$\text{vec}(\alpha A) = \alpha \text{vec}(A), \qquad \alpha \in \mathbb{C}.$$

This means the vec-operation is linear. An extension of $\text{vec}(A)$ is $\text{vech}(A)$, defined in the same way that $\text{vec}(A)$ is, except that for each column of A only that part of it which is on or below the diagonal of A is put into $\text{vech}(A)$ (vector-half of A). In this way, for A symmetric, $\text{vech}(A)$ contains only the distinct elements of A.

Example 1.31. Consider the square matrix

$$A = \begin{pmatrix} 1 & 7 & 6 \\ 7 & 3 & 8 \\ 6 & 8 & 2 \end{pmatrix} = A^T$$

Then

$$\text{vech}(A) = \begin{pmatrix} 1 \\ 7 \\ 6 \\ 3 \\ 8 \\ 2 \end{pmatrix}.$$

♣

The following theorems give useful properties of the vec operator. Proofs of the first depend on the elementary vector $\mathbf{e}_{j,n}$, the j-th column of the $n \times n$ unit matrix, i. e.

$$\mathbf{e}_{1,n} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \mathbf{e}_{2,n} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \ \mathbf{e}_{n,n} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

and

$$\mathbf{e}_{1,n}^T = (1, 0, \ldots, 0), \qquad \mathbf{e}_{2,n}^T = (0, 1, \ldots, 0), \qquad \ldots, \qquad \mathbf{e}_{n,n}^T = (0, \ldots, 0, 1).$$

Theorem 1.15. *Let A, B be $n \times n$ matrices. Then*

$$\mathrm{tr}(AB) = (\mathrm{vec}(A^T))^T \mathrm{vec}(B).$$

Proof. We have

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} \mathbf{e}_{i,n}^T AB\mathbf{e}_{i,n} = (\mathbf{e}_{1,n}^T A \cdots \mathbf{e}_{r,n}^T A) \begin{pmatrix} B\mathbf{e}_{1,n} \\ B\mathbf{e}_{2,n} \\ \vdots \\ B\mathbf{e}_{r,n} \end{pmatrix} = (\mathrm{vec}(A^T))^T \mathrm{vec}(B).$$

□

Theorem 1.16. *Let A be an $m \times m$ matrix. Then there is a permutation matrix P such that*

$$\mathrm{vec}(A) = P\mathrm{vec}(A^T).$$

The proof is left to the reader as an exercise.

Example 1.32. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

Then $\mathrm{vec}(A) = P\mathrm{vec}(A^T)$, where

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

The full power of the vec operator will be seen when we consider the Kronecker product and the vec operator (see section 2.12). ♣

1.14 Vector and Matrix Norms

Definition 1.43. Let V be an n dimensional vector space over the field $\mathbb{F}$ of scalars. A *norm* on V is a function $\|\cdot\| : V \to \mathbb{R}$ which satisfies the following properties:

$$\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = 0, \text{ and } \|\mathbf{v}\| \geq 0 \text{ for every } \mathbf{v} \in V$$

$$\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\| \text{ for every } \ \alpha \in \mathbb{F} \text{ and } \mathbf{v} \in V$$

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \text{ for every } \mathbf{u}, \mathbf{v} \in V.$$

The last property is known as the *triangle inequality*. A norm on V will also be called a *vector norm*. We call a vector space which is provided with a norm a *normed vector space*.

The norm induces a metric (distance) $\|\mathbf{u} - \mathbf{v}\|$ on the vector space V, where $\mathbf{u}, \mathbf{v} \in V$. We have

$$\|\mathbf{u} - \mathbf{v}\| \geq |\, \|\mathbf{u}\| - \|\mathbf{v}\| \,|\,.$$

Let V be a finite dimensional space. The following three norms are the ones most commonly used in practice

$$\|\mathbf{v}\|_1 := \sum_{j=1}^{n} |v_j|$$

$$\|\mathbf{v}\|_2 := \left(\sum_{j=1}^{n} |v_j|^2 \right)^{1/2} = (\mathbf{v}, \mathbf{v})^{1/2}$$

$$\|\mathbf{v}\|_\infty := \max_{1 \leq j \leq n} |v_j|.$$

The norm $\|\cdot\|_2$ is called the *Euclidean norm*. It is easy to verify directly that the two functions $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are indeed norms. As for the function $\|\cdot\|_2$, it is a particular case of the following more general result.

Theorem 1.17. *Let V be a finite dimensional vector space and $\mathbf{v} \in V$. For every real number $p \geq 1$, the function $\|\cdot\|_p$ defined by*

$$\|\mathbf{v}\|_p := \left(\sum_{j=1}^{n} |v_j|^p \right)^{1/p}$$

is a norm.

For the proof refer to Ciarlet [13].

The proof uses the following inequalities: For $p > 1$ and $1/p + 1/q = 1$, the inequality

$$\sum_{i=1}^{n} |u_j v_j| \le \left(\sum_{j=1}^{n} |u_j|^p \right)^{1/p} \left(\sum_{j=1}^{n} |v_i|^q \right)^{1/q}$$

is called *Hölder's inequality*. Hölder's inequality for $p = 2$,

$$\sum_{j=1}^{n} |u_j v_j| \le \left(\sum_{j=1}^{n} |u_j|^2 \right)^{1/2} \left(\sum_{j=1}^{n} |v_i|^2 \right)^{1/2}$$

is called *the Cauchy-Schwarz inequality*. The triangle inequality for the norm $\| \cdot \|_p$,

$$\left(\sum_{j=1}^{n} |u_j + v_j|^p \right)^{1/p} \le \left(\sum_{j=1}^{n} |u_j|^p \right)^{1/p} + \left(\sum_{j=1}^{n} |v_j|^p \right)^{1/p}$$

is called *Minkowski's inequality.*

The norms defined above are *equivalent*, this property being a particular case of the equivalence of norms in a finite dimensional space.

Definition 1.44. Two norms $\| \cdot \|$ and $\| \cdot \|'$, defined over the same vector space V, are equivalent if there exist two constants C and C' such that

$$\|\mathbf{v}\|' \le C\|\mathbf{v}\| \quad \text{and} \quad \|\mathbf{v}\| \le C'\|\mathbf{v}\|' \quad \text{for every} \quad \mathbf{v} \in V.$$

Let $\mathcal{A}_n$ be the ring of matrices of order n, with elements in the field $\mathbb{F}$.

Definition 1.45. A *matrix norm* is a function $\|\cdot\| : \mathcal{A}_n \to \mathbb{R}$ which satisfies the following properties

$$\begin{aligned} \|A\| = 0 \Leftrightarrow A = 0 \text{ and } \|A\| \ge 0 \text{ for every } A \in \mathcal{A}_n \\ \|\alpha A\| = |\alpha| \|A\| \text{ for every } \alpha \in \mathbb{F}, \quad A \in \mathcal{A}_n \\ \|A + B\| \le \|A\| + \|B\| \text{ for every } A, B \in \mathcal{A}_n \\ \|AB\| \le \|A\| \|B| \quad \text{for every } A, B \in \mathcal{A}_n. \end{aligned}$$

The ring $\mathcal{A}_n$ is itself a vector space of dimension n^2. Thus the first three properties above are nothing other than those of a vector norm, considering a matrix as a vector with n^2 components. The last property is evidently special to square matrices.

The result which follows gives a particularly simple means of constructing matrix norms.

Definition 1.46. Given a vector norm $\|\cdot\|$ on $\mathbb{C}^n$, the function $\|\cdot\| : \mathcal{A}_n(\mathbb{C}) \to \mathbb{R}$ defined by

$$\|A\| := \sup_{\substack{\mathbf{v}\in\mathbb{C}^n \\ \|\mathbf{v}\|=1}} \|A\mathbf{v}\|$$

is a matrix norm, called the *subordinate matrix norm* (subordinate to the given vector norm). Sometimes it is also called the *induced matrix norm.*

This is just one particular case of the usual definition of the norm of a linear transformation.

Example 1.33. Consider the matrix

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

Then we find

$$\|A\| = \sup_{\substack{\mathbf{v}\in\mathbb{C}^n \\ \|\mathbf{v}\|=1}} \|A\mathbf{v}\| = \sqrt{10}.$$

This result can be found by using the method of the *Lagrange multiplier.* The constraint is $\|\mathbf{v}\| = 1$. Furthermore we note that the eigenvalues of the matrix

$$A^T A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix}$$

are given by $\lambda_1 = 10$ and $\lambda_2 = 0$. Thus the norm of A is the square root of the largest eigenvalue of $A^T A$. ♣

Example 1.34. Let U_1 and U_2 be unitary $n \times n$ matrices with

$$\|U_1 - U_2\| \le \epsilon.$$

Let $\mathbf{v}$ be a normalized vector in $\mathbb{C}^n$. We have

$$\begin{aligned} \|U_1\mathbf{v} - U_2\mathbf{v}\| &= \|(U_1 - U_1)\mathbf{v}\| \\ &\le \max_{\|\mathbf{y}\|=1} \|(U_1 - U_2)\mathbf{y}\| \\ &= \|U_1 - U_2\| \\ &\le \epsilon. \end{aligned}$$

♣

It follows from the definition of a subordinate norm that

$$\|A\mathbf{v}\| \le \|A\|\|\mathbf{v}\| \quad \text{for every } \mathbf{v} \in \mathbb{C}^n.$$

A subordinate norm always satisfies $\|I_n\| = 1$, where I_n is the $n \times n$ unit matrix. Let us now calculate each of the subordinate norms of the vector norms $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$.

Theorem 1.18. *Let $A = (a_{ij})$ be a square matrix. Then*

$$\|A\|_1 := \sup_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|_1 = \max_{1\le j\le n} \sum_{i=1}^n |a_{ij}|$$

$$\|A\|_2 := \sup_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|_2 = \sqrt{\varrho(A^*A)} = \|A^*\|_2$$

$$\|A\|_\infty := \sup_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|_\infty = \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|$$

where $\varrho(A^*A)$ is the spectral radius of A^*A. The norm $\|\cdot\|_2$ is invariant under unitary transformations

$$UU^* = I \Rightarrow \|A\|_2 = \|AU\|_2 = \|UA\|_2 = \|U^*AU\|_2.$$

Furthermore, if the matrix A is normal, i.e $AA^* = A^*A$

$$\|A\|_2 = \varrho(A).$$

The invariance of the norm $\|\cdot\|_2$ under unitary transformations is nothing more than the interpretation of the equalities

$$\varrho(A^*A) = \varrho(U^*A^*AU) = \varrho(A^*U^*UA) = \varrho(U^*A^*UU^*AU).$$

If the matrix A is normal, there exists a unitary matrix U such that

$$U^*AU = \operatorname{diag}(\lambda_i(A)) := D.$$

Accordingly,

$$A^*A = (UDU^*)^*UDU^* = UD^*DU^*$$

which proves that

$$\varrho(A^*A) = \varrho(D^*D) = \max_{1\le i\le n} |\lambda_i(A)|^2 = (\varrho(A))^2.$$

The norm $\|A\|_2$ is nothing other than the largest singular value of the matrix A. If a matrix A is Hermitian, or symmetric (and hence normal), we have $\|A\|_2 = \varrho(A)$. If a matrix A is unitary, or orthogonal (and hence normal), we have

$$\|A\|_2 = \sqrt{\varrho(A^*A)} = \sqrt{\varrho(I)} = 1.$$

There exist matrix norms which are not subordinate to any vector norm. An example of a matrix norm which is not subordinate is given in the following theorem.

Theorem 1.19. *The function* $\|\cdot\|_E : \mathcal{A}_n \to \mathbb{R}$ *defined by*

$$\|A\|_E = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = (\mathrm{tr}(A^*A))^{1/2}$$

for every matrix $A = (a_{ij})$ *of order* n *is a matrix norm which is not subordinate (for* $n \geq 2$*). Furthermore, the function is invariant under unitary transformations,*

$$UU^* = I \Rightarrow \|A\|_E = \|AU\|_E = \|UA\|_E = \|U^*AU\|_E$$

and satisfies

$$\|A\|_2 \leq \|A\|_E \leq \sqrt{n}\|A\|_2 \textit{ for every } A \in \mathcal{A}_n.$$

Proof. The fact that $\|A\|_E$ is invariant under unitary transformation of A follows from the cyclic invariance of the trace (see section 1.4). The eigenvalues of A^*A are real and nonnegative. Let

$$\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_1 \geq 0$$

be the eigenvalues of A^*A. Then $\varrho(A^*A) = \lambda_n$. Since

$$\|A\|_2 = \sqrt{\varrho(A^*A)} = \sqrt{\lambda_n}$$

and

$$\|A\|_E = \sqrt{\lambda_n + \lambda_{n-1} + \cdots + \lambda_1} \geq \sqrt{\lambda_n}$$

we have $\|A\|_2 \leq \|A\|_E$. Also

$$\|A\|_E = \sqrt{\lambda_n + \lambda_{n-1} + \cdots + \lambda_1} \leq \sqrt{n\lambda_n} = \sqrt{n}\sqrt{\varrho(A^*A)}$$

so that $\|A\|_E \leq \sqrt{n}\|A\|_2$. □

Example 1.35. Let I_2 be the 2×2 unit matrix. Then

$$\|I_2\|_E = \sqrt{2} = \sqrt{2}\|I_2\|_2 \,.$$

♣

Exercises. (1) Let A, B be $n \times n$ matrices over $\mathbb{C}$. Show that

$$(A, B) := \operatorname{tr}(AB^*)$$

defines a scalar product. Then

$$\|A\| = \sqrt{(A, A)}$$

defines a norm. Find $\|A\|$ for

$$A = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

(2) Given a diagonalizable matrix A, does a matrix norm $\| \cdot \|$ exist for which $\varrho(A) = \|A\|$?

(3) Let $\mathbf{v} = \exp(i\alpha)\mathbf{u}$, where $\alpha \in \mathbb{R}$. Show that $\|\mathbf{u}\| = \|\mathbf{v}\|$.

(4) What can be said about the norm of a nilpotent matrix?

(5) What can be said about the norm of an idempotent matrix?

(6) Let A be a Hermitian matrix. Find a necessary and sufficient condition for the function $\mathbf{v} \to (\mathbf{v}^* A \mathbf{v})^{1/2}$ to be a norm.

(7) Let $\| \cdot \|$ be a subordinate matrix norm and A an $n \times n$ matrix satisfying $\|A\| < 1$. Show that the matrix $I_n + A$ is invertible and

$$\|(I_n + A)^{-1}\| \leq \frac{1}{1 - \|A\|}.$$

(8) Prove that the function

$$\mathbf{v} \in \mathbb{C}^n \to \|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}$$

is not a norm when $0 < p < 1$ (unless $n = 1$).

(9) Find the smallest constants C for which

$$\|\mathbf{v}\| \leq C \|\mathbf{v}\|' \quad \text{for every } \mathbf{v} \in \mathbb{F}^n$$

when the distinct norms $\| \cdot \|$ and $\| \cdot \|'$ are chosen from the set $\{\| \cdot \|_1, \| \cdot \|_2, \| \cdot \|_\infty\}$.

1.15 Rank-k Approximations

Let M be an $m \times n$ matrix over $\mathbb{C}$. We consider the *rank-k approximation* problem, i.e. to find an $m \times n$ matrix A with $\mathrm{rank}(A) = k \leq \mathrm{rank}(M)$ such that

$$\|M - A\|$$

is a minimum and $\|\cdot\|$ denotes some norm. Different norms lead to different matrices A. For the remainder of this section we consider the Frobenius norm $\|\cdot\|_F$.

Certain transformations leave the Frobenius norm invariant, for example unitary transformations. Let U_m and U_n be $m \times m$ and $n \times n$ unitary matrices respectively. Using the property that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ we find

$$\begin{aligned}\|U_m M\|_F &= \sqrt{\mathrm{tr}((U_m M)^*(U_m M))} = \sqrt{\mathrm{tr}(M^* U_m^* U_m M)} \\ &= \sqrt{\mathrm{tr}(M^* M)} = \|M\|_F\end{aligned}$$

and

$$\begin{aligned}\|M U_n\|_F &= \sqrt{\mathrm{tr}((M U_n)^*(M U_n))} = \sqrt{\mathrm{tr}(U_n^* M^* M U_n)} \\ &= \sqrt{\mathrm{tr}(U_n U_n^* M^* M)} = \sqrt{\mathrm{tr}(M^* M)} = \|M\|_F.\end{aligned}$$

Using the singular value decomposition $M = U \Sigma V^*$ of M we obtain

$$\|M - A\|_F = \|U^*(M - A)V\|_F = \|\Sigma - A'\|_F$$

where $A' := U^* A V$. Since

$$\|\Sigma - A'\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |\Sigma_{ij} - (A')_{ij}|^2} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |\delta_{ij}\sigma_i - (A')_{ij}|^2}$$

we find that minimizing $\|M - A\|$ implies that A' must be "diagonal"

$$\|\Sigma - A'\|_F = \sqrt{\sum_{i=1}^{\min\{m,n\}} |\sigma_i - (A')_{ii}|^2}.$$

Since A must have rank k, A' must have rank k (since unitary operators are rank preserving). Since A' is "diagonal" only k entries can be nonzero. Thus to minimize $\|M - A\|_F$ we set

$$A' = \sum_{j=1}^{k} \sigma_j \mathbf{e}_{j,m} \mathbf{e}_{j,n}^T$$

where $\sigma_1 \geq \cdots \geq \sigma_k$ are the k largest singular values of M by convention. Finally $A = U A' V^*$.

1.16 Sequences of Vectors and Matrices

Definition 1.47. In a vector space V, equipped with a norm $\|\cdot\|$, a sequence (x_k) of elements of V is said to converge to an element $x \in V$, which is the limit of the sequence (x_k), if

$$\lim_{k\to\infty} \|x_k - x\| = 0$$

and one writes

$$x = \lim_{k\to\infty} x_k .$$

If the space is finite dimensional, the equivalence of the norms shows that the convergence of a sequence is independent of the norm chosen. The particular choice of the norm $\|\cdot\|_\infty$ shows that the convergence of a sequence of vectors is equivalent to the convergence of n sequences (n being equal to the dimension of the space) of scalars consisting of the components of the vectors.

Example 1.36. Let $V = \mathbb{C}^2$ and

$$\mathbf{u}_k := \begin{pmatrix} \exp(-k) \\ 1/(1+k) \end{pmatrix}, \qquad k = 0, 1, 2, \ldots .$$

Then

$$\mathbf{u} = \lim_{k\to\infty} \mathbf{u}_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

♣

By considering the set $\mathcal{A}_{m,n}(K)$ of matrices of type (m,n) as a vector space of dimension mn, one sees in the same way that the convergence of a sequence of matrices of type (m,n) is independent of the norm chosen, and that it is equivalent to the convergence of mn sequences of scalars consisting of the elements of these matrices. The following result gives necessary and sufficient conditions for the convergence of the particular sequence consisting of the successive powers of a given (square) matrix to the null matrix. From these conditions can be derived the fundamental criterion for the convergence of iterative methods for the solution of linear systems of equations.

Theorem 1.20. *Let B be a square matrix. The following conditions are equivalent:*

(1)
$$\lim_{k\to\infty} B^k = 0$$

(2)
$$\lim_{k\to\infty} B^k \mathbf{v} = \mathbf{0} \text{ for every vector } \mathbf{v}$$

(3)
$$\varrho(B) < 1$$

(4)
$$\|B\| < 1 \text{for at least one subordinate matrix norm } \|\cdot\|.$$

For the proof of the theorem refer to Ciarlet [13].

Example 1.37. Consider the matrix

$$B = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}.$$

Then

$$\lim_{k\to\infty} B^k = 0\,.$$

♣

The following theorem (Ciarlet [13]) is useful for the study of iterative methods, as regards the rate of convergence.

Theorem 1.21. *Let A be a square matrix and let $\|\cdot\|$ be any subordinate matrix norm. Then*

$$\lim_{k\to\infty} \|A^k\|^{1/k} = \varrho(A).$$

Example 1.38. Let $A = I_n$, where I_n is the $n \times n$ unit matrix. Then $I_n^k = I_n$ and therefore $||I_n^k||^{1/k} = 1$. Moreover $\varrho(A) = 1$. ♣

In theoretical physics and in particular in quantum mechanics a very important role is played by the exponential function of a square matrix. Let A be an $n \times n$ matrix. We set

$$A_k := I_n + \frac{A}{1!} + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!}, \qquad k \ge 1.$$

The sequence (A_k) converges. Its limit is denoted by $\exp(A)$. We have

$$\exp(A) := \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Let ϵ be a parameter ($\epsilon \in \mathbb{R}$). Then we have

$$\exp(\epsilon A) = \sum_{k=0}^{\infty} \frac{(\epsilon A)^k}{k!}.$$

We can also calculate $\exp(\epsilon A)$ using

$$\exp(\epsilon A) = \lim_{n\to\infty} \left(I_n + \frac{\epsilon A}{n} \right)^n .$$

Example 1.39. Let

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

Then

$$\exp(zA) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cosh(z) + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \sinh(z).$$

♣

Theorem 1.22. *Let A be an $n \times n$ matrix. Then*

$$\det(\exp(A)) \equiv \exp(\operatorname{tr}(A)).$$

For the proof refer to Steeb [56]. The theorem shows that the matrix $\exp(A)$ is always invertible. If A is the zero matrix, then we have

$$\exp(A) = I_n$$

where I_n is the identity matrix.

We can also define

$$\sin(A) := \sum_{k=0}^{\infty} \frac{(-1)^k A^{2k+1}}{(2k+1)!}, \qquad \cos(A) := \sum_{k=0}^{\infty} \frac{(-1)^k A^{2k}}{(2k)!} .$$

Example 1.40. Let $x, y \in \mathbb{R}$ and

$$A = \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} .$$

Then we find

$$\sin(A) = \begin{pmatrix} \sin x & y \cos x \\ 0 & \sin x \end{pmatrix}, \qquad \cos(A) = \begin{pmatrix} \cos x & -y \sin x \\ 0 & \cos x \end{pmatrix} .$$

♣

Exercises. (1) Let

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

Calculate $\exp(\epsilon A)$, where $\epsilon \in \mathbb{R}$.

(2) Let

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

Find $\cos(A)$.

(3) Let A be an $n \times n$ matrix such that $A^2 = I_n$, where I_n is the $n \times n$ unit matrix. Let $\epsilon \in \mathbb{R}$. Show that

$$\exp(\epsilon A) = I_n \cosh(\epsilon) + A \sinh(\epsilon).$$

(4) Let A be a square matrix such that the sequence $(A^k)_{k \geq 1}$ converges to an invertible matrix. Find A.

(5) Let B be a square matrix satisfying $\|B\| < 1$. Prove that the sequence $(C_k)_{k \geq 1}$, where

$$C_k = I + B + B^2 + \cdots + B^k$$

converges and that

$$\lim_{k \to \infty} C_k = (I - B)^{-1}.$$

(6) Prove that

$$AB = BA \Rightarrow \exp(A + B) = \exp(A)\exp(B).$$

(7) Let (A_k) be a sequence of $n \times n$ matrices. Show that the following conditions are equivalent:
(i) the sequence (A_k) converges;
(ii) for every vector $\mathbf{v} \in \mathbb{R}^n$, the sequence of vectors $(A_k \mathbf{v})$ converges in $\mathbb{R}^n$.

(8) Let A and B be square matrices. Assume that $\exp(A)\exp(B) = \exp(A + B)$. Show that in general $[A, B] \neq 0$.

(9) Extend the Taylor expansion for $\ln(1 + x)$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad -1 < x \leq 1$$

to $n \times n$ matrices.

1.17 Gram-Schmidt Orthonormalization

The *Gram-Schmidt algorithm* is as follows: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis in $\mathbb{C}^n$. We define

$$\begin{aligned}
\mathbf{w}_1 &:= \mathbf{v}_1 \\
\mathbf{w}_2 &:= \mathbf{v}_2 - \frac{(\mathbf{w}_1, \mathbf{v}_2)}{(\mathbf{w}_1, \mathbf{w}_1)}\mathbf{w}_1 \\
\mathbf{w}_3 &:= \mathbf{v}_3 - \frac{(\mathbf{w}_2, \mathbf{v}_3)}{(\mathbf{w}_2, \mathbf{w}_2)}\mathbf{w}_2 - \frac{(\mathbf{w}_1, \mathbf{v}_3)}{(\mathbf{w}_1, \mathbf{w}_1)}\mathbf{w}_1 \\
&\vdots \\
\mathbf{w}_n &:= \mathbf{v}_n - \frac{(\mathbf{w}_{n-1}, \mathbf{v}_n)}{(\mathbf{w}_{n-1}, \mathbf{w}_{n-1})}\mathbf{w}_{n-1} - \cdots - \frac{(\mathbf{w}_1, \mathbf{v}_n)}{(\mathbf{w}_1, \mathbf{w}_1)}\mathbf{w}_1.
\end{aligned}$$

Then the vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ form an orthogonal basis in $\mathbb{C}^n$. Normalizing these vectors yields an orthonormal basis in $\mathbb{C}^n$.

Example 1.41. Let

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

We find

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{w}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{w}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

These vectors are already normalized. ♣

Example 1.42. Let

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We find

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \quad \mathbf{w}_3 = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \end{pmatrix}, \quad \mathbf{w}_4 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

which after normalization gives the orthonormal basis

$$\left\{ \frac{1}{2}(1,1,1,1)^T, \frac{1}{\sqrt{12}}(1,1,1,-3)^T, \frac{1}{\sqrt{6}}(1,1,-2,0)^T, \frac{1}{\sqrt{2}}(1,-1,0,0)^T \right\}.$$ ♣

1.18 Groups

In the representation of groups as $n \times n$ matrices the Kronecker product plays a central role. We give a short introduction to group theory and then discuss the connection with the Kronecker product. For further reading in group theory we refer to the books of Miller [42], Baumslag and Chandler [6] and Steeb [56]. In sections 2.13 and 2.14 we give a more detailed introduction to representation theory and the connection with the Kronecker product.

Definition 1.48. A *group* G is a set of objects $\{g, h, k, \ldots\}$ (not necessarily countable) together with a binary operation which associates with any ordered pair of elements $g, h \in G$ a third element $gh \in G$. The binary operation (called *group multiplication)* is subject to the following requirements:
(1) There exists an element e in G called the *identity element* such that $ge = eg = g$ for all $g \in G$.
(2) For every $g \in G$ there exists in G an *inverse element* g^{-1} such that $gg^{-1} = g^{-1}g = e$.
(3) *Associative law.* The identity $(gh)k = g(hk)$ is satisfied for all $g, h, k \in G$.

Thus, any set together with a binary operation which satisfies conditions (1) - (3) is called a group.

If $gh = hg$ we say that the elements g and h *commute.* If all elements of G commute then G is a *commutative* or *abelian* group. If G has a finite number of elements it has *finite order* $n(G)$, where $n(G)$ is the number of elements. Otherwise, G has *infinite order.*

A *subgroup* H of G is a subset which is itself a group under the group multiplication defined in G. The subgroups G and $\{e\}$ are called *improper* subgroups of G. All other subgroups are *proper.*

If a group G consists of a finite number of elements, then G is called a *finite group*; otherwise, G is called an *infinite group.*

Example 1.43. The set of integers $\mathbb{Z}$ with addition as group composition is an infinite additive group with $e = 0$. ♣

Example 1.44. The set $\{ 1, -1 \}$ with multiplication as group composition is a finite abelian group with $e = 1$. ♣

Definition 1.49. Let G be a finite group. The number of elements of G is called the *dimension* or *order* of G.

Example 1.45. The order of the group of the $n \times n$ permutation matrices under matrix multiplication is $n!$. ♣

Theorem 1.23. *The order of a subgroup of a finite group divides the order of the group.*

This theorem is called *Lagrange's theorem.* For the proof we refer to the literature (Miller [42]).

A way to partition G is by means of *conjugacy classes.*

Definition 1.50. A group element h is said to be conjugate to the group element $k, h \sim k$, if there exists a $g \in G$ such that

$$k = ghg^{-1}.$$

It is easy to show that conjugacy is an equivalence relation, i.e., (1) $h \sim h$ (reflexive), (2) $h \sim k$ implies $k \sim h$ (symmetric), and (3) $h \sim k, k \sim j$ implies $h \sim j$ (transitive). Thus, the elements of G can be divided into *conjugacy classes* of mutually conjugate elements. The class containing e consists of just one element since

$$geg^{-1} = e$$

for all $g \in G$. Different conjugacy classes do not necessarily contain the same number of elements.

Let G be an abelian group. Then each conjugacy class consists of one group element each, since

$$ghg^{-1} = h, \quad \text{for all} \quad g \in G.$$

Let us now give a number of examples to illustrate the definitions given above.

Example 1.46. A *field* $\mathbb{F}$ is an (infinite) abelian group with respect to addition. The set of nonzero elements of a field forms a group with respect to multiplication, which is called a multiplicative group of the field. ♣

Example 1.47. A *linear vector space* over a field $\mathbb{F}$ (such as the real numbers $\mathbb{R}$) is an abelian group with respect to the usual addition of vectors. The group composition of two elements (vectors) $\mathbf{a}$ and $\mathbf{b}$ is their vector sum $\mathbf{a}+\mathbf{b}$. The identity element is the zero vector and the inverse of an element is its negative. ♣

Example 1.48. Let N be an integer with $N \geq 1$. The set

$$\{\, e^{2\pi i n/N} \; : \; n = 0, 1, \ldots, N-1\,\}$$

is an abelian (finite) group under multiplication since

$$\exp(2\pi i n/N)\exp(2\pi i m/N) = \exp(2\pi i(n+m)/N)$$

where $n, m = 0, 1, \ldots, N-1$. Note that $\exp(2\pi i n) = 1$ for $n \in \mathbb{N}$. We consider some special cases of N: For $N = 2$ we find the set $\{1, -1\}$ and for $N = 4$ we find $\{1, i, -1, -i\}$. These are elements on the unit circle in the complex plane. For $N \to \infty$ the number of points on the unit circle increases. As $N \to \infty$ we find the unitary group

$$U(1) := \left\{\; e^{i\alpha} \; : \; \alpha \in \mathbb{R} \;\right\}.$$

♣

Example 1.49. The two matrices

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}$$

form a finite abelian group of order two with matrix multiplication as group composition. The closure can easily be verified

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

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

The identity element is the 2×2 unit matrix. ♣

Example 1.50. Let $M = \{1, 2, \ldots, n\}$. Let $Bi(M, M)$ be the set of bijective mappings $\sigma : M \to M$ so that

$$\sigma : \{1, 2, \ldots, n\} \to \{p_1, p_2, \ldots, p_n\}$$

forms a group S_n under the composition of functions. Let S_n be the set of all the permutations

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}.$$

We say 1 is mapped into p_1, 2 into p_2, $\ldots$, n into p_n. The numbers $p_1, p_2, \ldots, p_n$ are a reordering of $1, 2, \ldots, n$ and no two of the p_j's $j = 1, 2 \ldots, n$ are the same. The inverse permutation is given by

$$\sigma^{-1} = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ 1 & 2 & \cdots & n \end{pmatrix}.$$

The product of two permutations σ and τ, with

$$\tau = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \\ 1 & 2 & \cdots & n \end{pmatrix}$$

is given by the permutation

$$\sigma \circ \tau = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}.$$

That is, the integer q_i is mapped to i by τ and i is mapped to p_i by σ, so q_i is mapped to p_i by $\sigma \circ \tau$. The identity permutation is

$$e = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}.$$

S_n has order $n!$. The group of all permutations on M is called the *symmetric group* on M which is nonabelian, if $n > 2$. ♣

Example 1.51. Let N be a positive integer. The set of all matrices

$$Z_{2\pi k/N} = \begin{pmatrix} \cos(2k\pi/N) & -\sin(2k\pi/N) \\ \sin(2k\pi/N) & \cos(2k\pi/N) \end{pmatrix}$$

where $k = 0, 1, 2, \ldots, N-1$, forms an abelian group under matrix multiplication. The elements of the group can be generated from the transformation

$$Z_{2k\pi/N} = \left(Z_{2\pi/N}\right)^k, \qquad k = 0, 1, 2, \ldots, N-1.$$

For example, if $N = 2$ the group consists of the elements

$$\{(Z_\pi)^0, (Z_\pi)^1\} \equiv \{-I_2, +I_2\}$$

where I_2 is the 2×2 unit matrix. This is an example of a cyclic group. ♣

Example 1.52. The set of all invertible $n \times n$ matrices form a group with respect to the usual multiplication of matrices. The group is called the *general linear group* over the real numbers $GL(n, \mathbb{R})$, or over the complex numbers $GL(n, \mathbb{C})$. This group together with its subgroups are the so-called *classical groups* which are Lie groups. ♣

Example 1.53. Let $\mathbb{C}$ be the complex plane. Let $z \in \mathbb{C}$. The set of Möbius transformations in $\mathbb{C}$ form a group called the *Möbius group* denoted by M where $m : \mathbb{C} \to \mathbb{C}$,

$$M := \{\ m(a,b,c,d)\ :\ a,b,c,d \in \mathbb{C}, \quad ad - bc \neq 0\ \}$$

and

$$m : z \mapsto z' = \frac{az+b}{cz+d}.$$

The condition $ad-bc \neq 0$ must hold for the transformation to be invertible. Here, $z = x+iy$, where $x, y \in \mathbb{R}$. This forms a group under the composition of functions: Let

$$m(z) = \frac{az+b}{cz+d}, \qquad \widetilde{m}(z) = \frac{ez+f}{gz+h}$$

where $ad-bc \neq 0$ and $eh-fg \neq 0$ ($e, f, g, h \in \mathbb{C}$). Consider the composition

$$\begin{aligned} m\left(\widetilde{m}(z)\right) &= \frac{a(ez+f)/(gz+h)+b}{c(ez+f)/(gz+h)+d} \\ &= \frac{aez+af+bgz+hb}{cez+cf+dgz+hd} \\ &= \frac{(ae+bg)z+(af+hb)}{(ce+dg)z+(cf+hd)}. \end{aligned}$$

Thus $m(\widetilde{m}(z))$ has the form of a Möbius transformation, since

$$\begin{aligned} &(ae+bg)(cf+hd) - (af+hb)(ce+dg) \\ &\qquad = aecf + aehd + bgcf + bghd - afce - afdg - hbce - hbdg \\ &\qquad = ad(eh-fg) + bc(gf-eh) \\ &\qquad = (ad-bc)(eh-fg) \neq 0. \end{aligned}$$

Thus we conclude that m is closed under composition. Associativity holds since we consider the multiplication of complex numbers. The identity element is given by

$$m(1,0,0,1) = z.$$

To find the inverse of $m(z)$ we assume that

$$m\left(\widetilde{m}(z)\right) = \frac{(ae+bg)z+(af+hb)}{(ce+dg)z+(cf+hd)} = z$$

so that

$$ae+bg = 1, \qquad af+hb = 0, \qquad ce+dg = 0, \qquad cf+hd = 1$$

and we find

$$e = \frac{d}{ad-bc}, \qquad f = -\frac{b}{ad-bc}, \qquad g = -\frac{c}{ad-bc}, \qquad h = \frac{a}{ad-bc}.$$

The inverse is thus given by

$$(z')^{-1} = \frac{dz-b}{-cz+a}.$$

♣

Example 1.54. Let $\mathbb{Z}$ be the abelian group of integers. Let E be the set of even integers. Obviously, E is an abelian group under addition and is a subgroup of $\mathbb{Z}$. Let C_2 be the cyclic group of order 2. Then

$$\mathbb{Z}/E \cong C_2 .$$

♣

We denote the mapping between two groups by ρ and present the following definitions

Definition 1.51. A mapping of a group G into another group G' is called a *homomorphism* if it preserves all combinatorial operations associated with the group G so that

$$\rho(a \cdot b) = \rho(a) * \rho(b)$$

$a, b \in G$ and $\rho(a),\ \rho(b) \in G'$. Here $\cdot$ and $*$ are the group compositions in G and G', respectively.

Example 1.55. There is a homomorphism ρ from $GL(2,\mathbb{C})$ into the Möbius group M given by

$$\rho : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto m(z) = \frac{az+b}{cz+d}.$$

We now check that ρ is indeed a homomorphism: Consider the 2×2 matrix

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

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. The matrices A form a group with matrix multiplication as group composition. We find

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

where $e, f, g, h \in \mathbb{C}$. Thus

$$\rho(AB) = \frac{(ae+bg)z + (af+bh)}{(ce+dg)z + (cf+dh)}$$

and

$$\rho(A) = \frac{az+b}{cz+d}, \qquad \rho(B) = \frac{ez+f}{gz+h}$$

so that

$$\rho(A) * \rho(B) = \frac{(ae+bg)z + (af+bh)}{(ce+dg)z+(cf+dh)}.$$

We have shown that $\rho(A \cdot B) = \rho(A) * \rho(B)$ and thus that ρ is a homomorphism. ♣

An extension of the Möbius group is as follows. Consider the transformation

$$\mathbf{v} = \frac{A\mathbf{w}+B}{C\mathbf{w}+D}$$

where $\mathbf{v} = (v_1, \ldots, v_n)^T$, $\mathbf{w} = (w_1, \ldots, w_n)^T$ (T transpose). A is an $n \times n$ matrix, B an $n \times 1$ matrix, C a $1 \times n$ matrix and D a 1×1 matrix. The $(n+1) \times (n+1)$ matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is invertible.

Example 1.56. An $n \times n$ permutation matrix is a matrix that has in each row and each column precisely one 1. There are $n!$ permutation matrices. The $n \times n$ permutation matrices form a group under matrix multiplication. Consider the symmetric group S_n given above. It is easy to see that the two groups are isomorphic. *Cayley's theorem* tells us that every finite group is isomorphic to a subgroup (or the group itself) of these permutation matrices. The six 3×3 permutation matrices are given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad F = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

We have

$$AA = A \quad AB = B \quad AC = C \quad AD = D \quad AE = E \quad AF = F$$

$$BA = B \quad BB = A \quad BC = D \quad BD = C \quad BE = F \quad BF = E$$

$$
\begin{array}{cccccc}
CA = C & CB = E & CC = A & CD = F & CE = B & CF = D \\
DA = D & DB = F & DC = B & DD = E & DE = A & DF = C \\
EA = E & EB = C & EC = F & ED = A & EE = D & EF = B \\
FA = F & FB = D & FC = E & FD = B & FE = C & FF = A\,.
\end{array}
$$

For the inverse we find

$$A^{-1} = A, \quad B^{-1} = B, \quad C^{-1} = C, \quad D^{-1} = E, \quad E^{-1} = D, \quad F^{-1} = F.$$

The order of a finite group is the number of elements of the group. Thus our group has order 6. *Lagrange's theorem* tells us that the order of a subgroup of a finite group divides the order of the group. Thus the subgroups must have order 3, 2, 1. From the group table we find the subgroups

$$\{A,\ D,\ E\}$$

$$\{A,\ B\}, \qquad \{A,\ C\}, \qquad \{A,\ F\}$$

$$\{A\}.$$

Cayley's theorem tells us that every finite group is isomorphic to a subgroup (or the group itself) of these permutation matrices. The *order of an element* $g \in G$ is the order of the cyclic subgroup generated by $\{g\}$, i.e. the smallest positive integer m such that

$$g^m = e$$

where e is the identity element of the group. The integer m divides the order of G. Consider, for example, the element D of our group. Then

$$D^2 = E, \qquad D^3 = A, \qquad A \quad \text{identity element.}$$

Thus $m = 3$. ♣

Example 1.57. Let $c \in \mathbb{R}$ and $c \neq 0$. The 2×2 matrices

$$\begin{pmatrix} c & c \\ c & c \end{pmatrix}$$

form a group under matrix multiplication. Multiplication of two such matrices yields

$$\begin{pmatrix} c_1 & c_1 \\ c_1 & c_1 \end{pmatrix} \begin{pmatrix} c_2 & c_2 \\ c_2 & c_2 \end{pmatrix} = \begin{pmatrix} 2c_1c_2 & 2c_1c_2 \\ 2c_1c_2 & 2c_1c_2 \end{pmatrix}.$$

The neutral element is the matrix

$$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

and the inverse element is the matrix

$$\begin{pmatrix} 1/(4c) & 1/(4c) \\ 1/(4c) & 1/(4c) \end{pmatrix}.$$

♣

Exercises. (1) Show that all $n \times n$ permutation matrices form a group under matrix multiplication.

(2) Find all subgroups of the group of the 4×4 permutation matrices. Apply Lagrange's theorem.

(3) Consider the matrices

$$A(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \qquad \alpha \in \mathbb{R}.$$

Show that these matrices form a group under matrix multiplication.

(4) Consider the matrices

$$B(\alpha) = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}, \qquad \alpha \in \mathbb{R}.$$

Show that these matrices form a group under matrix multiplication.

(5) Consider the matrices given in (3). Find

$$X = \frac{d}{d\alpha} A(\alpha)\bigg|_{\alpha=0}.$$

Show that

$$\exp(\alpha X) = A(\alpha).$$

(6) Let S be the set of even integers. Show that S is a group under addition of integers.

(7) Show that all 2×2 matrices

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \qquad a \in \mathbb{R}$$

form a group under matrix multiplication.

(8) Let S be the set of real numbers of the form $a + b\sqrt{2}$, where $a, b \in \mathbb{Q}$ and are not simultaneously zero. Show that S is a group under the usual multiplication of real numbers.

1.19 Lie Algebras

Lie algebras (Bäuerle and de Kerf [7], Humphreys [32], Jacobson [33]) play a central role in theoretical physics. They are also linked to Lie groups and Lie transformation groups (Steeb [56]). In this section we give the definition of a Lie algebra and some applications.

Definition 1.52. A vector space L over a field $\mathbb{F}$, with an operation $L \times L \to L$ denoted by

$$(x, y) \to [x, y]$$

and called the commutator of x and y, is called a *Lie algebra* over $\mathbb{F}$ if the following axioms are satisfied:

(L1) The bracket operation is bilinear.

(L2) $[x, x] = 0 \qquad \text{for all} \quad x \in L$

(L3) $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \qquad (x, y, z \in L).$

Remark: Axiom (L3) is called the *Jacobi identity.*

Notice that (L1) and (L2), applied to $[x+y, x+y]$, imply anticommutativity: (L2')

$$[x, y] = -[y, x].$$

Conversely, if $\text{char}\mathbb{F} \neq 2$ (for $\mathbb{R}$ and $\mathbb{C}$ we have $\text{char}\mathbb{F} = 0$), then (L2') will imply (L2).

Definition 1.53. Let X and Y be $n \times n$-matrices. Then the *commutator* $[X, Y]$ of X and Y is defined as

$$[X, Y] := XY - YX.$$

The $n \times n$ matrices over $\mathbb{R}$ or $\mathbb{C}$ form a Lie algebra under the commutator. This means we have the following properties (X, Y, V, W $n \times n$ matrices and $c \in \mathbb{C}$)

$$[cX, Y] = c[X, Y], \qquad [X, cY] = c[X, Y]$$

$$[X, Y] = -[Y, X]$$

$$[X + Y, V + W] = [X, V] + [X, W] + [Y, V] + [Y, W]$$

and

$$[X, [Y, V]] + [V, [X, Y]] + [Y, [V, X]] = 0.$$

The last equation is called the *Jacobi identity*..

Definition 1.54. Two Lie algebras L, L' are called *isomorphic* if there exists a vector space isomorphism $\phi : L \to L'$ satisfying

$$\phi([x,y]) = [\phi(x), \phi(y)]$$

for all $x, y \in L$.

The Lie algebra of all $n \times n$ matrices over $\mathbb{C}$ is also called $gl(n, \mathbb{C})$. A basis is given by the matrices

$$(E_{ij}), \qquad i, j = 1, 2, \ldots, n$$

where (E_{ij}) is the matrix having 1 in the (i,j) position and 0 elsewhere. Since

$$(E_{ij})(E_{kl}) = \delta_{jk}(E_{il})$$

it follows that the commutator is given by

$$[(E_{ij}), (E_{kl})] = \delta_{jk}(E_{il}) - \delta_{li}(E_{kj})\,.$$

Thus the coefficients are all ± 1 or 0.

The classical Lie algebras are sub-Lie algebras of $gl(n, \mathbb{C})$. For example, $sl(n, \mathbb{R})$ is the Lie algebra with the condition

$$\mathrm{tr}(X) = 0$$

for all $X \in gl(n, \mathbb{R})$. Furthermore, $so(n, \mathbb{R})$ is the Lie algebra with the condition that

$$X^T = -X \quad \text{and} \quad \mathrm{tr}(X) = 0$$

for all $X \in so(n, \mathbb{R})$. For $n = 2$ a basis element is given by

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

For $n = 3$ we have a basis (skew-symmetric matrices)

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Exercises. (1) Show that the 2×2 matrices with trace zero form a Lie algebra under the commutator. Show that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

form a basis. Hint: Show that

$$\operatorname{tr}([A, B]) = 0$$

for any two $n \times n$ matrices.

(2) Find all Lie algebras with dimension 2.

(3) Show that the set of all diagonal matrices form a Lie algebra under the commutator.

(4) Do all Hermitian matrices form a Lie algebra under the commutator?

(5) Do all skew-Hermitian matrices form a Lie algebra under the commutator?

(6) An *automorphism* of L is an isomorphism of L onto itself. Let $L = sl(n, \mathbb{R})$. Show that if $g \in GL(n, \mathbb{R})$ and if

$$gLg^{-1} = L$$

then the map

$$x \mapsto gxg^{-1}$$

is an automorphism.

(7) The *center* of a Lie algebra L is defined as

$$Z(L) := \{\, z \in L \,:\, [z, x] = 0 \quad \text{for all}\, x \in L \,\}\,.$$

Find the center for the Lie algebra $sl(2, \mathbb{R})$.

1.20 Commutators and Anti-Commutators

Let A and B be $n \times n$ matrices. Then we define the *commutator* of A and B as

$$[A, B] := AB - BA .$$

For all $n \times n$ matrices A, B, C we have the *Jacobi identity*

$$[A, [B, C]] + [C, [A, B]] + [B, [C, A]] = 0_n$$

where 0_n is the $n \times n$ zero matrix. Since $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ we have

$$\mathrm{tr}([A, B]) = 0 .$$

If

$$[A, B] = 0_n$$

we say that the matrices A and B *commute*. For example, if A and B are diagonal matrices then the commutator is the zero matrix 0_n.

Let A and B be $n \times n$ matrices. Then we define the *anticommutator* of A and B as

$$[A, B]_+ := AB + BA .$$

We have

$$\mathrm{tr}([A, B]_+) = 2\mathrm{tr}(AB) .$$

The anticommutator plays a role for Fermi operators.

Example 1.58. Consider the Pauli spin matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then

$$[\sigma_x, \sigma_z] = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} = 2i\sigma_y$$

and

$$[\sigma_x, \sigma_z]_+ = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

♣

1.21 Functions of Matrices

Let A be an $n \times n$ matrix over $\mathbb{C}$. Let $f : \mathbb{R} \to \mathbb{R}$ be an analytic function. Then we can consider the matrix valued function $f(A)$. The simplest example of such a function is a polynomial. Let p be a polynomial

$$p(x) = \sum_{j=0}^{n} c_j x^j$$

then the corresponding matrix function of an $n \times n$ matrix A can be defined by

$$p(A) = \sum_{j=0}^{n} c_j A^j$$

with the convention $A^0 = I_n$. If a function f of a complex variable z has a *MacLaurin series expansion*

$$f(z) = \sum_{j=0}^{\infty} c_j z^j$$

which converges for $|z| < R$ $(R > 0)$, then the matrix series

$$f(A) = \sum_{j=0}^{\infty} c_j A^j$$

converges, provided A is square and each of its eigenvalues has absolute value less than R. More generally if a function f of a complex variable z has a *power series expansion*

$$f(z) = \sum_{j=0}^{\infty} c_j z^j$$

which converges for $z \in Z \subseteq \mathbb{C}$, then the matrix series

$$f(A) = \sum_{j=0}^{\infty} c_j A^j$$

converges, provided A is square and each of its eigenvalues are in Z.

Example 1.59. The most used matrix function is $\exp(A)$ defined by

$$\exp(A) := \sum_{j=0}^{\infty} \frac{A^j}{j!} \equiv \lim_{n \to \infty} \left(I_n + \frac{A}{n} \right)^n$$

which converges for all square matrices A. ♣

Example 1.60. The matrix functions $\sin(A)$ defined by

$$\sin(A) := \sum_{j=0}^{\infty}(-1)^j \frac{A^{2j+1}}{(2j+1)!}, \qquad \cos(A) := \sum_{j=0}^{\infty}(-1)^j \frac{A^{2j}}{(2j)!}$$

converge for all square matrices A. ♣

Example 1.61. The matrix function $\arctan(A)$ defined by

$$\arctan(A) := \sum_{j=0}^{\infty}(-1)^j \frac{A^{2j+1}}{2j+1}$$

converges for square matrices A with eigenvalues λ satisfying $|\lambda| < 1$. ♣

Example 1.62. The matrix function $\ln(A)$ defined by

$$\ln(A) := -\sum_{j=1}^{\infty} \frac{(I_n - A)^j}{j}$$

converges for square matrices A with eigenvalues λ where $|\lambda| \in (0, 2)$. ♣

Example 1.63. We calculate

$$\exp\begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix}$$

using MacLaurin series expansion to obtain the exponential. Since the square of the matrix is the 3×3 zero matrix we find from the MacLaurin series expansion

$$\exp\begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix} = \begin{pmatrix} 1 & i\pi & 0 \\ 0 & 1 & 0 \\ 0 & -i\pi & 1 \end{pmatrix}.$$

♣

We can also use the Cayley-Hamilton theorem (see section 1.6). Since A satisfies it's own characteristic equation

$$A^n - \mathrm{tr}(A)A^{n-1} + \cdots + (-1)^n \det(A)A^0 = 0_n$$

where 0_n is the $n \times n$ zero matrix, we find that A^n can be expressed in terms of A^{n-1}, ..., A and I_n

$$A^n = \mathrm{tr}(A)A^{n-1} - \cdots - (-1)^n \det(A)I_n.$$

Then the MacLaurin series contracts to

$$f(A) = \sum_{j=0}^{\infty} c_j A^j = \sum_{j=0}^{n-1} \alpha_j A^j.$$

The values $\alpha_0, \ldots, \alpha_{n-1}$ can be determined from the eigenvalues of A. Let λ be an eigenvalue of A corresponding to the eigenvector $\mathbf{x}$. Then $f(A)\mathbf{x} = f(\lambda)\mathbf{x}$ and

$$f(A)\mathbf{x} = \sum_{j=0}^{\infty} c_j A^j \mathbf{x} = \sum_{j=0}^{n-1} \alpha_j A^j \mathbf{x} = \sum_{j=0}^{n-1} \alpha_j \lambda_j^j \mathbf{x}.$$

Since $\mathbf{x}$ is nonzero we have the equation

$$f(\lambda) = \sum_{j=0}^{n-1} \alpha_j \lambda_j^j.$$

Thus we have a system of linear equations for the α_j. If an eigenvalue λ of A has multiplicity greater than 1, these equations will be insufficient to determine all of the α_j. Since f is analytic the derivative f' is also analytic and

$$f'(A) = \sum_{j=1}^{\infty} j c_j A^{j-1} = \sum_{j=1}^{n-1} j \alpha_j A^{j-1}$$

which provides the equation

$$f'(\lambda) = \sum_{j=1}^{n-1} j \alpha_j \lambda_j^j.$$

If the multiplicity is of λ is k then we have the equations

$$\begin{aligned}
f(\lambda) &= \sum_{j=0}^{n-1} \alpha_j \lambda_j^j \\
f'(\lambda) &= \sum_{j=1}^{n-1} j \alpha_j \lambda_j^{j-1} \\
&\vdots \\
f^{(k-1)}(\lambda) &= \sum_{j=k-1}^{n-1} \left(\prod_{m=0}^{k-1} (j-m) \right) \alpha_j \lambda_j^{j-k-1}.
\end{aligned}$$

Example 1.64. We calculate

$$\exp \begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix}.$$

using the Cayley-Hamilton theorem. The eigenvalue is 0 with multiplicity 3. Thus we solve the equations

$$\begin{aligned} e^\lambda &= \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 \\ e^\lambda &= \alpha_1 + 2\alpha_2\lambda \\ e^\lambda &= 2\alpha_2 \end{aligned}$$

where $\lambda = 0$. Thus $\alpha_2 = \frac{1}{2}$ and $\alpha_0 = \alpha_1 = 1$. Consequently

$$\exp\begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & i\pi & 0 \\ 0 & 0 & 0 \\ 0 & -i\pi & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & i\pi & 0 \\ 0 & 1 & 0 \\ 0 & -i\pi & 1 \end{pmatrix}.$$ ♣

Example 1.65. Consider the matrix

$$A = \begin{pmatrix} \frac{\pi}{\sqrt{2}} & 0 & 0 & \frac{\pi}{\sqrt{2}} \\ 0 & \frac{\pi}{\sqrt{2}} & \frac{\pi}{\sqrt{2}} & 0 \\ 0 & \frac{\pi}{\sqrt{2}} & -\frac{\pi}{\sqrt{2}} & 0 \\ \frac{\pi}{\sqrt{2}} & 0 & 0 & -\frac{\pi}{\sqrt{2}} \end{pmatrix}$$

with eigenvalues are π, π, $-\pi$ and $-\pi$. We wish to calculate $\sec(A)$ where $\sec(z)$ which is analytic on the intervals $(\pi/2, 3\pi/2)$ and $(-3\pi/2, -\pi/2)$ amongst others. The eigenvalues lie within these intervals. However, the power series for $\sec(z)$ on these two intervals are different, the methods above do not apply directly. We can restrict ourselves to the action on subspaces. The eigenvalue π corresponds to the eigenspace spanned by

$$\left\{ \frac{1}{\sqrt{4-2\sqrt{2}}}(1,0,0,\sqrt{2}-1)^T,\ \frac{1}{\sqrt{4-2\sqrt{2}}}(0,1,\sqrt{2}-1,0)^T \right\}.$$

The projection onto this eigenspace is given by

$$\Pi_1 := \frac{1}{4-2\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 & \sqrt{2}-1 \\ 0 & 1 & \sqrt{2}-1 & 0 \\ 0 & \sqrt{2}-1 & 3-2\sqrt{2} & 0 \\ \sqrt{2}-1 & 0 & 0 & 3-2\sqrt{2} \end{pmatrix}.$$

On this two dimensional subspace, A has one eigenvalue with multiplicity 2. We solve the equation $\sec(\pi) = -1 = \alpha_0 + \alpha_1\pi$ and $\sec'(\pi) = \sec(\pi)\tan(\pi) = 0 = \alpha_1$. Thus $\alpha_1 = 0$ and $\alpha_0 = -1$. The solution on this subspace is $(\alpha_0 I_4 + \alpha_1 A)\Pi_1 = -\Pi_1$. We perform a similar calculation for the eigenvalue $-\pi$ and the projection onto the eigenspace Π_2. We solve the equation $\sec(-\pi) = -1 = \beta_0 - \beta_1\pi$ and $\sec'(-\pi) = \sec(-\pi)\tan(-\pi) = 0 = -\beta_1$. Thus $\beta_1 = 0$ and $\beta_0 = -1$. The solution on this subspace is $(\beta_0 I_4 + \beta_1 A)\Pi_2 = -\Pi_2$. The solution on $\mathbb{C}^4$ is given by

$$\sec(A) = (-\Pi_1) + (-\Pi_2) = -I_4.$$ ♣