Linear Algebra Theorems and Definitions

Alexander J. Clarke

February 22, 2024

Contents

Chapter 0

List of Symbols

A_{ij}	the ij-th entry of the matrix A
A^{-1}	the inverse of the matrix A
$A^{\dagger}$	the pseudoinverse of the matrix A
A^*	the adjoint of the matrix A
$\tilde{A}_{ij}$	the matrix A with row i and column j deleted
A^t	the transpose of the matrix A
$(A|B)$	the matrix A augmented by the matrix B
$B_1 \bigoplus \cdots \bigoplus B_k$	the direct sum of matrices B_1 through B_k
$\mathcal{B}(V)$	the set of bilinear forms on V
β^*	the dual basis of β
β_x	the T-cyclic basis generated by x
$\mathbb{C}$	the field of complex numbers
$\mathbb{C}_i$	the ith Gerschgorin disk
$\operatorname{cond}(A)$	the condition number of the matrix A
$\mathsf{C}^n(\mathbb{R})$	set of functions f on $\mathbb{R}$ with $f^{(n)}$ continuous
C^{∞}	set of functions with derivatives of every order
$\mathsf{C}(\mathbb{R})$	the vector space of continuous functions on $\mathbb{R}$
$\mathsf{C}([0,1])$	the vector space of continuous functions on $[0,1]$
C_x	the T-cyclic subspaces generated by x
D	the derivative operator on C^{∞}
$\det(A)$	the determinant of the matrix A
δ_{ij}	the Kronecker delta
$\dim(V)$	the dimension of V
e^A	$\displaystyle \lim_{m \to \infty} \left(I + A + \frac{A^2}{2!} + \cdots + \frac{A^m}{m!} \right)$
e_i	the ith standard vector of $\mathbb{F}^n$

E_λ	the eigenspace of T corresponding to λ
$\mathbb{F}$	a field
$f(A)$	the polynomial $f(x)$ evaluated at the matrix A
F^n	the set of n-tuples with entries in a field $\mathbb{F}$
$f(T)$	the polynomial $f(x)$ evaluated at the operator T
$\mathcal{F}(S,\mathbb{F})$	the set of functions from S to a field $\mathbb{F}$
H	space of continuous complex functions on $[0,2\pi]$
I_n or I	the $n \times n$ identity matrix
$\mathbb{I}_V$ or $\mathbb{I}$	the identity operator on V
K_λ	generalized eigenspace of T corresponding to λ
K_ϕ	$\{x : (\phi(T))^p(x) = 0, \text{ for some positive integer } p\}$
L_A	left-multiplication transformation by matrix A
$\lim\limits_{m \to \infty} A_m$	the limit of a sequence of matrices
$\mathcal{L}(V)$	the space of linear transformations from V to V
$\mathcal{L}(V,W)$	the space of linear transformations from V to W
$M_{m \times n}(\mathbb{F})$	the set of $m \times n$ matrices with entries in $\mathbb{F}$
$v(A)$	the column sum of the matrix A
$v_j(A)$	the jth column sum of the matrix A
$N(T)$	the null space of T
nullity (T)	the dimension of the null space of T
O	the zero matrix
per (M)	the permanent of the 2×2 matrix M
$P(\mathbb{F})$	the space of polynomials with coefficients in $\mathbb{F}$
$P_n(\mathbb{F})$	the polynomials in $P(\mathbb{F})$ of degree at most n
ϕ_β	the standard representation with respect to basis β
$\mathbb{R}$	the field of real numbers
rank (A)	the rank of the matrix A
rank (T)	the rank of the linear transformation T
$\rho(A)$	the row sum of the matrix A
$\rho_i(A)$	the ith row sum of the matrix A
$R(T)$	the range of the linear transformation T
$S_1 + S_2$	the sum of sets S_1 and S_2
span (S)	the span of the set S
$S^\perp$	the orthogonal complement of the set S
$[T]_\beta$	the matrix representation of T in basis β
$[T]_\beta^\gamma$	the matrix representation of T in bases β and γ
T^{-1}	the inverse of the linear transformation T

$T^\dagger$	the pseudoinverse of the linear transformation T
T^*	the adjoint of the linear operator T
T_0	the zero transformation
T^t	the transpose of the linear transformation T
T_θ	the rotation transformation by θ
T_W	the restriction of T to a subspace W
$\text{tr}\,(A)$	the trace of the matrix A
V^*	the dual space of the vector space V
V/W	the quotient space of V modulo W
$W_1 + \cdots + W_k$	the sum of subspaces W_1 through W_k
$\displaystyle\sum_{i=1}^{k} W_i$	the sum of subspaces W_i through W_k
$W_1 \bigoplus W_2$	the direct sum of subspaces W_1 and W_2
$W_1 \bigoplus \cdots \bigoplus W_k$	the direct sum of subspaces W_1 through W_k
$\|x\|$	the norm of the vector $\vec{x}$
$[x]_\beta$	the coordinate vector of x relative to β
$\langle x, y \rangle$	the inner product of $\vec{x}$ and $\vec{y}$
$\mathbb{Z}_2$	the field consisting of 0 and 1
$\overline{\vec{z}}$	the complex conjugate of $\vec{z}$
$\vec{0}$	the zero vector

Chapter 1

Vector Spaces

1   Introduction

Definition 1.1 (Parallelogram Law for Vector Addition).
The sum of two vectors x and y that act at the same point P is the vector beginning at P that is represented by the diagonal of parallelogram having x and y as adjacent sides.

Definition 1.2.
Two nonzero vectors x and y are called **parallel** if $y = tx$ for some nonzero real number t. (Thus nonzero vectors having the same or opposite directions are parallel.)

2   Vector Spaces

Definition 1.3.
A **vector space** (or **linear space**) V over a field $\mathbb{F}$ consists of a set on which two operations (called **addition** and **scalar multiplication**, respectively) are defined so that for each pair of elements x and y in V there is a unique element a in $\mathbb{F}$ and each element x in V there is a unique element ax in V, such that the following conditions hold:

(VS 1) For all x, y in V, $x + y = y + x$ (commutativity of addition).

(VS 2) For all x, y in V, $(x + y) + z = x + (y + z)$ (associativity of addition).

(VS 3) There exists an element in V denoted by 0 such that $x + 0 = x$ for each x in V.

(VS 4) For each element x in V there exists an element y in V such that $x + y = 0$.

(VS 5) For each element x in V, $1x = x$.

(VS 6) For each pair of elements a, b in $\mathbb{F}$ and each element x in V, $(ab)x = a(bx)$.

(VS 7) For each element a in $\mathbb{F}$ and each pair of elements x, y in V, $a(x + y) = ax + ay$.

(VS 8) For each pair of elements a, b in $\mathbb{F}$ and each element x in V, $(a + b)x = ax + bx$.

The elements $x + y$ and ax are called the **sum** of x and y and the **product** of a and x, respectively.

The elements of the field $\mathbb{F}$ are called **scalars** and the elements of the vector space V are called **vectors**.

Note: The reader should not confuse this use of the word "vector" with the physical entity discussed in section 1.1: the word "vector" is now being used to describe any element of a vector space.

Definition 1.4.
An object of the form $(a_1, a_2, \ldots, a_n)$, where the entries $a_1, a_2, \ldots, a_n$ are elements of a field $\mathbb{F}$, is called an ***n*-tuple** with entries from $\mathbb{F}$. The elements $a_1, a_2, \ldots, a_n$ are called the **entries** or **components** of the n-tuple. Two n-tuples $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ with entries from a field $\mathbb{F}$ are called **equal** if $a_i = b_i$ for $i = 1, 2, \ldots, n$.

Definition 1.5.
Vectors in $\mathbb{F}^n$ may be written as **column vectors**

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

rather than as **row vectors** $(a_1, a_2, \ldots, a_n)$. Since a 1-tuple whose only entry is from $\mathbb{F}$ can be regarded as an element of $\mathbb{F}$, we usually write $\mathbb{F}$ rather than $\mathbb{F}^1$ for the vector space of 1-tuples with entry from $\mathbb{F}$.

Definition 1.6.
An $m \times n$ **matrix** with entries from a field $\mathbb{F}$ is a rectangular array of the form

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

where each entry a_{ij} ($1 \le i \le m$, $1 \le j \le n$) is an element of $\mathbb{F}$. We call the entries a_{ij} with $i = j$ the **diagonal entries** of the matrix. The entries $a_{i1}, a_{i2}, \ldots, a_{in}$ compose the ***i*th row** of the matrix, and the entries $a_{1j}, a_{2j}, \ldots, a_{mj}$ compose the ***j*th column** of the matrix. The rows of the preceding matrix are regarded as vectors in $\mathbb{F}^n$, and the columns are regarded as vectors in $\mathbb{F}^m$. The $m \times n$ matrix in which each entry equals zero is called the **zero matrix** and is denoted by O.

In this book, we denote matrices by capital italic letters (e.g. A, B, and C), and we denote the entry of a matrix A that lies in row i and column j by A_{ij}. In addition, if the number of rows and columns of a matrix are equal, the matrix is called **square**.

Two $m \times n$ matrices A and B are called **equal** if all their corresponding entries are equal, that is, if $A_{ij} = B_{ij}$ for $1 \le i \le m$ and $1 \le j \le n$.

Definition 1.7.
The set of all $m \times n$ matrices with entries from a field $\mathbb{F}$ is a vector space which we denote by $M_{m \times n}(\mathbb{F})$, with the following operations of **matrix addition** and **scalar multiplication**: For $A, B \in M_{m \times n}(\mathbb{F})$ and $c \in \mathbb{F}$,

$$(A + B)_{ij} = A_{ij} + B_{ij} \quad \text{and} \quad (cA)_{ij} = cA_{ij}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Definition 1.8.
Let S be any nonempty set and $\mathbb{F}$ be any field, and let $\mathcal{F}(S, \mathbb{F})$ denote the set of all functions from S to $\mathbb{F}$. Two functions f and g in $\mathcal{F}(S, \mathbb{F})$ are called **equal** if $f(s) = g(s)$ for each $s \in S$. The set $\mathcal{F}(S, \mathbb{F})$ is a vector space with the operations of addition and scalar multiplication defined for $f, g \in \mathcal{F}(S, \mathbb{F})$ and $c \in \mathbb{F}$ defined by

$$(f + g)(s) = f(s) + g(s) \quad \text{and} \quad (cf)(s) = c[f(s)]$$

for each $s \in S$. Note that these are the familiar operations of addition and scalar multiplication for functions used in algebra and calculus.

Definition 1.9.
A **polynomial** with coefficients from a field $\mathbb{F}$ is an expression of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where n is a non-negative integer and each a_k, called the **coefficient** of x^k, is in $\mathbb{F}$. If $f(x) = 0$, that is, if $a_n = a_{n-1} = \cdots = a_0 = 0$, then $f(x)$ is called the **zero polynomial** and, for convenience, its degree is defined to be -1; otherwise, the **degree** of a polynomial is defined to be the largest exponent of x that appears in the representation

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with a nonzero coefficient. Note that the polynomials of degree zero may be written in the form $f(x) = c$ for some nonzero scalar c. Two polynomials,

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

and

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_1 x + b_0,$$

are called **equal** if $m = n$ and $a_i = b_i$ for $i = 1, 2, \ldots, n$.

Definition 1.10.
Let $\mathbb{F}$ be any field. A **sequence** in $\mathbb{F}$ is a function σ from the positive integers into $\mathbb{F}$. In this book, the sequence σ such that $\sigma(n) = a_n$ for $n = 1, 2, \ldots$ is denoted $\{a_n\}$. Let V consist of all sequences $\{a_n\}$ in $\mathbb{F}$ that have only a finite number of nonzero terms a_n. If $\{a_n\}$ and $\{b_n\}$ are in V and $t \in \mathbb{F}$, define

$$\{a_n\} + \{b_n\} = \{a_n + b_n\} \quad \text{and} \quad t\{a_n\} = \{ta_n\}$$

Theorem 1.1 (Cancellation Law for Vector Addition).
If x, y and z are vectors in a vector space V such that $x + z = y + z$, then $x = y$.

Corollary 1.1.
The vector 0 described in (VS 3) is unique.

Corollary 1.2.
The vector y described in (VS 4) is unique.

Definition 1.11.
The vector 0 in (VS 3) is called the **zero vector** of V, and the vector y in (VS 4) (that is, the unique vector such that $x + y = 0$) is called the **additive inverse** of x and is denoted by $-x$.

Theorem 1.2.
In any vector space V, the following statements are true:

1. $0x = 0$ for each $x \in V$.

2. $(-a)x = -(ax) = a(-x)$ for each $a \in \mathbb{F}$ and each $x \in V$.

3. $a0 = 0$ for each $a \in \mathbb{F}$.

Definition 1.12.
Let $V = \{0\}$ consist of a single vector 0 and define $0 + 0 = 0$ and $c0 = 0$ for each scalar $c \in \mathbb{F}$. Then V is called the **zero vector space**.

Definition 1.13.
A real-valued function f defined on the real line is called an **even function** if $f(-t) = f(t)$ for each real number t, and is called an **odd function** if $f(-t) = -f(t)$ for each real number t.

Definition 1.14.
If S_1 and S_2 are nonempty subsets of a vector space V, then the **sum** of S_1 and S_2, denoted $S_1 + S_2$, is the set $\{x + y : x \in S_1, \text{ and } y \in S_2\}$.

Definition 1.15.
A vector space V is called the **direct sum** of W_1 and W_2 if W_1 and W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. We denote that V is the direct sum of W_1 and W_2 by writing $V = W_1 \oplus W_2$.

Definition 1.16.
A matrix M is called **skew-symmetric** if $M^t = -M$.

Definition 1.17.
Let W be a subspace of a vector space V over a field $\mathbb{F}$. For any $v \in V$, the set $\{v\} + W = \{v + w : w \in W\}$ is called the **coset of W containing** v. It is customary to denote this coset by $v + W$ rather than $\{v\} + W$.

Definition 1.18.
Let W be a subspace of a vector space V over a field $\mathbb{F}$, and let $S := \{v + W : v \in V\}$ be the set of all cosets of W. Then S is called the **quotient space of V modulo W**, and is denoted by V/W. Addition and scalar multiplication by the scalars of $\mathbb{F}$ can be defined as follows:

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$$

for all $v_1, v_2 \in V$, and

$$a(v + W) = av + W$$

for all $v \in V$ and $a \in \mathbb{F}$.

3   Subspaces

Definition 1.19.
A subset W of a vector space V over a field $\mathbb{F}$ is called a **subspace** of V if W is a vector space over $\mathbb{F}$ with the operations of addition and scalar multiplication defined on V.

In any vector space V, note that V and $\{0\}$ are subspaces. The latter is called the **zero subspace** of V.

Fortunately, it is not necessary to verify all of the vector space properties to prove that a subset is a subspace. Because properties (VS 1), (VS 2), (VS 5), (VS 6), (VS 7) and (VS 8) hold for all vectors in the vector space, these properties automatically hold for the vectors in any subset. Thus a subset W of a vector space V is a subspace of V if and only if the following four properties hold:

1. $x + y \in W$ whenever $x \in W$ and $y \in W$. (W is **closed under addition**).

2. $cx \in W$ whenever $c \in \mathbb{F}$ and $x \in W$. (W is **closed under scalar multiplication**).

3. W has a zero vector.

4. Each vector in W has an additive inverse in W.1

Theorem 1.3.
Let V be a vector space and W a subset of V. Then W is a subspace of V if and only if the following three conditions hold for the operations defined in V.

1. $0 \in W$.

2. $x + y \in W$ whenever $x \in W$ and $y \in W$.

3. $cx \in W$ whenever $c \in \mathbb{F}$ and $x \in W$.

Definition 1.20.
The **transpose** A^t of an $m \times n$ matrix A is the $n \times m$ matrix obtained from A by interchanging the rows with the columns; that is, $(A^t)_{ij} = A_{ji}$.

Definition 1.21.
A **symmetric matrix** is a matrix A such that $A^t = A$.

Definition 1.22.
An $n \times n$ matrix M is called a **diagonal matrix** if $M_{ij} = 0$ whenever $i \neq j$; that is, if all nondiagonal entries are zero.

Definition 1.23.
The **trace** of an $n \times n$ matrix M, denoted $\text{tr}(M)$, is the sum of the diagonal entries of M; that is,

$$\text{tr}(M) = M_{11} + M_{22} + \cdots + M_{nn}.$$

Theorem 1.4.
Any intersection of subspaces of a vector space V is a subspace of V.

Definition 1.24.
An $m \times n$ matrix A is called **upper triangular** if all entries lying below the diagonal entries are zero; that is, if $A_{ij} = 0$ whenever $i > j$.

4 Linear Combinations and Systems of Linear Equations

Definition 1.25.
Let V be a vector space and S a nonempty subset of V. A vector $v \in V$ is called a **linear combination** of vectors of S if there exist a finite number of vectors $v_1, v_2, \ldots, v_n$ in S and scalars $a_1, a_2, \ldots, a_n$ in $\mathbb{F}$ such that $v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$. In this case we also say that v is a linear combination of $v_1, v_2, \ldots, v_n$ and call $a_1, a_2, \ldots, a_n$ the **coefficients** of the linear combination.

Definition 1.26.
Let S be a nonempty subset of a vector space V. The **span** of S, denoted $\text{span}(S)$, is the set consisting of all linear combinations of the vectors in S. For convenience, we define $\text{span}(\emptyset) = \{0\}$.

Theorem 1.5.
The span of any subset S of a vector space V is a subspace of V. Moreover, any subspace of V that contains S must also contain the span of S.

Definition 1.27.
A subset S of a vector space V **generates** (or **spans**) V if $\text{span}(S) = V$. In this case, we also say that the vectors of S generate (or span) V.

5 Linear Dependence and Linear Independence

Definition 1.28.
A subset S of a vector space V is called **linearly dependent** if there exist a finite number of distinct vectors $v_1, v_2, \ldots, v_n$ in S and scalars $a_1, a_2, \ldots, a_n$ not all zero, such that

$$a_1 v_2 + a_2 v_2 + \cdots + a_n v_n = 0$$

In this case, we also say that the vectors of S are linearly dependent.

For any vectors $v_1, v_2, \ldots, v_n$, we have $a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = 0$ if $a_1 = a_2 = \cdots = a_n = 0$. We call this the **trivial representation** of 0 as a linear combination of $v_1, v_2, \ldots, v_n$. Thus, for a set tot be linearly dependent, there must exist a nontrivial representation of 0 as a linear combination of vectors in the set. Consequently, any subset of a vector space that contains the zero vector is linearly dependent, because $0 = 1 \cdot 0$ is a nontrivial representation of 0 as a linear combination of vectors in the set.

Definition 1.29.
A subset S of a vector space that is not linearly dependent is called **linearly independent**. As before, we also say that the vectors of S are linearly independent.

The following facts about linearly independent sets are true in any vector space.

1. The empty set is linearly independent, for linearly dependent sets must be nonempty.

2. A set consisting of a single nonzero vector is linearly independent. For if $\{v\}$ is linearly dependent, then $av = 0$ for some nonzero scalar a. thus

$$v = a^{-1}(av) = a^{-1}0 = 0.$$

3. A set is linearly independent if and only if the only representations of 0 as linear combinations of its vectors are trivial representations.

Theorem 1.6.
Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then S_2 is linearly dependent.

Corollary 1.3.
Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_2 is linearly independent, then S_1 is linearly independent.

Theorem 1.7.
Let S be a linearly independent subset of a vector space V, and let v be a vector in V that is not in S. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \text{span}(S)$.

6 Bases and Dimension

Definition 1.30.
A **basis** β for a vector space V is a linearly independent subset of V that generates V. If β is a basis for V, we also say that the vectors of β form a basis for V.

Theorem 1.8.
Let V be a vector space and $\beta = \{v_1, v_2, \ldots, v_n\}$ be a subset of V. Then β is a basis for V if and only if each $v \in V$ can be uniquely expressed as a linear combination of vectors of β, that is, can be expressed in the form

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

for unique scalars $a_1, a_2, \ldots, a_n$.

Theorem 1.9.
If a vector space V is generated by a finite set S, then some subset of S is a basis for V. Hence V has a finite basis.

Theorem 1.10 (Replacement Theorem).
Let V be a vector space that is generated by a set G containing exactly n vectors, and let L be a linearly independent subset of V containing exactly m vectors. Then $m \leq n$ and there exists a subset H of G containing exactly $n - m$ vectors such that $L \cup H$ generates V.

Corollary 1.4. Let V be a vector space having a finite basis. Then every basis for V contains the same number of vectors.

Definition 1.31.
A vector space is called **finite-dimensional** if it has a basis consisting of a finite number of vectors. The unique number of vectors in each basis for V is called the **dimension** of V and is denoted by $\dim(V)$. A vector space that is not finite-dimensional is called **infinite-dimensional**.

Corollary 1.5.
Let V be a vector space with dimension n.

1. Any finite generating set for V contains at least n vectors, and a generating set for V that contains exactly n vectors is a basis for V.

2. Any linearly independent subset of V that contains exactly n vectors is a basis for V.

3. Every linearly independent subset of V can be extended to a basis for V.

The Dimension of Subspaces

Theorem 1.11.
Let W be a subspace of a finite-dimensional vector space V. Then W is finite-dimensional and $\dim(W) \leq \dim(V)$. Moreover, if $\dim(W) = \dim(V)$, then $V = W$.

Corollary 1.6.
If W is a subspace of a finite-dimensional vector space V, then any basis for W can be extended to a basis for V.

The Lagrange Interpolation Formula

Definition 1.32 (The Lagrange Interpolation Formula).
Corollary 1.5 of the replacement theorem can be applied to obtain a useful formula. Let $c_0, c_1, \ldots, c_n$ be distinct scalars in an infinite field $\mathbb{F}$. The polynomials $f_0(x), f_1(x), \ldots, f_n(x)$ defined by

$$f_i(x) = \frac{(x - c_0) \ldots (x - c_{i-1})(x - c_{i+1}) \ldots (x - c_n)}{(c_i - c_0) \ldots (c_i - c_{i-1})(c_i - c_{i+1}) \ldots (c_i - c_n)} = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - c_k}{c_i - c_k}$$

are called the **Lagrange polynomials** (associated with $c_0, c_1, \ldots, c_n$). Note that each $f_i(x)$ is a polynomial of degree n and hence is in $P_n(\mathbb{F})$. By regarding $f_i(x)$ as a polynomial function $f_i : \mathbb{F} \to \mathbb{F}$, we see that

$$f_i(c_j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases} \tag{1.1}$$

This property of Lagrange polynomials can be used to show that $\beta = \{f_0, f_1, \ldots, f_n\}$ is a linearly independent subset of $P_n(\mathbb{F})$. Suppose that

$$\sum_{i=0}^{n} a_i f_i = 0 \quad \text{for some scalars } a_0, a_1, \ldots, a_n,$$

where 0 denotes the zero function. Then

$$\sum_{i=0}^{n} a_i f_i(c_j) = 0 \quad \text{for } j = 0, 1, \ldots, n.$$

But also

$$\sum_{i=0}^{n} a_i f_i(c_j) = a_j$$

by (1.1). Hence $a_j = 0$ for $j = 0, 1, \ldots, n$; so β is linearly independent. Since the dimension of $P_n(\mathbb{F})$ is $n + 1$, it follows from Corollary 2 of the replacement theorem that β is a basis for $P_n(\mathbb{F})$.

Because β is a basis for $P_n(\mathbb{F})$, every polynomial function g in $P_n(\mathbb{F})$ is a linear combination of polynomial functions of β, say,

$$g = \sum_{i=0}^{n} b_i f_i.$$

It follows that

$$g(c_j) = \sum_{i=0}^{n} b_i f_i(c_j) = b_j;$$

so

$$g = \sum_{i=0}^{n} g(c_i) f_i$$

is the unique representation of g as a linear combination of elements of β. This representation is called the **Lagrange interpolation formula**. Notice that the preceding argument shows that if $b_0, b_1, \ldots, b_n$ are any $n+1$ scalars in $\mathbb{F}$ (not necessarily distinct), then the polynomial function

$$g = \sum_{i=0}^{n} b_i f_i$$

is the unique polynomial in $P_n(\mathbb{F})$ such that $g(c_j) = b_j$. Thus we have found the unique polynomial of degree not exceeding n that has specified values b_j at given points c_j in its domain $(j = 0, 1, \ldots, n)$.

An important consequence of the Lagrange interpolation formula is the following result: If $f \in P_n(\mathbb{F})$ and $f(c_i) = 0$, for $n+1$ distinct scalars $c_0, c_1, \ldots, c_n$ in $\mathbb{F}$, then f is the zero function.

7   Maximal Linearly Independent Subsets

Definition 1.33.
Let $\mathcal{F}$ be a family of sets. A member M of $\mathcal{F}$ is called **maximal** (with respect to set inclusion) if M is contained in no member of $\mathcal{F}$ other than M itself.

Definition 1.34.
Let $\mathcal{F}$ be the family of all subsets of a nonempty set S. This family $\mathcal{F}$ is called the **power set** of S.

Definition 1.35.
A collection of sets $\mathcal{C}$ is called a **chain** (or **nest** or **tower**) if for each pair of sets A and B in $\mathcal{C}$, either $A \subseteq B$ or $B \subseteq A$.

Definition 1.36 (Maximal Principle).
Let $\mathcal{F}$ be a family of sets. If, for each chain $\mathcal{C} \subseteq \mathcal{F}$, there exists a member of $\mathcal{F}$ that contains each member of $\mathcal{C}$, then $\mathcal{F}$ contains a maximal member.

Note: The *Maximal Principle* is logically equivalent to the *Axiom of Choice*, which is an assumption in most axiomatic developments of set theory.

Definition 1.37.
Let S be a subset of a vector space V. A **maximal linearly independent subset** of S is a subset B of S satisfying both of the following conditions

1. B is linearly independent.

2. The only linearly independent subset of S that contains B is B itself.

Corollary 1.7.
Every vector space has a basis.

Chapter 2

Linear Transformations and Matrices

1   Linear Transformations, Null Spaces, and Ranges

Definition 2.1.
Let V and W be vector spaces (over $\mathbb{F}$). We call a function $T : V \to W$ a **linear transformation from V to W** if, for all $x, y \in V$, and $c \in \mathbb{F}$, we have

1. $T(x + y) = T(x) + T(y)$, and

2. $T(cx) = cT(x)$

If the underlying field $\mathbb{F}$ is the field of rational numbers, then (1) implies (2), but, in general (1) and (2) are logically independent.

We often simply call T **linear**.

Remark 2.1.
Let V and W be vector spaces (over $\mathbb{F}$). Let $T : V \to W$ be a linear transformation. Then the following properties hold:

1. If T is linear, then $T(0) = 0$.

2. T is linear if and only if $T(cx + y) = cT(x) + T(y)$ for all $x, y \in V$ and $c \in \mathbb{F}$.

3. If T is linear, then $T(x - y) = T(x) - T(y)$ for all $x, y \in V$.

4. T is linear if and only if, for $x_1, x_2, \ldots, x_n \in V$ and $a_1, a_2, \ldots, a_n \in \mathbb{F}$, we have

$$T\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i T(x_i).$$

We generally use property 2 to prove that a given transformation is linear.

Definition 2.2.
For any angle θ, define $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by the rule: $T_\theta(a_1, a_2)$ is the vector obtained by rotating (a_1, a_2) counterclockwise by θ if $(a_1, a_2) \neq (0, 0)$, and $T_\theta(0, 0) = (0, 0)$. Then $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation that is called the **rotation by** θ.

Definition 2.3.
Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(a_1, a_2) = (a_1, -a_2)$. T is called the **reflection about the x-axis**.

Definition 2.4.
For vector spaces V and W (over $\mathbb{F}$), we define the **identity transformation** $I_V : V \to V$ by $I_V(x) = x$ for all $x \in V$.

We define the **zero transformation** $T_0 : V \to W$ by $T_0(x) = 0$ for all $x \in V$.

Note: We often write I instead of I_V.

Definition 2.5.
Let V and W be vector spaces, and let $T : V \to W$ be linear. We define the **null space** (or **kernel**) $\mathrm{N}(T)$ to be the set of all vectors $x \in V$ such that $T(x) = 0$; that is,
$\mathrm{N}(T) = \{x \in V : T(x) = 0\}$.
We define the **range** (or **image**) $\mathrm{R}(T)$ of T to be the subset of W consisting of all images (under T) of vectors in V; that is, $\mathrm{R}(T) = \{T(x) : x \in V\}$.

Theorem 2.1.
Let V and W be vector spaces and $T : V \to W$ be linear. Then $\mathrm{N}(T)$ and $\mathrm{R}(T)$ are subspaces of V and W, respectively.

Theorem 2.2.
Let V and W be vector spaces, and let $T : V \to W$ be linear. If $\beta = \{v_1, v_2, \ldots, v_n\}$ is a basis for V, then

$$\mathrm{R}(T) = \mathrm{span}\,(T(\beta)) = \mathrm{span}\,(\{T(v_1), T(v_2), \ldots, T(v_n)\})\,.$$

Definition 2.6.
Let V and W be vector spaces, and let $T : V \to W$ be linear. If $\mathrm{N}(T)$ and $\mathrm{R}(T)$ are finite-dimensional, then we define the **nullity** of T, denoted $\mathrm{nullity}\,(T)$, and the **rank** of T, denoted $\mathrm{rank}\,(T)$, to be the dimensions of $\mathrm{N}(T)$ and $\mathrm{R}(T)$, respectively.

Theorem 2.3 (Dimension Theorem).
Let V and W be vector spaces, and let $T : V \to W$ be linear. If V is finite-dimensional, then

$$\mathrm{nullity}\,(T) + \mathrm{rank}\,(T) = \dim(V)$$

Theorem 2.4.
Let V and W be vector spaces, and let $T : V \to W$ be linear. Then T is one-to-one if and only if $\mathrm{N}(T) = \{0\}$.

Theorem 2.5.
Let V and W be vector spaces of equal (finite) dimension, and let $T : V \to W$ be linear. Then the following are equivalent.

1. T is one-to-one.

2. T is onto.

3. $\operatorname{rank}(T) = \dim(V)$.

Theorem 2.6.
Let V and W be vector spaces over $\mathbb{F}$, and suppose that $\{v_1, v_2, \ldots, v_n\}$ is a basis for V. For $w_1, w_2, \ldots, w_n$ in W, there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for $i = 1, 2, \ldots, n$.

Corollary 2.1.
Let V and W be vector spaces, and suppose that V has a finite basis $\{v_1, v_2, \ldots, v_n\}$. If $U, T : V \to W$ are linear and $U(v_i) = T(v_i)$, for $i = 1, 2, \ldots, n$, then $U = T$.

Definition 2.7.
Let V be a vector space and W_1 and W_2 be subspaces of V such that $V = W_1 \oplus W_2$. A function $T : V \to V$ is called the **projection on W_1 along W_2** if, for $x = x_1 + x_2$ with $x_1 \in W$ and $x_2 \in W_2$, we have $T(x) = x_1$.

Definition 2.8.
Let V be a vector space, and let $T : V \to W$ be linear. A subspace W of V is said to be **T-invariant** if $T(x) \in W$ for every $x \in W$, that is, $T(W) \subseteq W$. If W is T-invariant, we define the **restriction of T on W** to be the function $T_W : W \to W$ defined by $T_W(x) = T(x)$ for all $x \in W$.

2 The Matrix Representation of a Linear Transformation

Definition 2.9.
Let V be a finite-dimensional vector space. An **ordered basis** for V is a basis for V endowed with a specific order; that is, an ordered basis for V is a finite sequence of linearly independent vectors in V that generates V.

For the vector space $\mathbb{F}^n$, we call $\{e_1, e_2, \ldots, e_n\}$ the **standard ordered basis** for $\mathbb{F}^n$. Similarly, for the vector space $P_n(\mathbb{F})$, we call $\{1, x, \ldots, x^n\}$ the **standard ordered basis** for $P_n(\mathbb{F})$.

Definition 2.10.
Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for a finite-dimensional vector space V. For $x \in V$, let $a_1, a_2, \ldots, a_n$ be the unique scalar values such that

$$x = \sum_{i=1}^{n} a_i v_i.$$

We define the **coordinate vector of x relative to** β, denoted by $[x]_\beta$, by

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Notice that $[v_i]_\beta = e_i$ in the preceding definition. It can be shown that the correspondence $x \to [x]_\beta$ provides us with a linear transformation from V to $\mathbb{F}^n$.

Notation 2.1.

The following notation is used to construct a matrix representation of a linear transformation in the following definition.

Suppose that V and W are finite-dimensional vector spaces with ordered bases $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_m\}$, respectively. Let $T : V \to W$ be linear. Then for each j, $1 \le j \le n$, there exist unique scalars $a_{ij} \in \mathbb{F}$, $1 \le i \le m$, such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{for } 1 \le j \le n.$$

Definition 2.11.

Using the notation above, we call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$ the **matrix representation of T in the ordered bases β and γ.** and write $A = [T]_\beta^\gamma$. If $V = W$ and $\beta = \gamma$, then we write $A = [T]_\beta$.

Notice that the jth column of A is simply $[T(v_j)]_\gamma$. Also observe that if $U : V \to W$ is a linear transformation such that $[U]_\beta^\gamma = [T]_\beta^\gamma$, then $U = T$ by the corollary to Theorem 2.6 (Corollary 2.1).

Definition 2.12.

Let $T, U : V \to W$ be arbitrary functions, where V and W are vector spaces over $\mathbb{F}$, and let $a \in \mathbb{F}$. We define $T + U : V \to W$ by $(T + U)(x) = T(x) + U(x)$ for all $x \in V$, and $aT : V \to W$ by $(aT)(x) = aT(x)$ for all $x \in V$.

Theorem 2.7.

Let V and W be vector spaces over a field $\mathbb{F}$, and let $T, U : V \to W$ be linear.

1. For all $a \in \mathbb{F}$, $aT + U$ is linear.

2. Using the operations of addition and scalar multiplication in the preceding definition, the collection of all linear transformations from V to W is a vector space over $\mathbb{F}$.

Definition 2.13.

Let V and W be vector spaces over $\mathbb{F}$. We denote the vector space of all linear transformations from V to W by $\mathcal{L}(V, W)$. In the case that $V = W$, we write $\mathcal{L}(V)$ instead of $\mathcal{L}(V, W)$.

Theorem 2.8.

Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively, and let $T, U : V \to W$ be linear transformations. Then

1. $[T + U]^\gamma_\beta = [T]^\gamma_\beta + [U]^\gamma_\beta$ and

2. $[aT]^\gamma_\beta = a[T]^\gamma_\beta$ for all scalars a.

3 Compositions of Linear Transformations and Matrix Multiplication

Theorem 2.9.
Let V, W, and Z be vector spaces over the same field $\mathbb{F}$, and let $T : V \to U$ and $U : W \to Z$ be linear. Then $UT : V \to Z$ is linear.

Theorem 2.10.
Let V be a vector space. Let $T, U_1, U_2 \in \mathcal{L}(V)$. Then

1. $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1T + U_2T$

2. $T(U_1 U_2) = (TU_1)U_2$

3. $TI = IT = T$

4. $a(U_1 U_2) = (aU_1)U_2 = U_1(aU_2)$ for all scalars a.

Definition 2.14.
Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We define the **product** of A and B, denoted AB, to be the $m \times p$ matrix such that

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \quad \text{for} \ \ 1 \le i \le m, \ \ 1 \le j \le p.$$

Notice that $(AB)_{ij}$ is the sum of products of corresponding entries from the ith row of A and the jth column of B.

The reader should observe that in order for the product AB to be defined, there are restrictions regarding the relative sizes of A and B. The following mnemonic device is helpful: "$(m \times n) \cdot (n \times p) = (m \times p)$"; that is, in order for the product AB to be defined, the two "inner" dimensions must be equal, and the two "outer" dimensions yield the size of the product.

Theorem 2.11.
Let V, W, and Z be finite-dimensional vector spaces with ordered bases α, β, and γ, respectively. Let $T : V \to W$ and $U : W \to Z$ be linear transformations. Then

$$[UT]^\gamma_\alpha = [U]^\gamma_\beta [T]^\beta_\alpha$$

Corollary 2.2.
Let V be a finite-dimensional vector space with an ordered basis β. Let $T, U \in \mathcal{L}(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.

Definition 2.15.
We define the **Kronecker delta** δ_{ij} by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. The $n \times n$ **identity matrix** I_n is defined by $(I_n)_{ij} = \delta_{ij}$.

Theorem 2.12.
Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$ matrices. Then

1. $A(B + C) = AB + AC$ and $(D + E)A = DA + EA$.

2. $a(AB) = (aA)B = A(aB)$ for any scalar a.

3. $I_m A = A = A I_n$.

4. If V is an n-dimensional vector space with an ordered basis β, then $[I_V]_\beta = I_n$.

Corollary 2.3.
Let A be an $m \times n$ matrix, $B_1, B_2, \ldots, B_k$ be $n \times p$ matrices, $C_1, C_2, \ldots, C_k$ be $q \times m$ matrices, and $a_1, a_2, \ldots, a_k$ be scalars. Then

$$A\left(\sum_{i=1}^{k} a_i B_i\right) = \sum_{i=1}^{k} a_i A B_i$$

and

$$\left(\sum_{i=1}^{k} a_i C_i\right) A = \sum_{i=1}^{k} a_i C_i A.$$

Theorem 2.13.
Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. For each j ($1 \leq j \leq p$) let u_j and v_j denote the jth columns of AB and B, respectively. Then

1. $u_j = A v_j$.

2. $v_j = B e_j$, where e_j is the jth standard vector of $\mathbb{F}^p$.

Theorem 2.14.
Let V and W be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let $T : V \to W$ be linear. Then, for each $u \in V$, we have

$$[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta.$$

Definition 2.16.
Let A be an $m \times n$ matrix with entries from a field $\mathbb{F}$. We denote $L_A : \mathbb{F}^n \to \mathbb{F}^m$ defined by $L_A(x) = Ax$ (the matrix product of A and x) for each column vector $x \in \mathbb{F}^n$. We call L_A a **left-multiplication transformation**.

Theorem 2.15.
Let A be an $m \times n$ matrix with entries from $\mathbb{F}$. Then the left-multiplication transformation $L_A : \mathbb{F}^n \to \mathbb{F}^m$ is linear. Furthermore, if B is any other $m \times n$ matrix (with entries from $\mathbb{F}$) and β and γ are the standard ordered bases for $\mathbb{F}^n$ and $\mathbb{F}^m$, respectively, then we have the following properties.

1. $[L_A]_\beta^\gamma = A$.

2. $L_A = L_B$ if and only if $A = B$.

3. $L_{A+B} = L_A + L_B$ and $L_{aA} = aL_A$ for all $a \in \mathbb{F}$.

4. If $T : \mathbb{F}^n \to \mathbb{F}^m$ is linear, then there exists a unique $m \times n$ matrix C such that $T = L_C$. In fact, $C = [T]_\beta^\gamma$.

5. If E is an $n \times p$ matrix, then $L_{AE} = L_A L_E$.

6. If $m = n$, then $L_{I_n} = I_{\mathbb{F}^n}$.

Theorem 2.16.
Let A, B, and C be matrices such that $A(BC)$ is defined. Then $(AB)C$ is also defined and $A(BC) = (AB)C$; that is, matrix multiplication is associative.

Applications

Definition 2.17.
An **incidence matrix** is a square matrix in which all the entries are either zero or one and, for convenience, all the diagonal entries are zero. If we have a relationship on a set of n objects that we denote $1, 2, \ldots, n$, then we define the associated incidence matrix A by $A_{ij} = 1$ if i is related to j, and $A_{ij} = 0$ otherwise.

Definition 2.18.
A relationship among a group of people is called a **dominance relation** if the associated incidence matrix A has the property that for all distinct pairs i and j, $A_{ij} = 1$ if and only if $A_{ji} = 0$, that is, given any two people, exactly one of them *dominates* the other.

4 Invertibility and Isomorphisms

Definition 2.19.
Let V and W be vector spaces, and let $T : V \to W$ be linear. A function $U : W \to V$ is said to be an **inverse** of T if $TU = I_W$ and $UT = I_V$. If T has an inverse, then T is said to be **invertible**. If T is invertible, then the inverse of T is unique and is denoted by T^{-1}.

The following facts hold for invertible functions T and U.

1. $(TU)^{-1} = U^{-1}T^{-1}$.

2. $(T^{-1})^{-1} = T$; in particular, T^{-1} is invertible.

We often use the fact that a function is invertible if and only if it is one-to-one and onto. We can therefore restate Theorem 2.5 as follows:

3. Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional vector spaces of equal dimension. then T is invertible if and only if rank $(T) = \dim(T)$.

Theorem 2.17.
Let V and W be vector spaces, and let $T : V \to W$ be linear and invertible. Then $T^{-1} : W \to V$ is linear.

Definition 2.20.
Let A be an $n \times n$ matrix. Then A is **invertible** if there exists an $n \times n$ matrix B such that $AB = BA = I$.

If A is invertible, then the matrix B such that $AB = BA = I$ is unique. (If C were another such matrix, then $C = CI = C(AB) = (CA)B = IB = B$.) The matrix B is called the **inverse** of A and is denoted by A^{-1}.

Lemma 2.1.
Let T be an invertible linear transformation from V to W. Then V is finite-dimensional if and only if W is finite-dimensional. In this case, $\dim(V) = \dim(W)$

Theorem 2.18.
Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Let $T : V \to W$ be linear. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible. Furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Corollary 2.4.
Let V be a finite-dimensional vector space with an ordered bases β, and let $T : V \to V$ be linear. Then T is invertible if and only if $[T]_\beta$ is invertible. Furthermore, $[T^{-1}]_\beta = ([T]_\beta)^{-1}$.

Corollary 2.5.
Let A be and $n \times n$ matrix. Then A is invertible if and only if L_A is invertible. Furthermore, $(L_A)^{-1} = L_{A^{-1}}$.

Definition 2.21.
Let V and W be vector spaces. We say that V is **isomorphic** to W if there exists a linear transformation $T : V \to W$ that is invertible. Such a linear transformation is called an **isomorphism** from V onto W.

Theorem 2.19.
Let V and W be finite-dimensional vector spaces (over the same field). Then V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

Corollary 2.6.
Let V be a vector space over $\mathbb{F}$. Then V is isomorphic to $\mathbb{F}^n$ if and only if $\dim(V) = n$.

Theorem 2.20.
Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ of dimensions n and m, respectively, and let β and γ be ordered bases for V and W, respectively. Then the function $\Phi : \mathcal{L}(V, W) \to M_{m \times n}(\mathbb{F})$, defined by $\Phi(T) = [T]_\beta^\gamma$ for $T \in \mathcal{L}(V, W)$ is an isomorphism.

Corollary 2.7.
Let V and W be finite-dimensional vector spaces of dimension n and m, respectively. Then $\mathcal{L}(V, W)$ is finite-dimensional of dimension mn.

Definition 2.22.
Let β be an ordered basis for an n-dimensional vector space V over the field $\mathbb{F}$. The **standard representation of V with respect to** β is the function $\phi_\beta : V \to \mathbb{F}^n$ defined by $\phi_\beta(x) = [x]_\beta$ for each $x \in V$.

Theorem 2.21.
For any finite-dimensional vector space V with ordered basis β, ϕ_β is an isomorphism.

5   The Change of Coordinate Matrix

Theorem 2.22.
Let β and β' be two ordered bases for a finite-dimensional vector pace V, and let $Q = [I_V]_{\beta'}^\beta$. Then

1. Q is invertible.

2. For any $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

Definition 2.23.
The matrix $Q = [I_V]_{\beta'}^\beta$, defined in Theorem 2.22, is called a **change of coordinate matrix**. Because of part (2) of the theorem, we say that Q **changes β'-coordinates into β-coordinates**.

Definition 2.24.
A linear transformation that maps a vector space V into itself is called a **linear operator on V**.

Theorem 2.23.
Let T be a linear operator on a finite-dimensional vector space V, and let β and β' be ordered bases for V. Suppose that Q is the change of coordinate matrix that changes β'-coordinates into β-coordinates. Then

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q$$

Corollary 2.8.
Let $A \in M_{n \times n}(\mathbb{F})$, and let γ be an ordered basis for $\mathbb{F}^n$. Then $[L_A]_\gamma = Q^{-1}AQ$, where Q is the $n \times n$ matrix whose jth column is the jth vector of γ.

Definition 2.25.
Let A and B be matrices in $M_{n \times n}(\mathbb{F})$. We say that B is **similar** to A if there exists an invertible matrix Q such that $B = Q^{-1}AQ$.

Notice that the relation of similarity is an equivalence relation. So we need only say that A and B are similar.

6   Dual Spaces

Definition 2.26.
A linear transformation from a vector space V into its field of scalars $\mathbb{F}$, which is itself a vector space of dimension 1 over $\mathbb{F}$, is called a **linear functional** on V. We generally use the letters $\mathsf{f}, \mathsf{g}, \mathsf{h}, \dots$ to denote linear functionals.

Definition 2.27.
Let V be a vector space of continuous real-valued functions on the interval $[0, 2\pi]$. Fix a function $g \in V$. The function $\mathsf{h} : V \to \mathbb{R}$, defined by

$$\mathsf{h}(x) = \frac{1}{2\pi} \int_0^{2\pi} x(t)g(t)dt$$

is a linear functional on V. In the cases that $g(t)$ equals $\sin(nt)$ or $\cos(nt)$, $\mathsf{h}(x)$ is often called the ***n*th Fourier coefficient of** x.

Definition 2.28.
Let V be a finite dimensional vector space, and let $\beta = \{x_1, x_2, \dots, x_n\}$ be an ordered basis for V. For each $i = 1, 2, \dots, n$, define $\mathsf{f}_i(x) = a_i$, where

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

is the coordinate vector of x relative to β. Then f is a linear function on V called the ***i*th coordinate function with respect to the basis** β. Note that $\mathsf{f}_i(x_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta. These linear functionals play an important role in the theory of dual spaces (see Theorem 2.24).

Definition 2.29.
For a vector space V over $\mathbb{F}$, we define the **dual space** of V to be the vector space $\mathcal{L}(V, \mathbb{F})$, denoted by V^*.

Thus V^* is the vector space consisting of all linear functionals on V with the operations of addition and scalar multiplication. Note that if V is finite-dimensional, then by Corollary 2.7

$$\dim(V^*) = \dim(\mathcal{L}(V, \mathbb{F})) = \dim(V) \cdot \dim(\mathbb{F}) = \dim(V).$$

Hence by Theorem 2.19, V and V^* are isomorphic. We also define the **double dual** V^{**} of V to be the dual of V^*. In Theorem 2.26, we show, in fact, that there is a natural identification of V and V^{**} in the case that V is finite-dimensional.

Theorem 2.24.
Suppose that V is a finite-dimensional vector space with the ordered basis $\beta = \{x_1, x_2, \ldots, x_n\}$. Let f_i $(1 \le i \le n)$ be the ith coordinate function with respect to β as just defined, and let $\beta^* = \{f_1, f_2, \ldots, f_n\}$. Then β^* is an ordered basis for V^*, and, for any $f \in V^*$, we have

$$f = \sum_{i=1}^{n} f(x_i) f_i.$$

Definition 2.30.
Using the notation of Theorem 2.24, we call the ordered basis $\beta^* = \{f_1, f_2, \ldots, f_n\}$ of V^* that satisfies $f_i(x_j) = \delta_{ij}$ $(1 \le i,\ j \le n)$ the **dual basis** of β.

Theorem 2.25.
Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ with ordered bases β and γ, respectively. For any linear transformation $T : V \to W$, the mapping $T^t : W^* \to V^*$ defined by $T^t(g) = gT$ for all $g \in W^*$ is a linear transformation with the property that $[T^t]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.

Definition 2.31.
The linear transformation T^t defined in Theorem 2.25 is called the **transpose** of T. It is clear that T^t is the unique linear transformation U such that $[U]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.

Definition 2.32.
For a vector x in a finite-dimensional vector space V, we define the linear functional $\hat{x} : V^* \to \mathbb{F}$ on V^* by $\hat{x}(f) = f(x)$ for every $f \in V^*$. Since $\hat{x}$ is a linear functional on V^*, $\hat{x} \in V^{**}$.

The correspondence $x \leftrightarrow \hat{x}$ allows us to define the desired isomorphism between V^* and V^{**}.

Lemma 2.2.
Let V be a finite-dimensional vector space, and let $x \in V$. If $\hat{x}(f) = 0$ for all $f \in V^*$, then $x = 0$.

Theorem 2.26.
Let V be a finite-dimensional vector space, and define $\psi : V \to V^{**}$ by $\psi(x) = \hat{x}$. Then ψ is an isomorphism.

Corollary 2.9.
Let V be a finite-dimensional vector space with dual space V^*. Then every ordered basis for V^* is the dual basis for some basis V.

Definition 2.33.
Let V be a finite-dimensional vector space over $\mathbb{F}$. For every subset S of V, define the **annihilator** S^0 of S as

$$S^0 = \{f \in V^* : f(x) = 0,\ \text{for all } x \in S\}$$

7 Homogeneous Linear Differential Equations with Constant Coefficients

Definition 2.34.
A **differential equation** in an unknown function $y = y(t)$ is an equation involving y, t, and derivatives of y. If the differential equation is of the form

$$a_n y^{(n)} + a_{n-1} y^{(n-)} + \cdots + a_1 y^{(1)} + a_0 y = f, \tag{2.1}$$

where $a_0, a_1, \ldots, a_n$ and f are functions of t and $y^{(k)}$ denotes the kth derivative of y, then the equation is said to be **linear**. The functions a_i are called the **coefficients** of the differential equation. When f is identically zero, (2.1) is called **homogeneous**.

If $a_n \neq 0$, we say that differential equation (2.1) is of **order n**. In this case, we divide both sides by a_n to obtain a new, but equivalent, equation

$$y^{(n)} + b_{n-1} y^{(n-1)} + \cdots + b_1 y^{(1)} + b_0 y = 0,$$

where $b_i = a_i/a_n$ for $i = 0, 1, \ldots, n-1$. Because of this observation, we always assume that the coefficient a_n in (2.1) is 1.

A **solution** to (2.1) is a function that when substituted for y reduces (2.1) to an identity.

Definition 2.35.
Given a complex-valued function $x \in \mathcal{F}(\mathbb{R}, \mathbb{C})$ of a real variable t (where $\mathcal{F}(\mathbb{R}, \mathbb{C})$ is the vector space defined in Definition 1.8), there exist unique real-valued functions x_1 and x_2 of t, such that

$$x(t) = x_1(t) + ix_2(t) \quad \text{for} \quad t \in \mathbb{R},$$

where i is the imaginary number such that $i^2 = -1$. We call x_1 the **real part** and x_2 the **imaginary part** of x.

Definition 2.36.
Given a function $x \in \mathcal{F}(\mathbb{R}, \mathbb{C})$ with real part x_1 and imaginary part x_2, we say that x is **differentiable** if x_1 and x_2 are differentiable. If x is differentiable, we define the **derivative** x' of x by

$$x' = x_1' + ix_2'$$

Theorem 2.27.
Any solution to a homogeneous linear differential equation with constant coefficients has derivatives of all orders; that is, if x is a solution to such an equation, then $x^{(}k)$ exists for every positive integer k.

Definition 2.37.
We use $\mathbb{C}^\infty$ to denote the set of all functions in $\mathcal{F}(\mathbb{R}, \mathbb{C})$ that have derivatives of all orders.

Definition 2.38.
For any polynomial $p(t)$ over $\mathbb{C}$ of positive degree, $p(\mathsf{D})$ is called a **differential operator**. The **order** of the differential operator $p(\mathsf{D})$ is the degree of the polynomial $p(t)$.

Definition 2.39.
Given the differential equation

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y^{(1)} + a_0 y = 0,$$

the complex polynomial

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

is called the **auxiliary polynomial** associated with the equation.

Theorem 2.28.
The set of all solutions to a homogeneous linear differential equation with constant coefficients coincides with the null space of $p(\mathsf{D})$ where $p(t)$ is the auxiliary polynomial associated with the equation.

Corollary 2.10.
The set of all solutions to a homogeneous linear differential equation with constant coefficients is a subspace of $\mathbb{C}^\infty$.

Definition 2.40.
We call the set of solutions to a homogeneous linear differential equation with constant coefficients the **solution space** of the equation.

Definition 2.41.
Let $c = a + ib$ be a complex number with real part a and imaginary part b. Define

$$e^c = e^a(\cos(b) + i\sin(b)).$$

The special case

$$e^{ib} = \cos(b) + i\sin(a)$$

is called **Euler's formula**.

Definition 2.42.
A function $f : \mathbb{R} \to \mathbb{C}$ defined by $f(t) = e^{ct}$ for a fixed complex number c is called an **exponential function**.

Theorem 2.29.
For any exponential function $f(t) = e^{ct}$, $f'(t) = ce^{ct}$.

Theorem 2.30.
Recall that the **order** of a homogeneous linear differential equation is the degree of its auxiliary polynomial. Thus, an equation of order 1 is of the form

$$y' + a_0 y = 0. \tag{2.2}$$

The solution space for (2.2) is of dimension 1 and has $\{e^{-a_0 t}\}$ as a basis.

Corollary 2.11.

For any complex number c, the null space of the differential operator $\mathsf{D} - cI$ has $\{e^{ct}\}$ as a basis.

Theorem 2.31.

Let $p(t)$ be the auxiliary polynomial for a homogeneous linear differential equation with constant coefficients. For any complex number c, if c is a zero of $p(t)$, then e^{ct} is a solution to the differential equation.

Theorem 2.32.

For any differential operator $p(\mathsf{D})$ of order n, the null space of $p(\mathsf{D})$ is an n-dimensional subspace of $\mathbb{C}^\infty$.

Lemma 2.3.

The differential operator $\mathsf{D} - cI : \mathbb{C}^\infty \to \mathbb{C}^\infty$ is onto for any complex number c.

Lemma 2.4.

Let V be a vector space, and suppose that T and U are linear operators on V such that U is onto and the null spaces of T and U are finite-dimensional. Then the null space of TU is finite-dimensional, and

$$\dim\left(\mathrm{N}\left(TU\right)\right) = \dim\left(\mathrm{N}\left(T\right)\right) + \dim\left(\mathrm{N}\left(U\right)\right)$$

Corollary 2.12.

The solution space of any nth-order homogeneous linear differential equation with constant coefficients is an n-dimensional subspace of $\mathbb{C}^\infty$.

Theorem 2.33.

Given n distinct complex numbers $c_1, c_2, \ldots, c_n$, the set of exponential functions $\{e^{c_1 t}, e^{c_2 t}, \ldots, e^{c_n t}\}$ is linearly independent.

Corollary 2.13.

For any nth-order homogeneous linear differential equation with constant coefficients, if the auxiliary polynomial has n distinct zeros $c_1, c_2, \ldots, c_n$, then $\{e^{c_1 t}, e^{c_2 t}, \ldots, e^{c_n t}\}$ is a basis for the solution space of the differential equation.

Lemma 2.5.

For a given complex number c and a positive integer n, suppose that $(t - c)^n$ is the auxiliary polynomial of a homogeneous linear differential equation with constant coefficients. Then the set

$$\beta = \{e^{ct}, te^{ct}, \ldots, t^{n-1}e^{ct}\}$$

is a basis for the solution space of the equation.

Theorem 2.34.

Given a homogeneous linear differential equation with constant coefficients and auxiliary polynomial

$$(t - c_1)^{n_1}(t - c_2)^{n_2} \ldots (t - c_k)^{n_k},$$

where $n_1, n_2, \ldots, n_k$ are positive integers and $c_1, c_2, \ldots, c_k$ are distinct complex numbers, the following set is a basis for the solution space of the equation:

$$\{e^{c_1 t}, t e^{c_1 t}, \ldots, t^{n_1 - 1} e^{c_1 t}, \ldots, e^{c_k t}, t e^{c_k t}, \ldots, t^{n_k - 1} e^{c_k t}\}$$

Definition 2.43.
A differential equation

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

is called a **nonhomogeneous** linear differential equation with constant coefficients if the a_i's are constant and x is a function that is not identically zero.

Chapter 3

Elementary Matrix Operations and Systems of Linear Equations

1   Elementary Matrix Operations and Elementary Matrices

Definition 3.1.
Let A be an $m \times n$ matrix. Any one of the following three operations on the rows [columns] of A is called an **elementary row [column] operation**:

    1. interchanging any two rows [columns] of A;

    2. multiplying any row [column] of A by a nonzero scalar;

    3. adding any scalar multiple of a row [column] of A to another row [column].

Any of these three operations are called an **elementary operation**. Elementary operations are of **type 1**, **type 2**, or **type 3** depending on whether they are obtained by (1), (2), or (3).

Definition 3.2.
An $n \times n$ **elementary matrix** is a matrix obtained by performing an elementary operation on I_n. The elementary matrix is said to be of **type 1**, **2**, or **3** according to whether the elementary operation performed on I_n is a type 1, 2, or 3 operation, respectively.

Theorem 3.1.
Let $A \in M_{m \times n}(\mathbb{F})$, and suppose that B is obtained from A by performing an elementary row [column] operation. Then there exists an $m \times m$ $[n \times n]$ elementary matrix E such that $B = EA$ $[B = AE]$. In fact, E is obtained from I_m $[I_n]$ by performing the same elementary row [column] operation as that which was performed on A to obtain B. Conversely, if E is an elementary $m \times m$ $[n \times n]$ matrix, then EA $[AE]$ is the matrix obtained from A by performing the same elementary row [column] operation as that which produces E from I_m $[I_n]$.

Theorem 3.2.
Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix of the same type.

2   The Rank of a Matrix and Matrix Inverses

Definition 3.3.
If $A \in M_{m \times n}(\mathbb{F})$, we define the **rank** of A, denoted $\operatorname{rank}(A)$, to be the rank of the linear transformation $L_A : \mathbb{F}^n \to \mathbb{F}^m$.

Theorem 3.3.
Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces, and let β and γ be ordered bases for V and W, respectively. Then $\operatorname{rank}(T) = \operatorname{rank}\left([T]_\beta^\gamma\right)$.

Theorem 3.4.
Let A be an $m \times n$ matrix. if P and Q are invertible $m \times m$ and $n \times n$ matrices, respectively, then

1. $\operatorname{rank}(AQ) = \operatorname{rank}(A)$,

2. $\operatorname{rank}(PA) = \operatorname{rank}(A)$,
    and therefore

3. $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$.

Corollary 3.1.
Elementary row and column operations on a matrix are rank preserving.

Theorem 3.5.
The rank of any matrix equals the maximum number of its linearly independent columns; that is, the rank of a matrix is the dimension of the subspace generated by its columns.

Theorem 3.6.
Let A be an $m \times n$ matrix of rank r. Then $r \leq m$, $r \leq n$, and, by means of a finite number of elementary row and column operations, A can be transformed into the matrix

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

where O_1, O_2 and O_3 are the zero matrices. Thus $D_{ii} = 1$ for $i \leq r$ and $D_{ij} = 0$ otherwise.

Corollary 3.2.
Let A be an $m \times n$ matrix of rank r. Then there exist invertible matrices B and C of sizes $m \times m$ and $n \times n$, respectively, such that $D = BAC$, where

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

is the $m \times n$ matrix in which O_1, O_2, and O_3 are zero matrices.

Corollary 3.3.
Let A be an $m \times n$ matrix. Then

1. $\operatorname{rank}(A^t) = \operatorname{rank}(A)$.

2. The rank of any matrix equals the maximum number of its linearly independent rows; that is, the rank of a matrix is the dimension of the subspace generated by its rows.

3. The rows and columns of any matrix generate subspaces of the same dimension, numerically equal to the rank of the matrix.

Corollary 3.4.
Every invertible matrix is a product of elementary matrices.

Theorem 3.7.
Let $T : V \to W$ and $U : W \to Z$ be linear transformations on finite-dimensional vector spaces V, W, and Z, and let A and B be matrices such that the product AB is defined. Then

1. $\operatorname{rank}(UT) \leq \operatorname{rank}(U)$.

2. $\operatorname{rank}(UT) \leq \operatorname{rank}(T)$.

3. $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$.

4. $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$.

The Inverse of a Matrix

Definition 3.4.
Let A and B be $m \times n$ and $m \times p$ matrices, respectively. By the **augmented matrix** $(A|B)$, we mean the $m \times (n \times p)$ matrix $(A\ B)$, that is, the matrix whose first n columns are the columns of A, and whose last p columns are the columns of B.

3   Systems of Linear Equations – Theoretical Aspects

Definition 3.5.
The system of equations

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

where a_{ij} and b_i ($1 \leq i \leq m$ and $1 \leq j \leq n$) are scalars in a field $\mathbb{F}$ and $x_1, x_2, \ldots, x_n$ are n variables taking values in $\mathbb{F}$, is a called a **system of m linear equations in n unknowns over the field $\mathbb{F}$**.

The $m \times n$ matrix

$$
\begin{pmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & \vdots & & \vdots \\
a_{m1} & a_{m2} & \ldots & a_{mn}
\end{pmatrix}
$$

is called the **coefficient matrix** of the system (S).

If we let

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

then the system (S) may be rewritten as a single matrix equation

$$Ax = b.$$

To exploit the results that we have developed, we often consider a system of linear equations as a single matrix equation.

A **solution** to the system (S) is an n-tuple

$$s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \in \mathbb{F}^n$$

such that $As = b$. The set of all solutions to the system (S) is called the **solution set** of the system. System (S) is called **consistent** if its solution set is nonempty; otherwise it is called **inconsistent**.

Definition 3.6.
A system $Ax = b$ of m linear equations in n unknowns is said to be **homogeneous** if $b = 0$. Otherwise the system is said to be **nonhomogeneous**.

Any homogeneous system has at least one solution, namely, the zero vector.

Theorem 3.8.
Let $Ax = 0$ be a homogeneous system of m linear equations in n unknowns over a field $\mathbb{F}$. Let K denote the set of all solutions to $Ax = 0$. Then $K = \mathrm{N}\,(L_A)$; hence K is a subspace of $\mathbb{F}^n$ of dimension $n - \mathrm{rank}\,(L_A) = n - \mathrm{rank}\,(A)$.

Corollary 3.5.
If $m < n$, the system $Ax = 0$ has a nonzero solution.

Definition 3.7.
We refer to the equation $Ax = 0$ as the **homogeneous system corresponding to** $Ax = b$.

Theorem 3.9.
Let K be the solution set of a system of linear equations $Ax = b$, and let $\mathsf{K_H}$ be the solution set of the corresponding homogeneous system $Ax = 0$. Then for any solution s to $Ax = b$

$$K = \{s\} + \mathsf{K_H} = \{s + k : k \in \mathsf{K_H}\}.$$

Theorem 3.10.
Let $Ax = b$ be a system of n linear equations in n unknowns. If A is invertible, then the system has exactly one solution, namely, $A^{-1}b$. Conversely, if the system has exactly one solution, then A is invertible.

Definition 3.8.
The matrix $(A|b)$ is called the **augmented matrix of the system** $Ax = b$.

Theorem 3.11.
Let $Ax = b$ be a system of linear equations. Then the system is consistent if and only if $\operatorname{rank}(A) = \operatorname{rank}(A|b)$.

An Application

Definition 3.9. Consider a system of linear equations

$$a_{11}p_1 + a_{12}p_2 + \cdots + a_{1m}p_m = p_1$$
$$a_{21}p_1 + a_{22}p_2 + \cdots + a_{2m}p_m = p_2$$
$$\cdots$$
$$a_{n1}p_1 + a_{n2}p_2 + \cdots + a_{nm}p_m = p_m$$

This system can be written as $Ap = p$, where

$$p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$$

and A is the coefficient matrix of the system. In this context, A is called the **input-output (or consumption) matrix**, and $Ap = p$ is called the **equilibrium condition**.

For vectors $b = (b_1, b_2, \ldots, b_n)$ and $c = (c_1, c_2, \ldots, c_n)$ in $\mathbb{R}^n$, we use the notation $b \geq c$ $[b > c]$ to mean $b_i \geq c_i$ $[b_i > c_i]$ for all i. The vector b is called **non-negative** [**positive**] if $b \geq 0$ $[b > 0]$.

Theorem 3.12.
Let A be an $n \times n$ input-output matrix having the form

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix},$$

where D is a $1 \times (n-1)$ positive vector and C is an $(n-1) \times 1$ positive vector. Then $(I - A)x = 0$ has a one-dimensional solution set that is generated by a non-negative vector.

4 Systems of Linear Equations – Computational Aspects

Definition 3.10.
Two systems of linear equations are called **equivalent** if they have the same solution set.

Theorem 3.13.
Let $Ax = b$ be a system of m linear equations in n unknowns, and let C be an invertible $m \times n$ matrix. Then the system $(CA)x = Cb$ is equivalent to $Ax = b$.

a

Corollary 3.6.
Let $Ax = b$ be a system of m linear equations in n unknowns. If $(A'|b')$ is obtained from $(A|b)$ by a finite number of elementary row operations, then the system $A'x = b'$ is equivalent to the original system.

Definition 3.11.
A matrix is said to be in **reduced row echelon form** if the following three conditions are satisfied.

1. Any row containing a nonzero entry precedes any row in which all the entries are zero (if any).

2. The first nonzero entry in each row is the only nonzero entry in its column.

3. The first nonzero entry in each row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row.

Definition 3.12.
The following procedure for reducing an augmented matrix to reduced row echelon form is called **Gaussian elimination**. It consists of two separate parts.

1. In the *forward pass*, the augmented matrix is transformed into an upper triangular matrix in which the first nonzero entry of each row is 1, and it occurs in a column to the right of the first nonzero entry in the preceding row.

2. In the *backward pass* or *back-substitution*, the upper triangular matrix is transformed into reduced row echelon form by making the first nonzero entry of each row the only nonzero entry of its column.

Theorem 3.14.
Gaussian elimination transforms any matrix into its reduced row echelon form.

Definition 3.13. A solution to a system of equations of the form

$$s = s_0 + t_1 u_1 + t_2 u_2 + \cdots + t_{n-r} u_{n-r},$$

where r is the number of nonzero solutions in A' $(r \leq m)$, is called a **general solution** of the system $Ax = b$. It expresses an arbitrary solution s of $Ax = b$ in terms of $n - r$ parameters.

Theorem 3.15.

Let $Ax = b$ be a system of r nonzero equations in n unknowns. Suppose that $\text{rank}\,(A) = \text{rank}\,(A|b)$ and that $(A|b)$ is in reduced row echelon form. Then

1. $\text{rank}\,(A) = r$.

2. If the general solution obtained by the procedure above is of the form

$$s = s_0 + t_1 u_1 + t_2 u_2 + \cdots + t_{n-r} u_{n-r},$$

then $\{u_1, u_2, \ldots, u_{n-r}\}$ is a basis for the solution set of the corresponding homogeneous system, and s_0 is a solution to the original system.

An Interpretation of the Reduced Row Echelon Form

Theorem 3.16.

Let A be an $m \times n$ matrix of rank r, where $r > 0$, and let B be the reduced row echelon form of A. Then

1. The number of nonzero rows in B is r.

2. For each $i = 1, 2, \ldots, r$, there is a column b_{j_i} of B such that $b_{j_i} = e_i$.

3. The columns of A numbered $j_1, j_2, \ldots, j_r$ are linearly independent.

4. For each $k = 1, 2, \ldots, n$, if column k of B is $d_1 e_1 + d_2 e_2 + \cdots + d_r e_r$, then column k of A is $d_1 a_{j_1} + d_2 a_{j_2} + \cdots + d_r a_{j_r}$.

Corollary 3.7.

The reduced row echelon form of a matrix is unique.

Chapter 4

Determinants

1   Determinants of Order 2

Definition 4.1.
If

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

is a 2×2 matrix with entries from a field $\mathbb{F}$, then we define the **determinant** of A, denoted $\det(A)$ or $|A|$, to be the scalar $ad - bc$.

Theorem 4.1.
The function $\det : M_{2\times2}(\mathbb{F}) \to \mathbb{F}$ is a linear function of each row of a 2×2 matrix when the other row is held fixed. That is, if u, v and w are in $\mathbb{F}^2$ and k is a scalar, then

$$\det \begin{pmatrix} u + kv \\ w \end{pmatrix} = \det \begin{pmatrix} u \\ w \end{pmatrix} + k \det \begin{pmatrix} v \\ w \end{pmatrix}$$

and

$$\det \begin{pmatrix} w \\ u + kv \end{pmatrix} = \det \begin{pmatrix} w \\ u \end{pmatrix} + k \det \begin{pmatrix} w \\ v \end{pmatrix}.$$

Theorem 4.2.
Let $A \in M_{2\times2}(\mathbb{F})$. Then the determinant of A is nonzero if and only if A is invertible. Moreover, if A is invertible, then

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Definition 4.2.
By the **angle** between two vectors in $\mathbb{R}^2$, we mean the angle with measure θ ($0 \leq \theta < \pi$) that is formed by the vectors having the same magnitude and direction as the given vectors by emanating from the origin.

The Area of a Parallelogram

Definition 4.3.
If $\beta = \{u, v\}$ is an ordered basis for $\mathbb{R}^2$, we define the **orientation** of β to be the real number

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\det \begin{pmatrix} u \\ v \end{pmatrix}}{\left| \det \begin{pmatrix} u \\ v \end{pmatrix} \right|}$$

(The denominator of this fraction is nonzero by Theorem 4.2).

Definition 4.4.
A coordinate system $\{u, v\}$ is called **right-handed** if u can be rotated in a counterclockwise direction through an angle θ $(0 < \theta < \pi)$ to coincide with v. Otherwise, $\{u, v\}$ is called a **left-handed** system.

Definition 4.5.
Any ordered set $\{u, v\}$ in $\mathbb{R}^2$ determines a parallelogram in the following manner. Regarding u and v as arrows emanating from the origin of $\mathbb{R}^2$, we call the parallelogram having u and v as adjacent sides the **parallelogram determined by u and v**.

2 Determinants of Order n

Notation 4.1.
Given $A \in M_{n \times n}(\mathbb{F})$, for $n \geq 2$, denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j by $\tilde{A}_{ij}$. Thus for

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in M_{3 \times 3}(\mathbb{R})$$

we have

$$\tilde{A}_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, \qquad \tilde{A}_{13} = \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}, \qquad \tilde{A}_{32} = \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix}$$

and for

$$B = \begin{pmatrix} 1 & -1 & 2 & -1 \\ -3 & 4 & 1 & -1 \\ 2 & -5 & -3 & 8 \\ -2 & 6 & -4 & 1 \end{pmatrix}$$

we have

$$\tilde{B}_{23} = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -5 & 8 \\ -2 & 6 & 1 \end{pmatrix} \qquad \text{and} \qquad \tilde{B}_{42} = \begin{pmatrix} 1 & 2 & -1 \\ -3 & 1 & -1 \\ 2 & -3 & 8 \end{pmatrix}$$

Definition 4.6.

Let $A \in M_{n \times n}(\mathbb{F})$. If $n = 1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{A}_{1j}).$$

The scalar $\det(A)$ is called the **determinant** of A and is also denoted by $|A|$. The scalar

$$(-1)^{i+j} \det(\tilde{A}_{ij})$$

is called the **cofactor** of the entry of A in row i, column j.

Definition 4.7.

Letting

$$c_{ij} = (-1)^{i+j} \det(\tilde{A}_{ij})$$

denote the cofactor of the row i, column j entry of A, we can express the formula for the determinant of A as

$$\det(A) = A_{11}c_{11} + A_{12}c_{12} + \cdots + A_{1n}c_{1n}.$$

Thus the determinant of A equals the sum of the products of each entry in row 1 of A multiplied by its cofactor. This formula is called **cofactor expansion along the first row** of A.

Theorem 4.3.

the determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed. That is, for $1 \leq r \leq n$, we have

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u + kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + k \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

wherever k is a scalar and u, v and each a_i are row vectors in $\mathbb{F}^n$.

Corollary 4.1.

If $A \in M_{n \times n}(\mathbb{F})$ has a row consisting entirely of zeros, then $\det(A) = 0$.

Lemma 4.1.

Let $B \in M_{n \times n}(\mathbb{F})$, where $n \geq 2$. If row i of B equals e_k for some k ($1 \leq k \leq n$), then $\det(B) = (-1)^{i+k} \det(\tilde{B}_{ik})$.

Theorem 4.4.
The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if $A \in M_{n \times n}(\mathbb{F})$, then for any integer i ($1 \le i \le n$),

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}).$$

Corollary 4.2.
If $A \in M_{n \times n}(\mathbb{F})$ has two identical rows, then $\det(A) = 0$.

Theorem 4.5.
If $A \in M_{n \times n}(\mathbb{F})$ and B is a matrix obtained from A by interchanging any two rows of A, then $\det(B) = -\det(A)$.

Theorem 4.6.
Let $A \in M_{n \times n}(\mathbb{F})$, and let B be a matrix obtained by adding a multiple of one row of A to another row of A. Then $\det(B) = \det(A)$.

Corollary 4.3.
If $A \in M_{n \times n}(\mathbb{F})$ has rank less than n, then $\det(A) = 0$.

Remark 4.1.
The following rules summarize the effect of an elementary row operation on the determinant of a matrix A in $M_{n \times n}(\mathbb{F})$.

1. If B is a matrix obtained by interchanging any two rows of A, then $\det(B) = -\det(A)$.

2. If B is a matrix obtained by multiplying a row of A by a nonzero scalar k, then $\det(B) = k \det(A)$.

3. If B is a matrix obtained by adding a multiple of one row of A to another row of A, then $\det(B) = \det(A)$.

Lemma 4.2.
The determinant of an upper triangular matrix is the product of its diagonal entries.

3   Properties of Determinants

Remark 4.2.
Because the determinant of the $n \times n$ matrix is 1, we can interpret Remark 4.1 as the following facts about the determinants of elementary matrices.

1. If E is an elementary matrix obtained by interchanging any two rows of I, then $\det(E) = -1$.

2. If E is an elementary matrix obtained by multiplying some row of I by the nonzero scalar k, then $\det(E) = k$.

3. If E is an elementary matrix obtained by adding a multiple of some row of I to another row, then $\det(E) = 1$.

Theorem 4.7.
For any $A, B \in M_{n \times n}(\mathbb{F})$, $\det(AB) = \det(A) \cdot \det(B)$.

Corollary 4.4.
A matrix $A \in M_{n \times n}(\mathbb{F})$ is invertible if and only if $\det(A) \neq 0$. Furthermore, if A is invertible, then $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Theorem 4.8.
For any $A \in M_{n \times n}(\mathbb{F})$, $\det(A^t) = \det(A)$.

Theorem 4.9 (Cramer's Rule).
Let $Ax = b$ be the matrix form of a system of n linear equations in n unknowns, where $x = (x_1, x_2, \ldots, x_n)^t$. If $\det(A) \neq 0$, then this system has a unique solution, and for each k $(k = 1, 2, \ldots, n)$,

$$x_k = \frac{\det(M_k)}{\det(A)},$$

where M_k is the $n \times n$ matrix obtained from A by replacing column k of A by b.

Definition 4.8.
It is possible to interpret the determinant of a matrix $A \in M_{n \times n}(\mathbb{R})$ geometrically. If the rows of A are $a_1, a_2, \ldots, a_n$, respectively, then $|\det(A)|$ is the **n-dimensional volume** (the generalization of are in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$) of the parallelepiped having the vectors $a_1, a_2, \ldots, a_n$ as adjacent sides.

Definition 4.9.
A matrix $M \in M_{n \times n}(\mathbb{C})$ is called **nilpotent** if, for some positive integer k, $M^k = O$, where O is the $n \times n$ zero matrix.

Definition 4.10.
A matrix $M \in M_{n \times n}(\mathbb{C})$ is called **skew-symmetric** if $M^t = -M$.

Definition 4.11.
A matrix $Q \in M_{n \times n}(\mathbb{R})$ is called **orthogonal** if $QQ^t = I$.

Definition 4.12.
A matrix $Q \in M_{n \times n}(\mathbb{C})$ is called **unitary** if $QQ^* = I$, where $Q^* = \overline{Q^t}$.

Definition 4.13.
A matrix $A \in M_{n \times n}(\mathbb{F})$ is called **lower triangular** if $A_{ij} = 0$ for $1 \leq i < j \leq n$.

Definition 4.14.
A matrix of the form

$$\begin{pmatrix} 1 & c_0 & c_0^2 & \cdots & c_0^n \\ 1 & c_1 & c_1^2 & \cdots & c_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & c_n & c_n^2 & \cdots & c_n^n \end{pmatrix}$$

is called a **Vandermonde matrix**.

Definition 4.15.
Let $A \in M_{n \times n}(\mathbb{F})$ be nonzero. For any m $(1 \leq m \leq n)$, and $m \times m$ **submatrix** is obtained by deleting any $n - m$ rows and any $n - m$ columns of A.

Definition 4.16.
The **classical adjoint** of a square matrix A is the transpose of the matrix whose ij-entry is the ij-cofactor of A.

Definition 4.17.
Let $y_1, y_2, \ldots, y_n$ be linearly independent function in $\mathbb{C}^\infty$. For each $y \in \mathbb{C}^\infty$, define $T(y) \in \mathbb{C}^\infty$ by

$$[T(y)](t) = \det \begin{pmatrix} y(t) & y_1(t) & y_2(t) & \cdots & y_n(t) \\ y'(t) & y_1'(t) & y_2'(t) & \cdots & y_n'(t) \\ \vdots & \vdots & \vdots & & \vdots \\ y^{(n)}(t) & y_1^{(n)}(t) & y_2^{(n)}(t) & \cdots & y_n^{(n)}(t) \end{pmatrix}$$

The preceding determinant is called the **Wronskian** of $y, y_1, \ldots, y_n$.

4   Summary – Important Facts about Determinants

Definition 4.18.
The **determinant** of an $n \times n$ matrix A having entries from a field $\mathbb{F}$ is a scalar in $\mathbb{F}$, denoted by $\det(A)$ or $|A|$, and can be computed in the following manner:

1. If A is 1×1, then $\det(A) = A_{11}$, the single entry of A.

2. If A is 2×2, then $\det(A) = A_{11}A_{22} - A_{12}A_{21}$.

3. If A is $n \times n$ for $n > 2$, then

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

(if the determinant is evaluated by the entries of row i of A) or

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

(if the determinant is evaluated by the entries of column j of A), where $\tilde{A}_{ij}$ is the $(n-1) \times (n-1)$ matrix obtained by deleting row i and column j from A.

In the formulas above, the scalar $(-1)^{i+j} \det(\tilde{A}_{ij})$ is called the **cofactor** of the row i column j of A.

Definition 4.19 (Properties of the Determinant).

1. If B is a matrix obtained by interchanging any two rows or interchanging any two columns of an $n \times n$ matrix A, then $\det(B) = -\det(A)$.

2. If B is a matrix obtained by multiplying each entry of some row or column of an $n \times n$ matrix A by a scalar k, then $\det(B) = k \cdot \det(A)$.

3. If B is a matrix obtained from an $n \times n$ matrix A by adding a multiple of row i to row j or a multiple of column i to column j for $i \neq j$, then $\det(B) = \det(A)$.

5 A Characterization of the Determinant

Definition 4.20.
A function $\delta : M_{n \times n}(\mathbb{F}) \to \mathbb{F}$ is called an ***n*-linear function** if it is a linear function of each row of an $n \times n$ matrix when the remaining $n - 1$ rows are held fixed, that is, δ is n-linear if, for every $r = 1, 2, \ldots, n$, we have

$$
\delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u + kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + k\delta \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}
$$

whenever k is a scalar and u, v and each a_i are vectors in $\mathbb{F}^n$.

Definition 4.21.
An n-linear function $\delta : M_{n \times n}(\mathbb{F}) \to \mathbb{F}$ is called **alternating** if, for each $A \ in M_{n \times n}(\mathbb{F})$, we have $\delta(A) = 0$ whenever two adjacent rows of A are identical.

Theorem 4.10.
Let $\delta : M_{n \times n}(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear function.

1. If $A \in M_{n \times n}(\mathbb{F})$ and B is a matrix obtained from A by interchanging any two rows of A, then $\delta(B) = -\delta(A)$.

2. If $A \in M_{n \times n}(\mathbb{F})$ has two identical rows, then $\delta(A) = 0$.

Corollary 4.5.
Let $\delta : M_{n \times n}(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear function. If B is a matrix obtained from $A \in M_{n \times n}(\mathbb{F})$ by adding a multiple of some row of A to another row, then $\delta(B) = \delta(A)$.

Corollary 4.6.
Let $\delta : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear function. if $M \in M_{n\times n}(\mathbb{F})$ has rank less than n, then $\delta(M) = 0$.

Corollary 4.7.
Let $\delta : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear function, and let E_1, E_2 and E_3 in $M_{n\times n}(\mathbb{F})$ be elementary matrices of types 1, 2, and 3, respectively. Suppose that E_2 is obtained by multiplying some row of I by the nonzero scalar k. Then $\delta(E_1) = -\delta(I)$, $\delta(E_2) = k \cdot \delta(I)$, and $\delta(E_3) = \delta(I)$.

Theorem 4.11.
Let $\delta : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ be an alternating n-linear function such that $\delta(I) = 1$. For any $A, B \in M_{n\times n}(\mathbb{F})$, we have $\delta(AB) = \delta(A) \cdot \delta(B)$.

Theorem 4.12.
If $\delta : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is an alternating n-linear function such that $\delta(I) = 1$, then $\delta(A) = \det(A)$ for every $A \in M_{n\times n}(\mathbb{F})$.

Chapter 5

Diagonalization

1 Eigenvalues and Eigenvectors

Definition 5.1.
A linear operator T on a finite-dimensional vector space V is called **diagonalizable** if there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. A square matrix A is called **diagonalizable** if L_A is diagonalizable.

Definition 5.2.
Let T be a linear operator on a vector space V. A nonzero vector $v \in V$ is called an **eigenvector** of T if there exists a scalar λ such that $T(v) = \lambda v$. The scalar λ is called the **eigenvalue** corresponding to the eigenvector v.

Let A be in $M_{n \times n}(\mathbb{F})$. A nonzero vector $v \in \mathbb{F}^n$ is called an **eigenvector** of A if v is an eigenvector of L_A; that is, if $Av = \lambda v$ for some scalar λ. The scalar λ is called the **eigenvalue** of A corresponding to the eigenvector v.

The words *characteristic vector* and *proper vector* are also used in place of *eigenvector*. The corresponding terms for *eigenvalue* are *characteristic value* and *proper value*.

Theorem 5.1.
A linear operator T on a finite-dimensional vector space V is diagonalizable if and only if there exists an ordered basis β for V consisting of eigenvectors of T. Furthermore, if T is diagonalizable, $\beta = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of eigenvectors of T, and $D = [T]_\beta$, then D is a diagonal matrix and D_{ij} is the eigenvalue corresponding to v_j for $1 \le j \le n$.

Remark 5.1.
To *diagonalize* a matrix or linear operator is to find a basis of eigenvectors and the corresponding eigenvalues.

Theorem 5.2.
Let $A \in M_{n \times n}(\mathbb{F})$. Then a scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Definition 5.3.
Let $A \in M_{n \times n}(\mathbb{F})$. The polynomial $f(t) = \det(A - t I_n)$ is called the **characteristic polynomial** of A.

The observant reader may have noticed that the entries of the matrix $A - tI_n$ are not scalars in the field $\mathbb{F}$. They are, however, scalars in another field $\mathbb{F}(t)$, the field of quotients of polynomials in t with coefficients from $\mathbb{F}$. Consequently, any results proved about determinants in chapter 4 remain valid in this context.

Definition 5.4.
Let T be a linear operator on an n-dimensional vector space V with ordered basis β. We define the **characteristic polynomial** $f(t)$ of T to be the characteristic polynomial of $A = [T]_\beta$. That is,

$$f(t) = \det(A - tI_n).$$

Theorem 5.3.
Let $A \in M_{n \times n}(\mathbb{F})$.

1. The characteristic polynomial of A is a polynomial of degree n with leading coefficient $(-1)^n$.

2. A has at most n distinct eigenvalues.

Theorem 5.4.
Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. A vector v is an eigenvector of T corresponding to λ if and only if $v \neq 0$ and $v \in \mathrm{N}\,(T - \lambda I)$.

Definition 5.5.
A **scalar matrix** is a square matrix of the form λI for some scalar λ; that is, a scalar matrix is a diagonal matrix in which all the diagonal entries are equal.

2 Diagonalizability

Theorem 5.5.
Let T be a linear operator on a vector space V, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvectors of T. If $v_1, v_2, \ldots, v_k$ are eigenvectors of T such that λ_i corresponds to v_i $(1 \leq i \leq k)$, then $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Corollary 5.1.
Let T be a linear operator on an n-dimensional vector space V. If T has n distinct eigenvalues, then T is diagonalizable.

Definition 5.6.
A polynomial $f(t)$ in $P(\mathbb{F})$ **splits over** $\mathbb{F}$ if there are scalars $c, a_1, \ldots, a_n$ (not necessarily distinct) in $\mathbb{F}$ such that

$$f(t) = c(t - a_1)(t - a_2) \ldots (t - a_n).$$

Theorem 5.6.
The characteristic polynomial of any diagonalizable linear operator splits.

Definition 5.7.

Let λ be an eigenvalue of a linear operator or matrix with characteristic polynomial $f(t)$. The **(algebraic) multiplicity** of λ is the largest positive integer k for which $(t - \lambda)^k$ is a factor of $f(t)$.

Definition 5.8.

Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Define $E_\lambda = \{x \in V : T(x) = \lambda x\} = \mathrm{N}\,(T - \lambda I_V)$. The set E_λ is called the **eigenspace** of T corresponding to the eigenvalue λ. Analogously, we define the **eigenspace** of a square matrix A to be the eigenspace of L_A.

Theorem 5.7.

Let T be a linear operator on a finite-dimensional vector space V, and let λ be an eigenvalue of T having multiplicity m. Then $1 \leq \dim(E_\lambda) \leq m$.

Lemma 5.1.

Let T be a linear operator, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, 2, \ldots, k$, let $v_i \in E_{\lambda_i}$, the eigenspace corresponding to λ_i. If

$$v_1 + v_2 + \cdots + v_k = 0,$$

then $v_i = 0$ for all i.

Theorem 5.8.

Let T be a linear operator on a vector space V, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of T. For each $i = 1, 2, \ldots, k$, let S_i be a finite linearly independent subset of the eigenspace E_{λ_i}. Then $S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a linearly independent subset of V.

Theorem 5.9.

Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Let $\lambda_1, \lambda_2, \ldots \lambda_k$ be the distinct eigenvalues of T. Then

1. T is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for all i.

2. If T is diagonalizable and β_i is an ordered basis for E_{λ_i} for each i, then $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ for an ordered basis for V consisting of eigenvectors of T.

   Note: We regard $\beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ as an ordered basis in the natural way – the vectors in β_1 are listed first (in the same order as in β_1), then the vectors in β_2 (in the same order as β_2), etc.

Test for Diagonalization

Remark 5.2.

Let T be a linear operator on an n-dimensional vector space V. Then T is diagonalizable if and only if both of the following conditions hold.

1. The characteristic polynomial of T splits.

2. For each eigenvalue λ of T, the multiplicity of λ equals $n - \text{rank}\,(T - \lambda I)$.

Definition 5.9.
Let $W_1, W_2, \ldots, W_k$ be subspaces of a vector space V. We define the **sum** of these subspaces to be the set

$$\{v_1 + v_2 + \cdots + v_k : v_i \in W_i \text{ for } 1 \le i \le k\}$$

which we denote by $W_1 + W_2 + \cdots + W_k$ or $\displaystyle\sum_{i=1}^{k} W_i$.

Definition 5.10.
Let $W_1, W_2, \ldots, W_k$ be subspaces of a vector space V. We call V the **direct sum** of the subspaces $W_1, W_2, \ldots, W_k$ and write $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, if

$$V = \sum_{i=1}^{k} W_i$$

and

$$W_j \cap \sum_{i \ne j} W_i = \{0\} \quad \text{for each } j \ (1 \le j \le k)$$

Theorem 5.10.
Let $W_1, W_2, \ldots, W_k$ be subspaces of a finite-dimensional vector space V. The following conditions are equivalent.

1. $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

2. $V = \displaystyle\sum_{i=1}^{k} W_i$ and, for any vectors $v_1, v_2, \ldots, v_k$ such that $v_i \in W_i$ $(1 \le i \le k)$, if $v_1 + v_2 + \cdots + v_k = 0$, then $v_i = 0$ for all i.

3. Each vector $v \in V$ can be uniquely written as $v = v_1 + v_2 + \cdots + v_k$, where $v_i \in W_i$.

4. If γ_i is an ordered basis for W_i $(1 \le i \le k)$, then $\gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_k$ is an ordered basis for V.

5. For each $i = 1, 2, \ldots, k$, there exists an ordered basis γ_i for W_i such that $\gamma_i \cup \gamma_2 \cup \cdots \cup \gamma_k$ is an ordered basis for V.

Theorem 5.11.
A linear operator T on a finite-dimensional vector space V is diagonalizable if and only if V is the direct sum of the eigenspaces of T.

Definition 5.11.
Two linear operators T and U on a finite-dimensional vector space V are called **simultaneously diagonalizable** if there exists an ordered basis β or V such that both $[T]_\beta$ and $[U]_\beta$ are diagonal matrices. Similarly, $A, B \in M_{n \times n}(\mathbb{F})$ are called **simultaneously diagonalizable** if there exists an invertible matrix $Q \in M_{n \times n}(\mathbb{F})$ such that both $Q^{-1}AQ$ and $Q^{-1}BQ$ are diagonal matrices.

3   Matrix Limits and Markov Chains

Definition 5.12.
Let $L, A_1, A_2, \ldots$ be $n \times p$ matrices having complex entries. The sequence $A_1, A_2, \ldots$ is said to **converge** to the $n \times p$ matrix L, called the **limit** of the sequence, if

$$\lim_{m \to \infty} (A_m)_{ij} = L_{ij}$$

for all $1 \le i \le n$ and $1 \le j \le p$. To designate that L is the limit of the sequence, we write

$$\lim_{m \to \infty} A_m = L.$$

Theorem 5.12.
Let $A_1, A_2, \ldots$ be a sequence of $n \times p$ matrices with complex entries that converges to the matrix L. Then for any $P \in M_{r \times n}(\mathbb{C})$ and $Q \in M_{p \times s}(\mathbb{C})$,

$$\lim_{m \to \infty} PA_m = PL \quad \text{and} \quad \lim_{m \to \infty} A_m Q = LQ.$$

Corollary 5.2.
Let $A \in M_{n \times n}(\mathbb{C})$ be such that $\lim_{m \to \infty} A^m = L$. Then for any invertible matrix $Q \in M_{n \times n}(\mathbb{C})$,

$$\lim_{m \to \infty} (QAQ^{-1})^m = QLQ^{-1}.$$

Theorem 5.13.
Let A be a square matrix with complex entries. Then $\lim_{m \to \infty} A^m$ exists if and only if both of the following conditions hold.

1. Every eigenvalue of A is contained in S.

2. If 1 is an eigenvalue of A, then the dimension of the eigenspace corresponding to 1 equals the multiplicity of 1 as an eigenvalue of A.

Theorem 5.14.
Let $A \in M_{n \times n}(\mathbb{C})$ satisfy the following two conditions.

1. Every eigenvalue of A is contained in S.

2. A is diagonalizable.

Then $\lim_{m \to \infty} A^m$ exists.

Definition 5.13. Any square matrix having the following two properties is called a **transition matrix** or a **stochastic matrix**.

1. All entries are non-negative

2. The sum of entries in each column sums up to 1.

For an arbitrary $n \times n$ transition matrix M, the rows and columns correspond to n **states**, and the entry M_{ij} represents the probability of moving from state j to state i in one **stage**.

Definition 5.14. A vector with non-negative entries that add up to 1 is called a **probability vector**.

Theorem 5.15.
Let M be an $n \times n$ matrix having real non-negative entries, let v be a column vector in $\mathbb{R}^n$ having non-negative coordinates, and let $u \in \mathbb{R}^n$ be the column vector in which each coordinate equals 1. Then

1. M is a transition matrix if and only if $M^t u = u$;

2. v is a probability vector if and only if $u^t v = (1)$.

Corollary 5.3.

1. The product of two $n \times n$ transition matrices is an $n \times n$ transition matrix. In particular, any power of a transition matrix is a transition matrix.

2. The product of a transition matrix and a probability vector is a probability vector.

Definition 5.15. A process in which elements of a set are each classified as being in one of several fixed states that can switch over time is called a **stochastic process**. The switching to a particular state is described by the probability, and in general this probability depends on such factors as the state in question, the time in question, some or all of the previous states in which the object has been (including the current state), and the states that other objects are in or have been in.

If, however, the probability that an object in one state changes to a different state in a fixed interval of time depends only on the two states (and not on the time, earlier states, or other factors), then the stochastic process is called a **Markov process**. If, in addition, the number of possible states is finite, then the Markov process is called a **Markov chain**.

Definition 5.16.
The vector that describes the initial probability of being in each state of a Markov chain is called the **initial probability vector** for the Markov chain.

Definition 5.17.
A transition matrix is called **regular** if some power of the matrix contains only positive entries.

Definition 5.18.
Let $A \in M_{n \times n}(\mathbb{C})$. For $1 \leq i, j \leq n$, define $\rho_i(A)$ to be the sum of the absolute values of the entries of row i of A, and define $\nu_j(A)$ to be equal to the sum of the absolute values of the entries of column j and A. Thus

$$\rho_i(A) = \sum_{j=1}^{n} |A_{ij}| \quad \text{for } i = 1, 2, \ldots, n$$

54

and

$$\nu_j(A) = \sum_{i=1}^{n} |A_{ij}| \ \text{ for } j = 1, 2, \ldots n.$$

The **row sum** of A, denoted $\rho(A)$, and the **column sum** of A, denoted $\nu(A)$, are defined as

$$\rho(A) = \max\{\rho_i(A) : 1 \le i \le n\} \quad \text{and} \quad \nu(A) = \max\{\nu_j(A) : 1 \le j \le n\}.$$

Definition 5.19.
For an $n \times n$ matrix A, we define the ith **Gerschgorin disk** $\mathbb{C}_i$ to be the disk in the complex plane with center A_{ij} and radius $r_i = \rho_i(A) - |A_{ii}|$; that is,

$$\mathbb{C}_i = \{z \in \mathbb{C} : |z - A_{ii}| < r_i\}$$

Theorem 5.16 (Gerschgorin's Disk Theorem).
Let $A \in M_{n \times n}(\mathbb{C})$. Then ever eigenvalue of A is contained in a Gerschgorin disk.

Corollary 5.4.
Let λ be any eigenvalue of $A \in M_{n \times n}(\mathbb{C})$. Then $|\lambda| \le \rho(A)$.

Corollary 5.5.
Let λ be any eigenvalue of $A \in M_{n \times n}(\mathbb{C})$. Then

$$|\lambda| \le \min\{\rho(A), \nu(A)\}.$$

Corollary 5.6.
If λ is an eigenvalue of a transition matrix, then $|\lambda| \le 1$.

Theorem 5.17.
Every transition matrix has 1 as an eigenvalue.

Theorem 5.18.
Let $A \in M_{n \times n}(\mathbb{C})$ be a matrix in which each entry is positive, and let λ be an eigenvalue of A such that $|\lambda| = \rho(A)$. Then $\lambda = \rho(A)$ and $\{u\}$ is a basis for E_λ, where $u \in \mathsf{C}^n$ is the column vector in which each coordinate equals 1.

Corollary 5.7.
Let $A \in M_{n \times n}(\mathbb{C})$ be a matrix in which each entry is positive, and let λ be an eigenvalue of A such that $|\lambda| = \nu(A)$. Then $\lambda = \nu(A)$, and the dimension of $E_\lambda = 1$.

Corollary 5.8.
Let $A \in M_{n \times n}(\mathbb{C})$ be a transition matrix in which each entry is positive, and let λ be any eigenvalue of A other than 1. Then $|\lambda| < 1$. Moreover, the eigenspace corresponding to the eigenvalue 1 has dimension 1.

Theorem 5.19.
Let A be a regular transition matrix, and let λ be an eigenvalue of A. Then

1. $|\lambda| \leq 1$.

2. If $|\lambda| = 1$, then $\lambda = 1$, and $\dim(E_\lambda) = 1$.

Corollary 5.9.
Let A be a regular transition matrix that is diagonalizable. Then $\lim\limits_{m \to \infty} A^m$ exists.

Theorem 5.20.
Let A be an $n \times n$ regular transition matrix. Then

1. The multiplicity of 1 as an eigenvalue of A is 1.

2. $\lim\limits_{m \to \infty} A^m$ exists.

3. $\mathrm{L} = \lim\limits_{m \to \infty} A^m$ is a transition matrix.

4. $AL = LA = L$.

5. The columns of L are identical. In fact, each column of L is equal to the unique probability vector v that is also an eigenvector of A corresponding to the eigenvalue 1.

6. For any probability vector w, $\lim\limits_{m \to \infty} (A^m w) = v$.

Definition 5.20.
The vector v in Theorem 5.20(5) is called the **fixed probability vector** or **stationary vector** of the regular transition matrix A.

Definition 5.21.
Consider transition matrices that can be represented in the following form:

$$\begin{pmatrix} I & B \\ O & \mathbb{C} \end{pmatrix}$$

where I is an identity matrix and O is a zero matrix. (Such transition matrices are not regular since the lower left block remains O in any power of the matrix.) The states corresponding to the identity submatrix are called **absorbing states** because such a state is never left once it is entered. A Markov chain is called an **absorbing Markov chain** if it is possible to go from an arbitrary state into an absorbing state in a finite number of stages.

Definition 5.22.
For $A \in M_{n \times n}(\mathbb{C})$, define $e^A = \lim\limits_{m \to \infty} B_m$, where

$$B_m = I + A + \frac{A^2}{2!} + \cdots + \frac{A^m}{m!}$$

Thus e^A is the sum of the infinite series

$$I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots,$$

and B_m is the mth partial sum of this series. (Note the analogy with the power series

$$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \dots,$$

which is valid for all complex numbers a.)

4   Invariant Subspaces and the Cayley-Hamilton Theorem

Definition 5.23.
Let T be a linear operator on a vector space V. A subspace W of V is called a T-**invariant subspace** of V if $T(W) \subseteq W$, that is, if $T(v) \in W$ for all $v \in W$.

Definition 5.24.
Let T be a linear operator on a vector space V, and let x be a nonzero vector in V. The subspace

$$W = \text{span}\left(\{x, T(x), T^2(x), \dots\}\right)$$

is called the T-**cyclic subspace of V generated by** x.

Theorem 5.21.
Let T be a linear operator on a finite-dimensional vector space V, and let W be a T-invariant subspace of V. Then the characteristic polynomial of T_W divides the characteristic polynomial of T.

Theorem 5.22.
Let T be a linear operator on a finite-dimensional vector space V, and let W denote the T-cyclic subspace of V generated by a nonzero vector $v \in V$. Let $k = \dim\left(()W\right)$. Then

1. $\{v, T(v), T^2(v), \dots, T^{k-1}(v)\}$ is a basis for W.

2. If $a_0 v + a_1 T(v) + \dots + a_{k-1} T^{k-1}(v) + T^k(v) = 0$, then the characteristic polynomial of T_W is $f(t) = (-1)^k (a_o + a_1 t + \dots + a_{k-1} t^{k-1} + t^k)$.

The Cayley-Hamilton Theorem

Theorem 5.23 (Cayley-Hamilton).
Let T be a linear operator on a finite-dimensional vector space V, and let $f(t)$ be the characteristic polynomial of T. Then $f(T) = T_0$, the zero transformation. That is, T "satisfies" its characteristic equation.

Corollary 5.10 (Cayley-Hamilton Theorem for Matrices).
Let A be an $n \times n$ matrix, and let $f(t)$ be the characteristic polynomial of A. Then $f(A) = O$, the $n \times n$ zero matrix.

Invariant Subspaces and Direct Sums

Theorem 5.24.
Let T be a linear operator on a finite-dimensional vector space V, and suppose that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$, where W_i is a T-invariant subspace of V for each i ($1 \le i \le k$). Suppose that $f_i(t)$ is the characteristic polynomial of T_{W_i} ($1 \le i \le k$). Then $f_1(t) \cdot f_2(t) \cdots \cdots f_k(t)$ is the characteristic polynomial of T.

Definition 5.25.
Let $B_1 \in M_{m \times m}(\mathbb{F})$, and let $B_2 \in M_{n \times n}(\mathbb{F})$. We define the **direct sum** of B_1 and B_2, denoted $B_1 \oplus B_2$, as the $(m + n) \times (m + n)$ matrix A such that

$$A_{ij} = \begin{cases} (B_1)_{ij} & \text{for } 1 \le i, j \le m \\ (B_2)_{(i-m),(j-m)} & \text{for } m + 1 \le i, j \le n + m \\ 0 & \text{otherwise.} \end{cases}$$

If $B_1, B_2, \ldots, B_k$ are square matrices with entries from $\mathbb{F}$, then we define the **direct sum** of $B_1, B_2, \ldots, B_k$ recursively by

$$B_1 \oplus B_2 \oplus \cdots \oplus B_k = (B_1 \oplus B_2 \oplus \cdots \oplus B_{k-1}) \oplus B_k.$$

If $A = B_1 \oplus B_2 \oplus \cdots \oplus B_k$, then we often write

$$A = \begin{pmatrix} B_1 & O & \ldots & O \\ O & B_2 & \ldots & O \\ \vdots & \vdots & & \vdots \\ O & O & \ldots & B_k \end{pmatrix}$$

Theorem 5.25.
Let T be a linear operator on a finite-dimensional vector space V, and let $W_1, W_2, \ldots, W_k$ be T-invariant subspaces of V such that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. For each i, let β_i be an ordered basis for W_i, and let $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$. Let $A = [T]_\beta$ and $B_i = [T_{W_i}]_{\beta_i}$ for $i = 1, 2, \ldots, k$. Then $A = B_1 \oplus B_2 \oplus \cdots \oplus B_k$.

Definition 5.26.
Let T be a linear operator on a vector space V, and let W be a T-invariant subspace of V. Define $\overline{T} : V/W \to V/W$ by

$$\overline{T}(v + W) = T(v) + W \quad \text{for any } v + W \in V/W.$$

Chapter 6

Inner Product Spaces

1   Inner Products and Norms

Definition 6.1.
Let V be a vector space over $\mathbb{F}$. An **inner product** on V is a function that assigns, to every ordered pair of vectors x and y in V, a scalar in $\mathbb{F}$, denoted by $\langle x, y \rangle$, such that for all x, y and z in V and all c in $\mathbb{F}$, the following hold:

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$.

2. $\langle cx, y \rangle = c\langle x, y \rangle$.

3. $\overline{\langle x, y \rangle} = \langle y, x \rangle$, where the bar denoted complex conjugation.

4. $\langle x, x \rangle > 0$ if $x \neq 0$.

Definition 6.2.
If x and y are in the vector space $\mathbb{C}^n$, define their inner product to be

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k \overline{y_k}.$$

This is called the (complex) **standard inner product**.

When $\mathbb{F} = \mathbb{R}$ the conjugations are not needed, and in early courses this standard inner product is usually called the *dot product* and is denoted by $x \cdot y$ instead of $\langle x, y \rangle$.

Definition 6.3.
Let $A \in M_{m \times n}(\mathbb{F})$. We define the **conjugate transpose** or **adjoint** of A to be the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{(A_{ij})}$ for all i, j.

Definition 6.4.
Given two complex number-valued $n \times m$ matrices A, and B, written explicitly as

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{pmatrix}, \quad \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1m} \\ B_{21} & B_{22} & \dots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \dots & B_{nm} \end{pmatrix}$$

the **Frobenius inner product** is defined as,

$$\langle A, B \rangle_{\mathbb{F}} = \sum_{i,j} \overline{A_{ij}} B_{ij} = \operatorname{tr}(A^*A)$$

Where the overline denotes the complex conjugate, and A^* denotes the conjugate transpose (Definition 6.3). Explicitly this sum is

$$\begin{aligned} \langle A, B \rangle_{\mathbb{F}} = &\overline{A_{11}} B_{11} + \overline{A_{12}} B_{12} + \dots + \overline{A_{1m}} B_{1m} \\ &+ \overline{A_{21}} B_{21} + \overline{A_{22}} B_{22} + \dots + \overline{A_{2m}} B_{2m} \\ &\vdots \\ &+ \overline{A_{n1}} B_{n1} + \overline{A_{n2}} B_{n2} + \dots + \overline{A_{nm}} B_{nm} \end{aligned}$$

Definition 6.5.
A vector space V over $\mathbb{F}$ endowed with a specific inner product is called an **inner product space**. If $\mathbb{F} = \mathbb{C}$, we call V a **complex inner product space**, whereas if $\mathbb{F} = \mathbb{R}$, we call V a **real inner product space**.

Notation 6.1.
For the remainder of this chapter, F^n denotes the inner product space with the standard inner product as defined in Definition 6.2. Likewise, $M_{n \times n}(F)$ denotes the inner product space with the Frobenius inner product as defined in Definition 6.4.

Theorem 6.1.
Let V be an inner product space. Then for $x, y, z \in V$ and $c \in \mathbb{F}$ the following statements are true.

1. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.

2. $\langle x, cy \rangle = \overline{c} \langle x, y \rangle$.

3. $\langle x, 0 \rangle = \langle 0, x \rangle = 0$.

4. $\langle x, x \rangle = 0$ if and only if $x = 0$.

5. If $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in V$, then $y = z$.

Remark 6.1.
It should be observed that the (1) and (2) of Theorem 6.1 show that the inner product is **conjugate linear** in the second component.

Definition 6.6.
Let V be an inner product space. For $x \in V$, we define the **norm** or **length** of x by
$||x|| = \sqrt{\langle x, x \rangle}$.

Theorem 6.2.
Let V be an inner product space over $\mathbb{F}$. Then for all $x, y \in V$ and $c \in \mathbb{F}$, the following statements are true.

1. $||cx|| = |c| \cdot ||x||$.

2. $||x|| = 0$ if and only if $x = 0$. In any case, $||x|| \geq 0$.

3. (Cauchy-Schwarz Inequality) $|\langle x, y \rangle| \leq ||x|| \cdot ||y||$.

4. (Triangle Inequality) $||x + y|| \leq ||x|| + ||y||$.

Definition 6.7.
Let V be an inner product space. Vectors x and y in V are **orthogonal** (**perpendicular**) if $\langle x, y \rangle = 0$. A subset S of V is **orthogonal** if any two distinct vectors in S are orthogonal. A vector x in V is a **unit vector** if $||x|| = 1$. Finally, a subset S of V is **orthonormal** if S is orthogonal and consists entirely of unit vectors.

Note that if $S = \{v_1, v_2, \dots\}$, then S is orthonormal if and only if $\langle v_i, v_j \rangle = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. Also, observe that multiplying vectors by nonzero scalars does not affect their orthogonality and that if x is any nonzero vector, then $(1/||x||)x$ is a unit vector. The process of multiplying a nonzero vector by the reciprocal of its length is called **normalizing**.

Lemma 6.1 (Parallelogram Law).
Let V be an inner product space. Then

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2 \quad \text{for all } x, y \in V$$

Definition 6.8.
Let V be a vector space $\mathbb{F}$, where $\mathbb{F}$ is either $\mathbb{R}$ of $\mathbb{C}$. Regardless of whether V is or is not an inner product space, we may still define a norm $|| \cdot ||$ as a real-values function on V satisfying the following three conditions for all $x, y \in V$ and $a \in \mathbb{F}$:

1. $||x|| \geq 0$, and $||x|| = 0$ if and only if $x = 0$.

2. $||ax|| = |a| \cdot ||x||$.

3. $||x + y|| \leq ||x|| + ||y||$.

Definition 6.9.
Let $|| \cdot ||$ be a norm on a vector space V, and define, for each ordered pair of vectors, the scalar $d(x, y) = ||x - y||$. This is called the **distance** between x and y.

2 The Gram-Schmidt Orthogonalization Process and Orthogonal Complements

Definition 6.10.
Let V be an inner product space. A subset of V is an **orthonormal basis** for V if it is an ordered basis that is orthonormal.

Theorem 6.3.
Let V be an inner product space and $S = \{v_1, v_2, \ldots, v_k\}$ be an orthogonal subset of V consisting of nonzero vectors. If $y \in \text{span}(S)$, then

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{||v_i||^2} v_i.$$

Corollary 6.1.
If, in addition to the hypotheses of Theorem 6.3, S is orthonormal and $y \in \text{span}(()\, S)$, then

$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i.$$

Corollary 6.2.
Let V be an inner product space, and let S be an orthogonal subset of V consisting of nonzero vectors. Then S is linearly independent.

Theorem 6.4.
Let V be an inner product space and $S = \{w_1, w_2, \ldots, w_n\}$ be a linearly independent subset of V. Define $S' = \{v_1, v_2, \ldots, v_n\}$ where $v_1 = w_1$ and

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{||v_j||^2} v_j \quad \text{for } 2 \leq k \leq n.$$

Then S' is an orthogonal set of nonzero vectors such that $\text{span}(S') = \text{span}(S)$.

Definition 6.11.
The **Gram-Schmidt process** is a way of making two or more vectors perpendicular to each other. It is a method of constructing an orthonormal basis from a set of vectors in an inner product space. It takes a finite, linearly independent set of vectors $S = \{v_1, v_2, \ldots, v_k\}$ for $k \leq n$ and generates an orthogonal set $S' = \{u_1, u_2, \ldots, u_k\}$ that spans the same k-dimensional subspace of $\mathbb{R}^n$ as S.

The **vector projection** of a vector v on a nonzero vector u is defined as

$$\text{proj}_u(v) = \frac{\langle v, u \rangle}{\langle u, u \rangle} u,$$

This means that $\text{proj}_u(v)$ is the orthogonal projection of v onto the line spanned by u. If u is the zero vector, then $\text{proj}_u(v)$ is defined as the zero vector.

Given k vectors $v_1, v_2, \ldots, v_k$, the Gram-Schmidt process defines the vectors $u_1, u_2, \ldots, u_k$ as follows:

$$u_1 = v_1, \qquad\qquad e_1 = \frac{u_1}{||u_1||}$$

$$u_2 = v_2 - \text{proj}_{u_1}(v_2), \qquad\qquad e_2 = \frac{u_2}{||u_2||}$$

$$u_3 = v_3 - \text{proj}_{u_1}(v_3) - \text{proj}_{u_2}(v_3), \qquad\qquad e_3 = \frac{u_3}{||u_3||}$$

$$u_4 = v_4 - \text{proj}_{u_1}(v_4) - \text{proj}_{u_2}(v_4) - \text{proj}_{u_3}(v_4), \qquad e_4 = \frac{u_4}{||u_4||}$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$u_k = v_k = \sum_{j=1}^{k-1} \text{proj}_{u_j}(v_k), \qquad\qquad e_k = \frac{u_k}{||u_k||}.$$

The sequence $u_1, u_2, \ldots, u_k$ is the required system of orthogonal vectors, and the normalized vectors $e_1, e_2, \ldots, e_k$ form an orthonormal set.

Theorem 6.5.
Let V be a nonzero finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, v_2, \ldots, v_n\}$ and $x \in V$, then

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i.$$

Corollary 6.3.
Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then for any i and j, $A_{ij} = \langle T(v_j), v_i \rangle$.

Definition 6.12.
Let β be an orthonormal subset (possibly infinite) of an inner product space V, and let $x \in V$. We define the **Fourier coefficients** of x relative to β to be the scalars $\langle x, y \rangle$, where $y \in \beta$.

Definition 6.13.
Let S be a nonempty subset of an inner product space V. We define $S^\perp$ (read "S perp") to be the set of all vectors in V that are orthogonal to every vector in S; that is, $S^\perp ='$ $x \in V : \langle x, y \rangle = 0, \ \forall y \in S$. The set $S^\perp$ is called the **orthogonal complement of** S.

Theorem 6.6.
Let W be a finite-dimensional subspace of an inner product space V, and let $y \in V$. Then there exist unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal basis for W, then

$$u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i.$$

63

Corollary 6.4.
In the notation of Theorem 6.6, the vector u is the unique vector in W that is "closest" to y; that is, for any $x \in W$, $||y - x|| \geq ||y - u||$, and this inequality is an equality if and only if $x = u$.

Theorem 6.7.
Suppose that $S = \{v_1, v_2, \ldots, v_k\}$ is an orthonormal set in an n-dimensional inner product space V. Then

1. S can be extended to an orthonormal basis $\{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V.

2. If $W = \text{span}(S)$, then $S_1 = \{v_{k+1}, v_{k+2}, \ldots, v_n\}$ is an orthonormal basis for $W^\perp$ (using the preceding notation).

3. If W is any subspace of V, then $\dim(V) = \dim(W) + \dim(W^\perp)$.

Definition 6.14 (Parseval's Identity).
Let V be a finite-dimensional inner product space over $\mathbb{F}$, and let $\{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis for V. Then for any $x, y \in V$,

$$\langle x, y \rangle = \sum_{i=1}^{n} \langle x, v_i \rangle \overline{\langle y, v_i \rangle}.$$

Definition 6.15 (Bessel's Inequality). Let V be an inner product space, and let $S = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal subset of V. For any $x \in V$,

$$||x||^2 \geq \sum_{i=1}^{n} |\langle x, v_i \rangle|^2.$$

3 The Adjoint of a Linear Operator

Definition 6.16.
For a linear operator T on an inner product space V, we define a related linear operator on V called the **adjoint** of T, whose matrix representation with respect to any orthonormal basis β for V is $[T]_\beta^*$.

Theorem 6.8.
Let V be a finite-dimensional inner product space over $\mathbb{F}$, and let $\mathbf{g} : V \to \mathbb{F}$ be a linear transformation. Then there exists a unique vector $y \in V$ such that $\mathbf{g}(x) = \langle x, y \rangle$ for all $x \in V$.

Theorem 6.9.
Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Then there exists a unique function $T^* : V \to V$ such that $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ for all $x, y \in V$. Furthermore, T^* is linear.

Remark 6.2.
Important Note: For the remainder of this chapter we adopt the convention that a reference to the adjoint of a linear operator on an infinite-dimensional inner product space assumes its existence.

Theorem 6.10.
Let V be a finite-dimensional inner product space, and let β be an orthonormal basis for V. If T is a linear operator on V, then

$$[T^*]_\beta = [T]_\beta^*.$$

Corollary 6.5.
Let A be an $n \times n$ matrix. Then $L_{A^*} = (L_A)^*$.

Theorem 6.11.
Let V be an inner product space, and let T and U be linear operators on V. Then

1. $(T + U)^* = T^* + U^*$;

2. $(cT)^* = \bar{c}T^*$ for any $c \in \mathbb{F}$;

3. $(TU)^* = U^*T^*$;

4. $T^{**} = T$;

5. $I^* = I$.

Corollary 6.6.
Let A and B be $n \times n$ matrices. Then

1. $(A + B)^* = A^* + B^*$;

2. $(cA)^* = \bar{c}A^*$ for all $c \in \mathbb{F}$;

3. $(AB)^* = B^*A^*$;

4. $A^{**} = A$;

5. $I^* = I$.

Least Squares Approximation

Lemma 6.2.
Let $A \in M_{m \times n}(\mathbb{F}), x \in F^n$, and $y \in F^m$. Then

$$\langle Ax, y \rangle_m = \langle x, A^*y \rangle_n.$$

Lemma 6.3.
Let $A \in M_{m \times n}(\mathbb{F})$. Then $\mathrm{rank}\,(A^*A) = \mathrm{rank}\,(A)$.

Corollary 6.7.
If A is an $m \times n$ matrix such that $\mathrm{rank}\,(A) = n$, then A^*A is invertible.

Theorem 6.12.
Let $A \in M_{m \times n}(\mathbb{F})$ and $y \in F^m$. Then there exists $x_0 \in F^n$ such that $(A^*A)x_0 = A^*y$ and $||Ax_0 = y|| \leq ||Ax - y||$ for all $x \in F^n$. Furthermore, if $\mathrm{rank}\,(A) = n$, then $x_0 = (A^*A)^{-1}A^*y$.

Definition 6.17.
A solution s to a system of linear equations $Ax = b$ is called a **minimal solution** if $||s|| \leq ||u||$ for all other solutions u.

Minimal Solutions to Systems of Linear Equations

Theorem 6.13.
Let $A \in M_{m \times n}(\mathbb{F})$ and $b \in F^m$. Suppose that $Ax = b$ is consistent. Then the following statements are true.

1. There exists exactly one minimal solution s of $Ax = b$, and $x \in \mathrm{R}\,(L_{A^*})$.

2. The vector s is the only solution to $Ax = b$ that lies in $\mathrm{R}\,(L_{A^*})$; that is, if u satisfies $(AA^*)u = b$, then $s = A^*u$.

Definition 6.18.
Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional inner product spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, respectively. A function $T^* : W \to V$ is called an **adjoint** of T if $\langle T(x), y \rangle_2 = \langle x, T^*y \rangle_1$ for all $x \in V$ and $y \in W$.

   This is an extension of the definition of the *adjoint* of a linear operator (Definition 6.16)

Definition 6.19.
In physics, *Hooke's law* states that (within certain limits) there is a linear relationship between the length x of a spring and the force y applied to (or exerted by) the spring. That is, $y = cx + d$, where c is called the **spring constant**.

4 Normal and Self-Adjoint Operators

Lemma 6.4.
Let T be a linear operator on a finite-dimensional inner product space V. If T has an eigenvector, then so does T^*.

Theorem 6.14 (Schur).
Let T be a linear operator on a finite-dimensional inner product space V. Suppose that the characteristic polynomial of T splits. Then there exists an orthonormal basis β for V such that the matrix $[T]_\beta$ is upper triangular.

Definition 6.20.
Let V be an inner product space, and let T be a linear operator on V. We say that T is **normal** if $TT^* = T^*T$. An $n \times n$ real or complex matrix A is **normal** if $AA^* = A^*A$.

Theorem 6.15.
Let V be an inner product space, and let T be a normal operator on V. Then the following statements are true.

1. $||T(x)|| = ||T^*(x)||$ for all $x \in V$.

2. $T - cI$ is normal for every $c \in \mathbb{F}$.

3. If x is an eigenvector of T, then x is also an eigenvector of T^*. In fact, if $T(x) = \lambda x$, then $T^*(x) = \bar{\lambda} x$.

4. If λ_1 and λ_2 are distinct eigenvalues of T with corresponding eigenvectors x_1 and x_2, then x_1 and x_2 are orthogonal.

Theorem 6.16.
Let T be a linear operator on a finite-dimensional complex inner product space V. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

Definition 6.21.
Let T be a linear operator on an inner product space V. We say that T is **self-adjoint** (**Hermitian**) if $T = T^*$. An $n \times n$ real or complex matrix A is **self-adjoint** (**Hermitian**) if $A = A^*$.

Lemma 6.5.
Let T be a self-adjoint operator on a finite-dimensional inner product space V. Then

1. Every eigenvalue of T is real.

2. Suppose that V is a real inner product space. Then the characteristic polynomial of T splits.

Theorem 6.17.
Let T be a linear operator on a finite-dimensional real inner product space V. Then T is self-adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.

Definition 6.22.
An $n \times n$ real matrix A is said to be a **Gramian** matrix if there exists a real (square) matrix B such that $A = B^t B$.

Definition 6.23.
A linear operator T on a finite-dimensional inner product space is called **positive definite** [**positive semidefinite**] if T is self-adjoint and $\langle T(x), x \rangle > 0$ [$\langle T(x), x \rangle \geq 0$] for all $x \neq 0$.

An $n \times n$ matrix A with entries from $\mathbb{R}$ or $\mathbb{C}$ is called **positive definite** [**positive semidefinite**] if L_A is positive definite [positive semidefinite].

5   Unitary and Orthogonal Operators and Their Matrices

Definition 6.24.
Let T be a linear operator on a finite-dimensional inner product space V (over $\mathbb{F}$). If $||T(x)|| = ||x||$ for all $x \in V$, we call T a **unitary operator** if $\mathbb{F} = \mathbb{C}$ and an **orthogonal operator** if $\mathbb{F} = \mathbb{R}$.

It should be noted that, in the infinite-dimensional case, an operator satisfying the preceding norm requirement is generally called an **isometry**. If, in addition, the operator is onto (the condition guarantees one-to-one), then the operator is called a **unitary** or **orthogonal operator**,

Theorem 6.18.
Let T be a linear operator on a finite-dimensional inner product space V. Then the following statements are equivalent.

1. $TT^* = T^*T = I$.

2. $\langle T(x), T(y) \rangle = \langle x, y \rangle$ for all $x, y \in V$.

3. If β is an orthonormal basis for V, then $T(\beta)$ is an orthonormal basis for V.

4. There exists an orthonormal basis β for V such that $T(\beta)$ is an orthonormal basis for V.

5. $||T(x)|| = ||x||$ for all $x \in V$.

Lemma 6.6.
Let U be a self-adjoint operator on a finite-dimensional inner product space V. If $\langle x, U(x) \rangle = 0$ for all $x \in V$, then $U = T_0$.

Corollary 6.8.
Let T be a linear operator on a finite-dimensional real inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is both self-adjoint and orthogonal.

Corollary 6.9.
Let T be a linear operator on a finite-dimensional complex inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is unitary.

Definition 6.25.
Let L be a one-dimensional subspace of $\mathbb{R}^2$. We may view L as a line in the plane through the origin. A linear operator T on $\mathbb{R}^2$ is called a **reflection of $\mathbb{R}^2$ about** L if $T(x) = x$ for all $x \in L$ and $T(x) = -x$ for all $x \in L^\perp$.

Definition 6.26.
A square matrix A is called an **orthogonal matrix** if $A^t A = A A^t = I$ and **unitary** if $A^*A = AA^* = I$.

Since for a real matrix A we have $A^* = A^t$, a real unitary matrix is also orthogonal. In this case, we call A **orthogonal** rather than unitary.

Definition 6.27.
We know that, for a complex normal [real symmetric] matrix A, there exists an orthonormal basis β for F^n consisting of eigenvectors of A. Hence A is similar to a diagonal matrix D. By Corollary 2.8, the matrix Q whose columns are the vectors in β is such that $D = Q^{-1}AQ$. But since the columns of Q are an orthonormal basis for F^n, it follows that Q is unitary [orthogonal]. In this case, we say that A is **unitarily equivalent** [**orthogonally equivalent**] to D. It is easily seen that this relation is an equivalence relation on $M_{n \times n}(\mathbb{C})$ [$M_{n \times n}(\mathbb{R})$]. More generally, *A and B are unitarily equivalent [orthogonally equivalent]* if and only if there exists a unitary [orthogonal] matrix P such that $A = P^*BP$.

Theorem 6.19.
Let A be a complex $n \times n$ matrix. Then A is normal if and only if A is unitarily equivalent to a diagonal matrix.

Theorem 6.20.
Let A be a real $n \times n$ matrix. Then A is symmetric if and only if A is orthogonally equivalent to a real diagonal matrix.

Theorem 6.21 (Schur).
Let $A \in M_{n \times n}(\mathbb{F})$ be a matrix whose characteristic polynomial splits over $\mathbb{F}$.

1. If $\mathbb{F} = \mathbb{C}$, then A is unitarily equivalent to a complex upper triangular matrix.

2. If $\mathbb{F} = \mathbb{R}$, then A is orthogonally equivalent to a real upper triangular matrix.

Rigid Motions

Definition 6.28.
Let V be a real inner product space. A function $f : V \to V$ is called a **rigid motion** if

$$||f(x) - f(y)|| = ||x - y||$$

for all $x, y \in V$.

Definition 6.29.
Let V be a real inner product space. A function $g : V \to V$ is called a **translation** if there exists a vector $v_0 \in V$ such that $g(x) = x + v_0$ for all $x \in V$. We say that g is the *translation by v_0*.

Theorem 6.22.
Let $f : V \to V$ be a rigid motion on a finite-dimensional real inner product space V. Then there exists a unique orthogonal operator T on V and a unique translation g on V such that $f = g \circ T$.

Orthogonal Operators on $\mathbb{R}^2$

Theorem 6.23.
Let T be an orthogonal operator on $\mathbb{R}^2$, and let $A = [T]_\beta$ where β is the standard ordered basis for $\mathbb{R}^2$. Then exactly one of the following conditions is satisfied:

1. T is a rotation, and $\det(A) = 1$.

2. T is a reflection about a line through the origin, and $\det(A) = -1$.

Corollary 6.10.
Any rigid motion on $\mathbb{R}^2$ is either a rotation followed by a translation or a reflection about a line through the origin followed by a translation.

Conic Sections

Definition 6.30. Consider the quadratic equation

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0. \tag{6.1}$$

The expression

$$ax^2 + 2bxy + cy^2$$

is called the **associated quadratic form** of (6.1)

Definition 6.31.
Let V be a finite-dimensional inner product space. A linear operator U on V is called a **partial isometry** if there exists a subspace W of V such that $||U(x)|| = ||x||$ for all $x \in W$ and $U(x) = 0$ for all $x \in W^\perp$.

Definition 6.32. Let V be a finite-dimensional complex [real] inner product space, and let u be a unit vector in V. Define the **Householder** operator $\mathsf{H}_u : V \to V$ by $\mathsf{H}_u(x) = x - 2\langle x, u \rangle u$ for all $x \in V$.

6 Orthogonal Projections and the Spectral Theorem

Definition 6.33.
Let V be an inner product space, and let $T : V \to V$ be a projection. We say that T is an **orthogonal projection** if $\mathrm{R}\,(T)^\perp = \mathrm{N}\,(T)$ and $\mathrm{N}\,(T)^\perp = \mathrm{R}\,(T)$.

Definition 6.34.
Let W be a finite-dimensional subspace of an inner product space V, and let T be an orthogonal projection on W. We call T the **orthogonal projection of V on W**.

Definition 6.35. A **trigonometric polynomial of degree** n is a function $g \in \mathsf{H}$ of the form

$$g(t) = \sum_{j=-n}^{n} a_j f_j(t) = \sum_{j=-n}^{n} a_j e^{ijt},$$

where a_n or a_{-n} is nonzero.

Theorem 6.24.
Let V be an inner product space, and let T be a linear operator on V. Then T is an orthogonal projection if and only if T has an adjoint T^* and $T^2 = T = T^*$.

Theorem 6.25 (The Spectral Theorem).
Suppose that T is a linear operator on a finite-dimensional inner product space V over $\mathbb{F}$ with the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Assume that T is normal if $\mathbb{F} = \mathbb{C}$ and that T is self-adjoint if $\mathbb{F} = \mathbb{R}$. For each i ($1 \le i \le k$), let W_i be the eigenspace of T corresponding to the eigenvalue λ_i, and let T_i be the orthogonal projection of V on W_i. Then the following statements are true.

    1. $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

2. If W_i' denotes the direction sum of the subspaces W_j for $j \neq i$, then $W_i^\perp = W_i'$.

3. $T_i T_j = \delta_{ij} T_i$ for $1 \leq i, j \leq k$.

4. $I = T_1 + T_2 + \cdots + T_k$.

5. $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$.

Definition 6.36.
In the context of Theorem 6.25:

1. The set $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ of eigenvalues of T is called the **spectrum** of T.

2. The sum $I = T_1 + T_2 + \cdots + T_k$ in (4) is called the **resolution of the identity operator** induced by T.

3. The sum $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ in (5) is called the **spectral decomposition** of T. The spectral decomposition of T is unique up to the order of its eigenvalues.

Corollary 6.11.
If $\mathbb{F} = \mathbb{C}$, then T is normal if and only if $T^* = g(T)$ for some polynomial g.

Corollary 6.12.
If $\mathbb{F} = \mathbb{C}$, then T is unitary if and only if T is normal and $|\lambda| = 1$ for every eigenvalue of T.

Corollary 6.13.
If $\mathbb{F} = \mathbb{C}$ and T is normal, then T is self-adjoint if and only if every eigenvalue of T is real.

Corollary 6.14.
Let T be as in Theorem 6.25 with spectral decomposition $T = \lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$. Then each T_j is a polynomial in T.

7   The Singular Value Decomposition and the Pseudoinverse

Theorem 6.26 (Singular Value Theorem for Linear Transformations).
Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be a linear transformation of rank r. Then there exist orthonormal bases $\{v_1, v_2, \ldots, v_n\}$ for V and $\{u_1, u_2, \ldots, u_m\}$ for W and positive scalars $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ such that

$$T(v_i) = \begin{cases} \sigma_i u_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}$$

Conversely, suppose that the preceding conditions are satisfied. Then for $1 \leq i \leq n$, v_i is an eigenvector of T^*T with corresponding eigenvalue σ_i^2 if $1 \leq i \leq r$ and 0 if $i > r$. Therefore the scalars $\sigma_1, \sigma_2, \ldots, \sigma_r$ are uniquely determined by T.

Definition 6.37.
The unique scalars $\sigma_1, \sigma_2, \ldots \sigma_r$ in Theorem 6.26 are called the **singular values** of T. If r is less than both m and n, then the term *singular value* is extended to include $\sigma_{r+1} = \cdots = \sigma_k = 0$, where k is the minimum of m and n.

Definition 6.38.
Let A be an $m \times n$ matrix. We define the **singular values** of A to be the singular values of the linear transformation L_A.

Theorem 6.27 (Singular Value Decomposition Theorem for Matrices).
Let A be an $m \times n$ matrix of rank r with the positive singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$, and let Σ be the $m \times n$ matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if } i = j \leq r \\ 0 & \text{otherwise.} \end{cases}$$

Then there exists an $m \times m$ unitary matrix U and an $n \times n$ unitary matrix V such that

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

Definition 6.39.
Let A be an $m \times n$ matrix of rank r with positive singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. A factorization $A = U\Sigma V^*$ where U and V are unitary matrices and Σ is the $m \times n$ matrix defined as in Theorem 6.27 is called a **singular value decomposition** of A.

The Polar Decomposition of a Square Matrix

Theorem 6.28 (Polar Decomposition).
For any square matrix A, there exists a unitary matrix W and a positive semidefinite matrix P such that

$$A = WP.$$

Furthermore, if A is invertible, then the representation is unique.

Definition 6.40.
The factorization of a square matrix A as WP where W is unitary and P is positive semidefinite is called a **polar decomposition** of A.

The Pseudoinverse

Definition 6.41.
Let V and W be finite-dimensional inner product spaces over the same field, and let $T : V \to W$ be a linear transformation. Let $L : \mathrm{N}(T)^\perp \to \mathrm{R}(T)$ be the linear transformation defined by $L(x) = T(x)$ for all $x \in \mathrm{N}(T)^\perp$. The **pseudoinverse** (or *Moore-Penrose generalized inverse*) of T, denoted by $T^\dagger$, is defined as the unique linear transformation from W to V such that

$$T^\dagger(y) = \begin{cases} L^{-1}(y) & \text{for } y \in \mathrm{R}(T) \\ 0 & \text{for } y \in \mathrm{R}(T)^\perp. \end{cases}$$

Definition 6.42.

Let A be an $m \times n$ matrix. Then there exists a unique $n \times m$ matrix B such that $(L_A)^\dagger : F^m \to F^n$ is equal to the left-multiplication transformation L_B. We call B the **pseudoinverse** of A and denote it by $B = A^\dagger$. Thus

$$(L_A)^\dagger = L_{A^\dagger}$$

Theorem 6.29.

Let A be an $m \times n$ matrix of rank r with a singular value decomposition $A = U\Sigma V^*$ and nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Let $\Sigma^\dagger$ be the $n \times m$ matrix defined by

$$\Sigma^\dagger_{ij} = \begin{cases} \frac{1}{\sigma_i} & \text{if } i = j \leq r \\ 0 & \text{otherwise.} \end{cases}$$

Then $A^\dagger = V\Sigma^\dagger U^*$, and this is a singular value decomposition of $A^\dagger$.

The Pseudoinverse and Systems of Linear Equations

Lemma 6.7.

Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be linear. Then

1. $T^\dagger T$ is the orthogonal projection of V on $\text{N}(T)^\dagger$.

2. $TT^\dagger$ is the orthogonal projection of W on $\text{R}(T)$.

Theorem 6.30.

Consider the system of linear equations $Ax = b$, where A is an $m \times n$ matrix and $b \in F^m$. If $z = A^\dagger b$, then z has the following properties.

1. If $Ax = b$ is consistent, then z is the unique solution to the system having minimum norm. That is, z is a solution to the system, and if y is any solution to the system, then $||z|| \leq ||y||$ with equality if and only if $z = y$.

2. If $Ax = b$ is inconsistent, then z is the unique best approximation to a solution having minimum norm. That is, $||Az - b|| \leq ||Ay - b||$ for any $y \in F^n$, with equality if and only if $Az = Ay$. Furthermore, if $Az = Ay$, then $||z|| \leq ||y||$ with equality if and only if $z = y$.

Lemma 6.8 (Penrose Conditions).

Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be linear. Then the following conditions hold:

1. $TT^\dagger T = T$.

2. $T^\dagger TT^\dagger = T^\dagger$.

3. Both $T^\dagger T$ and $TT^\dagger$ are self-adjoint.

These conditions characterize the pseudoinverse of a linear transformation.

8   Bilinear and Quadratic Forms

Bilinear Forms

Definition 6.43.
Let V be a vector space over a field $\mathbb{F}$. A function H from the set $V \times V$ if ordered pairs of vectors to $\mathbb{F}$ is called a **bilinear form** on V if H is linear in each variable when the other variable is held fixed; that is, H is a bilinear form on V if

1. $H(ax_1 + x_2, y) = aH(x_1, y) + H(x_2, y)$ for all $x_1, x_2, y \in V$ and $a \in \mathbb{F}$.

2. $H(x, ay_1 + y_2) = aH(x, y_1) + H(x, y_2)$ for all $x, y_1, y_2 \in V$ and $a \in \mathbb{F}$.

Notation 6.2.
We denote the set of all bilinear forms on V by $\mathcal{B}(V)$.

Definition 6.44.
Let V be a vector space, let H_1 and H_2 be bilinear forms on V, and let a be a scalar. We define the **sum** $H_1 + H_2$ and the **scalar product** aH_1 by the equations

$$(H_1 + H_2)(x, y) = H_1(x, y) + H_2(x, y)$$

and

$$(aH_1)(x, y) = a(H_1(x, y)) \quad \text{for all } x, y \in V.$$

Theorem 6.31.
For any vector space V, the sum of two bilinear forms and the product of a scalar and a bilinear form on V are again bilinear forms on V. Furthermore, $\mathcal{B}(V)$ is a vector space with respect to these operations.

Definition 6.45. let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for an n-dimensional vector space V, and let $H \in \mathcal{B}(V)$. We can associate H with an $n \times n$ matrix A whose entry in row i and column j is defined by

$$A_{ij} = H(v_i, v_j) \quad \text{for } i, j = 1, 2, \ldots, n.$$

The matrix A is called the **matrix representation** of H with respect to the ordered basis β and is denoted by $\psi_\beta(H)$.

Theorem 6.32.
For any n-dimensional vector space V over $\mathbb{F}$ and any ordered basis β for V, $\psi_\beta : \mathcal{B}(V) \to M_{n \times n}(\mathbb{F})$ is an isomorphism.

Corollary 6.15.
For any n-dimensional vector space V, $\mathcal{B}(V)$ has dimension n^2.

Corollary 6.16.
Let V be an n-dimensional vector space over $\mathbb{F}$ with ordered basis β. If $H \in \mathcal{B}(V)$ and $A \in M_{n \times n}(\mathbb{F})$, then $\psi_\beta(H) = A$ if and only if $H(x, y) = [\phi_\beta(x)]^t A[\phi_\beta(y)]$ for all $x, y \in V$.

Corollary 6.17.
Let $\mathbb{F}$ be a field, n a positive integer, and β be the standard ordered basis for F^n. Then for any $H \in \mathcal{B}(F^n)$, there exists a unique square matrix $A \in M_{n \times n}(\mathbb{F})$, names, $A = \psi_\beta(H)$, such that

$$H(x, y) = x^t A y \quad \text{for all } x, y \in F^n.$$

Definition 6.46.
Let $A, B \in M_{n \times n}(\mathbb{F})$. Then B is said to be **congruent** to A if there exists an invertible matrix $Q \in M_{n \times n}(\mathbb{F})$ such that $B = Q^t A Q$.

Theorem 6.33.
Let V be a finite-dimensional vector space with ordered bases $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_n\}$, and let Q be the change of coordinate matrix changing γ-coordinates into β-coordinates. Then, for any $H \in \mathcal{B}(V)$, we have $\psi_\gamma(H) = Q^t \psi_\beta(H) Q$. Therefore $\psi_\gamma(H)$ is congruent to $\psi_\beta(H)$.

Corollary 6.18.
Let V be an n-dimensional vector space with ordered basis β, and let H be a bilinear form on V. For any $n \times n$ matrix B, if B is congruent to $\psi_\beta(H)$, then there exists an ordered basis γ for V such that $\psi_\gamma(H) = B$. Furthermore, if $B = Q^t \psi_\beta(H) Q$ for some invertible matrix Q, then Q changes γ-coordinates into β-coordinates.

Symmetric Bilinear Forms

Definition 6.47.
A bilinear form H on a vector space V is **symmetric** if $H(x, y) = H(y, x)$ for all $x, y \in V$.

Theorem 6.34.
Let H be a bilinear form on a finite-dimensional vector space V, and let β be an ordered basis for V. Then H is symmetric if and only if $\psi_\beta(H)$ is symmetric.

Definition 6.48.
A bilinear form H on a finite-dimensional vector space V is called **diagonalizable** if there is an ordered basis β for V such that $\psi_\beta(H)$ is a diagonal matrix.

Corollary 6.19.
Let H be a diagonalizable bilinear form on a finite-dimensional vector space V. Then H is symmetric.

Definition 6.49.
The **characteristic** of a field $\mathbb{F}$ is the smallest positive integer p for which a sum of p 1's equals 0. If no such positive integer exists, then $\mathbb{F}$ is said to have **characteristic zero**.

Lemma 6.9.
Let H be a nonzero symmetric bilinear form on a vector space V over a field $\mathbb{F}$ not of characteristic two. Then there is a vector x in V such that $H(x, x) \neq 0$.

Theorem 6.35.
Let V be a finite-dimensional vector space over a field $\mathbb{F}$ not of characteristic two. Then every symmetric bilinear form on V is diagonalizable.

Corollary 6.20.
Let $\mathbb{F}$ be a field that is not of characteristic two. If $A \in M_{n \times n}(\mathbb{F})$ is a symmetric matrix, then A is congruent to a diagonal matrix.

Quadratic Forms

Definition 6.50.
Let V be a vector space over $\mathbb{F}$. A function $K : V \to \mathbb{F}$ is called a **quadratic form** if there exists a symmetric bilinear form $H \in \mathcal{B}(V)$ such that

$$K(x) = H(x, x) \quad \text{for all } x \in V.$$

Quadratic Forms Over the Field $\mathbb{R}$

Theorem 6.36.
Let V be a finite-dimensional real inner product space, and let H be a symmetric bilinear form on V. Then there exists an orthonormal basis β for V such that $\psi_\beta(H)$ is a diagonal matrix.

Corollary 6.21.
Let K be a quadratic form on a finite-dimensional real inner product space V. There exists an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$ for V and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ (not necessarily distinct) such that if $x \in V$ and

$$x = \sum_{i=1}^{n} s_i v_i, \quad s_i \in \mathbb{R}$$

then

$$K(x) = \sum_{i=1}^{n} \lambda_i s_i^2.$$

In fact, if H is the symmetric bilinear form determined by K, then β can be chosen to be any orthonormal basis for V such that $\psi_\beta(H)$ is a diagonal matrix.

The Second Derivative Test for Functions of Several Variables

Definition 6.51.
Let $z = f(t_1, t_2, \ldots, t_n)$ be a fixed real-valued function of n real variables for which all third-order partial derivatives exist and are continuous. The function f is said to have a **local maximum** at point $p \in \mathbb{R}^n$ if there exists a $\delta > 0$ such that $f(p) \geq f(x)$ whenever $||x - p|| < \delta$. Likewise, f has a **local minimum** at $p \in \mathbb{R}^n$ if there exists a $\delta > 0$ such that $f(p) \leq f(x)$ whenever $||x - p|| < \delta$. If f has either a local minimum or a local maximum at p, we say that f has

a **local extremum** at p. A point $p \in \mathbb{R}^n$ is called a **critical point** of f if $\dfrac{\partial f(p)}{\partial (t_i)} = 0$ for $i = 1, 2, \ldots, n$. It is a well known fact that if f has a local extremum at a point $p \in \mathbb{R}^n$, then p is a critical point of f. For, if f has a local extremum at $p = (p_1, p_2, \ldots, p_n)$, then for any $i = 1, 2, \ldots, n$, the function ϕ_i defined by $\phi_i(t) = f(p_1, p_2, \ldots, p_{i-1}, t, p_{i+1}, \ldots, p_n)$ has a local extremum at $t = p_i$. So, by an elementary single-variable argument,

$$\frac{\partial f(p)}{\partial t_i} = \frac{d\phi_i(p_i)}{dt} = 0.$$

Thus p is a critical point of f. But critical points are not necessarily local extrema.

The second-order partial derivatives of f at a critical point p can often be used to test for local extremum at p. These partials determine a matrix $A(p)$ in which the row i, column j entry is

$$\frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)}.$$

This matrix is called the **Hessian matrix** of f at p. Note that if the third order partial derivatives of f are continuous, then the mixed second-order partials of f at p are independent of the order in which they are taken, and hence $A(p)$ is a symmetric matrix. In this case, all of the eigenvalues of $A(p)$ are real.

Theorem 6.37 (The Second Derivative Test).
Let $f(t_1, t_2, \ldots, t_n)$ be a real-valued function in n real variables for which all third-order partial derivatives exist and are continuous. Let $p = (p_1, p_2, \ldots, p_n)$ be a critical point of f, and let $A(p)$ be the Hessian of f at p.

1. If all eigenvalues of $A(p)$ are positive, then f has a local minimum at p.

2. If all eigenvalues of $A(p)$ are negative, then f has a local maximum at p.

3. If $A(p)$ has at least one positive and at least one negative eigenvalue, then f has no local extremum at p (P is called a **saddle-point** of f).

4. If $\operatorname{rank}(A(p)) < n$ and $A(p)$ does not have both positive and negative eigenvalues, then the second derivative test is inconclusive.

Sylvester's Law of Inertia

Definition 6.52.
The **rank** of a bilinear form is the rank of any of its matrix representations.

Theorem 6.38 (Sylvester's Law of Inertia). Let H be a symmetric bilinear form on a finite-dimensional real vector space V. Then the number of positive diagonal entries and the number of negative diagonal entries in any diagonal matrix representation of H are each independent of the diagonal representation.

Definition 6.53.
The number of positive diagonal entries in a diagonal representation of a symmetric bilinear form on a real vector space is called the **index** of the form. The difference between the number of positive and the number of negative diagonal entries in a diagonal representation of a symmetric bilinear form is called the **signature** of the form. The three terms *rank*, *index*, and *signature* are called **invariants** of the bilinear form because they are invariant with respect to matrix representations. These same terms apply to the associated quadratic form. Notice that the values of any two of these invariants determine the value of the third.

Corollary 6.22 (Sylvester's Law of Inertia for Matrices).
Let A be a real symmetric matrix. Then the number of positive diagonal entries and the number of negative diagonal entries in any diagonal matrix congruent to A is independent of the choice of the diagonal matrix.

Definition 6.54.
Let A be a real symmetric matrix, and let D be a diagonal matrix that is congruent to A. The number of positive diagonal entries of D is called the **index** of A. The difference between the number of positive diagonal entries and the number of negative diagonal entries of D is called the **signature** of A. As before, the *rank*, *index*, and *signature* of a matrix are called the **invariants** of the matrix, and the values of any two of these invariants determine the value of the third.

Corollary 6.23.
Two real symmetric $n \times n$ matrices are congruent if and only if they have the same invariants.

Corollary 6.24.
A real symmetric $n \times n$ matrix A has index p and rank r if and only if A is congruent to J_{pr}, defined by

$$J_{pr} = \begin{pmatrix} I_p & O & O \\ O & -I_{r-p} & O \\ O & O & O \end{pmatrix}$$

Definition 6.55.
For a 2×2 matrix M with columns x and y, the bilinear form $H(M) = H(x, y)$ is called the **permanent** of M.

9 Einstein's Special Theory of Relativity

Axioms of the Special Theory of Relativity

Definition 6.56.
The basic problem is to compare two different inertial (non-accelerating) coordinate systems S and S' in three-space ($\mathbb{R}^3$) that are in motion relative to each other under the assumption that the speed of light is the same when measured in either system. We assume that S' moves at a constant velocity in relation to S as measured from S. To simplify matters, let us suppose that the following conditions hold:

1. The corresponding axes of S and S' (x and x', y and y', z and z') are parallel, and the origin of S' moves in the positive direction of the x-axis of S at a constant velocity $v > 0$ relative to S.

2. Two clocks C and C' are placed in space - the first stationary relative to the coordinate system S and the second stationary relative to the coordinate system S'. These clocks are designed to give real numbers in units of seconds as readings. The clocks are calibrated so that at the instant the origins of S and S' coincide, both clocks give the reading zero.

3. The unit of length is the **light second** (the distance light travels in 1 second), and the unit of time is the second. Note that, with respect to these units, the speed of light is 1 light second per second.

Given any event (something whose position and time of occurrence can be described), we may assign a set of *space-time coordinates* to it. For example, if p is an event that occurs at position

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

relative to S and at time t as read on clock C, we can assign to p the set of coordinates

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}.$$

This ordered 4-tuple is called the **space-time coordinates** of p relative to S and C. Likewise, p has a set of space-time coordinates

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

relative to S' and C'.

For a fixed velocity v, let $T_v : \mathbb{R}^4 \to \mathbb{R}^4$ be the mapping defined by

$$T_v \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix},$$

where

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

are the space-time coordinates of the same event with respect to S and C and with respect to S' and C', respectively.

Einstein made certain assumptions about T_v that led to his special theory of relativity. We formulate an equivalent set of assumptions.

(R 1) The speed of any light beam, when measured in either coordinate system using a clock stationary relative to that coordinate system, is 1.

(R 2) The mapping $T_v : \mathbb{R}^4 \to \mathbb{R}^4$ is an isomorphism.

(R 3) If

$$T_v \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

then $y' = y$ and $z' = z$.

(R 4) If

$$T_v \begin{pmatrix} x \\ y_1 \\ z_1 \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} \quad \text{and} \quad T_v \begin{pmatrix} x \\ y_2 \\ z_2 \\ t \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ z'' \\ t'' \end{pmatrix}$$

then $x'' = x'$ and $t'' = t'$.

(R 5) The origin of S moves in the negative direction of the x'-axis of S' at the constant velocity $-v < 0$ as measured from S'.

Axioms (R 3) and (R 4) tell us that for $p \in \mathbb{R}^4$, the second and third coordinates of $T_v(p)$ are unchanged and the first and fourth coordinates of $T_v(p)$ are independent of the second and third coordinates of p.

These five axioms completely characterize T_v. The operator T_v is called the **Lorentz transformation** in direction x.

Theorem 6.39.
On $\mathbb{R}^4$, the following statements are true.

1. $T_v(e_i) = e_i$ for $i = 2, 3$.

2. span $(\{e_2, e_3\})$ is T_v-invariant.

3. span $(\{e_1, e_4\})$ is T_v-invariant.

4. Both span $(\{e_2, e_3\})$ and span $(\{e_1, e_4\})$ are T_v^*-invariant.

5. $T_v^*(e_i) = e_i$ for $i = 2, 3$.

Theorem 6.40.
If $\langle L_A(w), w \rangle = 0$ for some $w \in \mathbb{R}^4$, then

$$\langle T_v^* L_A T_V(w), w \rangle = 0.$$

Theorem 6.41.
There exist nonzero scalars a and b such that

1. $T_v^* L_A T_v(w_1) = aw_2$.

2. $T_v^* L_A T_v(w_2) = bw_1$.

Corollary 6.25.
Let $B_v = [T_v]_\beta$, where β is the standard ordered basis for $\mathbb{R}^4$. Then

1. $B_v^* A B_v = A$.

2. $T_v^* L_A T_v = L_A$.

Theorem 6.42.
Let β be the standard ordered basis for $\mathbb{R}^4$. Then

$$[T_V]_\beta = B_v = \begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & 0 & 0 & \frac{-v}{\sqrt{1-v^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{-v}{\sqrt{1-v^2}} & 0 & 0 & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}$$

10   Conditioning and the Rayleigh Quotient

Definition 6.57.
For a system of linear equations $Ax = b$, two types of errors must be considered for real-world application. First, experimental errors arise in the collection of data since no instruments can provide completely accurate measurements. Second, computers introduce round-off errors. One might intuitively feel that small relative changes in the coefficients of the system cause small relative errors in the solution. A system that has this property is called **well-conditioned**; otherwise, the system is called **ill-conditioned**.

Notation 6.3.
Let $Ax = b$ be a system of linear equations. We use the notation δb to denote the vector $b' - b$, where b is the vector in the original system, and b' is the vector in the modified system.

Definition 6.58.
Let $Ax = b$ be a system of linear equations. We define the **relative change** in a vector b to be the scalar $||\delta b||/||b||$, where $||\cdot||$ denotes the standard norm on $\mathbb{C}^n$ (or $\mathbb{R}^n$); that is, $||b|| = \sqrt{\langle b, b \rangle}$. Similar definitions hold for the **relative change** of x.

Definition 6.59. Let A be a complex (or real) $n \times n$ matrix. Define the **(Euclidean) norm** of A by

$$||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||},$$

where $x \in \mathbb{C}^n$ or $x \in \mathbb{R}^n$.

Intuitively, $||A||$ represents the maximum *magnification* of a vector by the matrix A.

Definition 6.60. Let B be an $n \times n$ self-adjoint matrix. The **Rayleigh quotient** for $x \neq 0$ is defined to be the scalar $R(x) = \langle Bx, x \rangle / ||x||^2$.

Theorem 6.43.
For a self-adjoint matrix $B \in M_{n \times n}(\mathbb{F})$, we have that $\max\limits_{x \neq 0} R(x)$ is the largest eigenvalue of B and $\min\limits_{x \neq 0} R(x)$ is the smallest eigenvalue of B.

Corollary 6.26.
For any square matrix A, $||A||$ is finite and, in fact, equals $\sqrt{\lambda}$, where λ is the largest eigenvalue of A^*A.

Lemma 6.10.
For any square matrix A, λ is an eigenvalue of A^*A if and only if λ is an eigenvalue of AA^*.

Corollary 6.27.
Let A be an invertible matrix. Then $||A^{-1}|| = 1/\sqrt{\lambda}$, where λ is the smallest eigenvalue of A^*A.

Definition 6.61.
Let A be an invertible matrix. The number $||A|| \cdot ||A^{-1}||$ is called the **condition number** of A and is denoted $\text{cond}(A)$.

Theorem 6.44.
For the system $Ax = b$, where A is invertible and $b \neq 0$, the following statements are true.

1. For any norm, $|| \cdot ||$, we have $\dfrac{1}{\text{cond}(A)} \dfrac{||\delta b||}{||b||} \leq \dfrac{||\delta x||}{||x||} \leq \text{cond}(A) \dfrac{||\delta b||}{||b||}$.

2. If $|| \cdot ||$ is the Euclidean norm, then $\text{cond}(A) = \sqrt{\lambda_1 / \lambda_n}$, where λ_1 and λ_n are the largest and smallest eigenvalues, respectively, of A^*A.

11 The Geometry of Orthogonal Operators

Definition 6.62.
Let T be a linear operator on a finite-dimensional real inner product space V. The operator T is called a **rotation** if T is the identity on V or if there exists a two-dimensional subspace W of V, and orthonormal basis $\beta = \{x_1, x_2\}$ for W, and a real number θ such that

$$T(x_1) = (\cos(\theta))x_1 + (\sin(\theta))x_2, \quad T(x_2) = (-\sin(\theta))x_1 + (\cos(\theta))x_2,$$

and $T(y) = y$ for all $y \in W^\perp$. In this context, T is called a **rotation of W about $W^\perp$**. The subspace $W^\perp$ is called the **axis of rotation**.

Definition 6.63.

Let T be a linear operator on a finite-dimensional real inner product space V. The operator T is called a **reflection** if there exists a one-dimensional subspace W of V such that $T(x) = -x$ for all $x \in W$ and $T(y) = y$ for all $y \in W^\perp$. In this context, T is called a **reflection of V about $W^\perp$**.

Theorem 6.45.

Let T be an orthogonal operator on a two-dimensional real inner product space V. Then T is either a rotation or a reflection. Furthermore, T is a rotation if and only if $\det(T) = 1$, and T is a reflection if and only if $\det(T) = -1$.

Corollary 6.28.

Let V be a two-dimensional real inner product space. The composite of a reflection and a rotation on V is a reflection on V.

Lemma 6.11.

If T is a linear operator on a nonzero finite-dimensional real vector space V, then there exists a T-invariant subspace W of V such that $1 \leq \dim(W) \leq 2$.

Theorem 6.46.

Let T be an orthogonal operator on a nonzero finite-dimensional real inner product space V. Then there exists a collection of pairwise orthogonal T-invariant subspaces $\{W_1, W_2, \ldots, W_m\}$ of V such that

1. $1 \leq \dim(() W_i) \leq 2$ for $i = 1, 2, \ldots, m$.

2. $V = W_1 \oplus W_2 \oplus \cdots \oplus W_m$.

Theorem 6.47.

Let $T, V, W_1, \ldots, W_m$ be as in Theorem 6.46.

1. The number of W_i's for which T_{W_i} is a reflection is even or odd according to whether $\det(T) = 1$ or $\det(T) = -1$.

2. It is always possible to decompose V as in Theorem 6.46 so that the number of W_i's for which T_{W_i} is a reflection is zero or one according to whether $\det(T) = 1$ or $\det(T) = -1$. Furthermore, if T_{W_i} is a reflection, then $\dim(W_i) = 1$.

Corollary 6.29.

Let T be an orthogonal operator on a finite-dimensional real inner product space V. Then there exists a collection $\{T_1, T_2, \ldots, T_m\}$ of orthogonal operators on V such that the following statements are true.

1. For each i, T_i is either a reflection or a rotation.

2. For at most one i, T_i is a reflection.

3. $T_i T_j = T_j T_i$ for all i and j.

4. $T = T_1 T_2 \ldots T_m$.

5. $\det(T) = \begin{cases} 1 & \text{if } T_i \text{ is a rotation for each } i \\ -1 & \text{otherwise} \end{cases}$

Chapter 7

Canonical Forms

1 The Jordan Canonical Form I

Definition 7.1. In this section, we extend the definition of eigenspace to *generalized eigenspace*. From these subspaces, we select ordered bases whose union is an ordered basis β for V such that

$$[T]_\beta = \begin{pmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \vdots & \vdots & & \vdots \\ O & O & \ldots & A_k \end{pmatrix}$$

where each O is a zero matrix, and each A_i is a square matrix of the form (λ) or

$$\begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda & 1 \\ 0 & 0 & 0 & \ldots & 0 & \lambda \end{pmatrix}$$

for some eigenvalue λ of T. Such a matrix A_i is called a **Jordan block** corresponding to λ, and the matrix $[T]_\beta$ is called a **Jordan canonical form** of T. We also say that the ordered basis β is a **Jordan canonical basis** for T. Observe that each Jordan block A_i is "almost" a diagonal matrix – in fact, $[T]_\beta$ is a diagonal matrix if and only if each A_i is of the form (λ).

Definition 7.2.
Let T be a linear operator on a vector space V, and let λ be a scalar. A nonzero vector x in V is called a **generalized eigenvector of T corresponding to** λ if $(T - \lambda I)^p(x) = 0$ for some positive integer p.

Definition 7.3.
Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. The **generalized eigenspace of T corresponding to** λ, denoted K_λ, is the subset of V defined by

$$K_\lambda = \{x \in V : (T - \lambda I)^p(x) = 0 \text{ for some positive integer } p\}.$$

Theorem 7.1.

Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. then

1. K_λ is a T-invariant subspace of V containing E_λ (the eigenspace of T corresponding to λ).

2. for any scalar $\mu \neq \lambda$, the restriction of $T - \mu I$ to K_λ is one-to-one.

Theorem 7.2.

Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Suppose that λ is an eigenvalue of T with multiplicity m. Then

1. $\dim(K_\lambda) \leq m$.

2. $K_\lambda = N\left((T - \lambda I)^m\right)$.

Theorem 7.3.

Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T. Then, for every $x \in V$, there exist vectors $v_i \in K_\lambda$, $1 \leq i \leq k$, such that

$$x = v_1 + v_2 + \cdots + v_k.$$

Theorem 7.4.

Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of T with corresponding multiplicities $m_1, m_2, \ldots, m_k$. For $1 \leq i \leq k$, let β_i be an ordered basis for K_{λ_i}. Then the following statements are true.

1. $\beta_i \cap \beta_j = \emptyset$ for $i \neq j$.

2. $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an ordered basis for V.

3. $\dim(K_{\lambda_i}) = m_i$ for all i.

Corollary 7.1.

Let T be a linear operator on a finite-dimensional vector space V such that the characteristic polynomial of T splits. Then T is diagonalizable if and only if $E_\lambda = K_\lambda$ for every eigenvalue λ of T.

Definition 7.4.

Let T be a linear operator on a vector space V, and let x be a generalized eigenvector of T corresponding to the eigenvalue λ. Suppose that p is the smallest positive integer for which $(T - \lambda I)^p(x) = 0$. Then the ordered set

$$\{(T - \lambda I)^{p-1}(x), (T - \lambda I)^{p-2}(x), \ldots, (T - \lambda I)(x), x\}$$

is called a **cycle of generalized eigenvectors** of T corresponding to λ. The vectors $(T - \lambda I)^{p-1}(x)$ and x are called the **initial vector** and the **end vector** of the cycle, respectively. We say that the **length** of the cycle is p.

Theorem 7.5.
Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits, and suppose that β i a basis for V such that β is a disjoint union of cycles of generalized eigenvectors of T. Then the following statements are true.

1. For each cycle γ of generalized eigenvectors contained in β, $W = \text{span}(\gamma)$ is T-invariant, and $[T_W]_\gamma$ is a Jordan block.

2. β is a Jordan canonical basis for V.

Theorem 7.6.
Let T be a linear operator on a vector space V, and let λ be an eigenvalue of T. Suppose that $\gamma_1, \gamma_2, \ldots, \gamma_q$ are cycles of generalized eigenvectors of T corresponding to λ such that the initial vectors of the γ_i's are distinct and form a linearly independent set. Then the γ_i's are disjoint, and their union $\gamma = \bigcup_{i=1}^{q} \gamma_i$ is linearly independent.

Corollary 7.2.
Every cycle of generalized eigenvectors of a linear operator is linearly independent.

Theorem 7.7.
Let T be a linear operator on a finite-dimensional vector space V, and let λ be an eigenvalue of T. Then K_λ has an ordered basis consisting of a union of disjoint cycles of generalized eigenvectors corresponding to λ.

Corollary 7.3.
Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits. then T has Jordan canonical form.

Definition 7.5.
Let $A \in M_{n \times n}(\mathbb{F})$ be such that the characteristic polynomial of A (and hence of L_A) splits. Then the **Jordan canonical form** of A is defined to be the Jordan canonical form of the linear operator L_A on $\mathbb{F}^n$.

Corollary 7.4.
Let A be an $n \times n$ matrix whose characteristic polynomial splits. Then A has Jordan canonical form J, and A is similar to J.

Theorem 7.8.
Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits. Then V is the direct sum of the generalized eigenspaces of T.

2 The Jordan Canonical Form II

Definition 7.6.
For the purposes of this section, we fix a linear operator T on an n-dimensional vector space V such that the characteristic polynomial of T splits. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues

of T.

By Theorem 7.7, each generalized eigenspace K_{λ_i} contains an ordered basis β_i consisting of a union of disjoint cycles of generalized eigenvectors corresponding to λ_i. So by Theorem 7.4(2) and Theorem 7.5, the union $\beta = \bigcup_{i=1}^{k} \beta_i$ is a Jordan canonical basis for T. For each i, let T_i be the restriction of T to K_{λ_i}, and let $A_i = [T_i]_{\beta_i}$. Then A_i is the Jordan canonical form of T_{ij}, and

$$J = [T]_\beta = \begin{pmatrix} A_1 & O & \dots & O \\ O & A_2 & \dots & O \\ \vdots & \vdots & & \vdots \\ O & O & \dots & A_k \end{pmatrix}$$

is the Jordan canonical form of T. In this matrix, each O is a zero matrix of appropriate size.

Note: In this section, we compute the matrices A_i and the bases β_i, thereby computing J and β as well. To aid in formulating the uniqueness theorem for J, we adopt the following convention: The basis β_i for K_λ will henceforth be ordered in such a way that the cycles appear in order of decreasing length. That is, if β_i is a disjoint union of cycles $\gamma_1, \gamma_2, \dots, \gamma_{n_i}$ and if the length of the cycle γ_j is p_j, we index the cycles so that $p_1 \geq p_2 \geq \dots \geq p_{n_i}$.

To illustrate the discussion above, suppose that, for some i, the ordered basis β_i for K_{λ_i} is the union of four cycles $\beta_i = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$ with respective lengths $p_1 = 3, p_2 = 3, p_3 = 2$, and $p_4 = 1$. Then

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

To help us visualize each of the matrices A_i and ordered bases β_i, we use an array of dots called a **dot diagram** of T_i, where T_i is the restriction of T to K_{λ_i}. Suppose that β_i is a disjoint union of cycles of generalized eigenvectors $\gamma_1, \gamma_2, \dots, \gamma_{n_i}$ with lengths $p_1 \geq p_2 \geq \dots \geq p_{n_i}$, respectively. The dot diagram of T_i contains one dot for each vector in β_i, and the dots are configured according to the following rules.

1. The array consists of n_i columns (one column for each cycle).

2. Counting from left to right, the jth column consists of the p_j dots that correspond to the vectors of γ_j starting with the initial vector at the top and continuing down to the end vector.

Denote the end vectors of the cycles by $v_1, v_2, \ldots, v_{n_i}$. In the following dot diagram of T_i, each dot is labeled with the name of the vector in β_i to which it corresponds.

$$
\begin{array}{llll}
\bullet(T-\lambda_i I)^{p_1-1}(v_1) & \bullet(T-\lambda_i I)^{p_2-1}(v_2) & \cdots & \bullet(T-\lambda_i I)^{p_{n_i}-1}(v_{n_i}) \\
\bullet(T-\lambda_i I)^{p_1-2}(v_1) & \bullet(T-\lambda_i I)^{p2-2}(v_2) & \cdots & \bullet(T-\lambda_i I)^{p_{n_i}-2}(v_{n_i}) \\
\vdots & \vdots & & \vdots \\
& & & \bullet(T-\lambda_i I)(v_{n_i}) \\
& & & \bullet(v_{n_i}) \\
& \bullet(T-\lambda_i I)(v_2) & & \\
& \bullet v_2 & & \\
\bullet(T-\lambda_i I)(v_1) & & & \\
\bullet v_1 & & &
\end{array}
$$

Notice that the dot diagram of T_i has n_i columns (one for each cycle) and p_1 rows. Since $p_1 \geq p_2 \geq \cdots \geq p_{n_i}$, the columns of the dot diagram become shorter (or at least not longer) as we move from left to right.

Now let r_j denote the number of dots in the jth row of the dot diagram. Observe that $r_1 \geq r_2 \geq \cdots \geq r_{p_1}$. Furthermore, the diagram can be reconstructed from the values of the r_i's.

In the above example, with $n_i = 4$, $p_1 = p_2 = 3$, $p_3 = 2$, and $p_4 = 1$, the dot diagram of T_i is as follows:

$$
\begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \\
\bullet & \bullet & &
\end{array}
$$

Here $r_1 = 4$, $r_2 = 3$ and $r_3 = 2$.

We now devise a method for computing the dot diagram of T_i using the ranks of linear operators determined by T and λ_i. Hence the dot diagram is completely determined by T, from which it follows that it is unique. On the other hand β_i is not unique.

To determine the dot diagram of T_i, we devise a method for computing each r_j, the number of dots in the jth row of the dot diagram, using only T_i and λ_i. The next three result give us the required method. To facilitate our arguments, we fix a basis β_i for K_{λ_i} so that β_i is a disjoint union of n_i cycles of generalized eigenvectors with lengths $p_1 \geq p_2 \geq \cdots \geq p_{n_i}$.

Theorem 7.9.
For any positive integer r, the vectors in β_i that are associated with the dots in the first r rows of the dot diagram of T_i constitute a basis for $\mathrm{N}\left((T-\lambda_i I)^r\right)$. Hence the number of dots in the first r rows of the dot diagram equals nullity $\left((T-\lambda_i I)^r\right)$.

Corollary 7.5.
The dimension of E_{λ_i} is n_i. Hence in a Jordan canonical form of T, the number of Jordan blocks corresponding to λ_i equals the dimension of E_{λ_i}.

Theorem 7.10.
Let r_j denote the number of dots in the jth row of the dot diagram of T_i, the restriction of T to K_{λ_i}. Then the following statements are true.

1. $r_1 = \dim(V) - \mathrm{rank}\,(T - \lambda_i I)$.

2. $r_j = \mathrm{rank}\,((T - \lambda_i I)^{j-1}) - \mathrm{rank}\,((T - \lambda_i I)^j)$ if $j > 1$.

Corollary 7.6.
For any eigenvalue λ_i of T, the dot diagram of T_i is unique. Thus, subject to the convention that cycles of generalized eigenvectors for the bases of each generalized eigenspace are listed in order of decreasing length, the Jordan canonical form of a linear operator or a matrix is unique up to the ordering of the eigenvalues.

Theorem 7.11.
Let A and B be $n \times n$ matrices, each having Jordan canonical forms computed according to the conventions of this section. Then A and B are similar if and only if they have (up to an ordering of their eigenvalues) the same Jordan canonical form.

Lemma 7.1. A linear operator T on a finite-dimensional vector space V is diagonalizable if and only if its Jordan canonical form is a diagonal matrix. Hence T is diagonalizable if and only if the Jordan canonical basis for T consists of eigenvectors of T.

Definition 7.7.
A linear operator T on a vector space V is called **nilpotent** if $T^p = T_0$ for some positive integer p. An $n \times n$ matrix A is called **nilpotent** if $A^p = O$ for some positive integer p.

Definition 7.8.
For any $A \in M_{n \times n}(\mathbb{C})$, define the norm of A by

$$||A|| = \max\{|A_{ij}| : 1 \le i, j \le n\}.$$

3 The Minimal Polynomial

Definition 7.9.
A polynomial $f(x)$ with coefficients from a field $\mathbb{F}$ is called **monic** if its leading coefficient is 1. If $f(x)$ has positive degree and cannot be expressed as a product of polynomials with coefficients from $\mathbb{F}$ each having positive degree, then $f(x)$ is called **irreducible**.

Definition 7.10. Let T be a linear operator on a finite-dimensional vector space. A polynomial $p(t)$ is called a **minimal polynomial** of T if $p(t)$ is a monic polynomial of least positive degree for which $p(T) = T_0$.

Theorem 7.12.
Let $p(t)$ be a minimal polynomial of a linear operator T on a finite-dimensional vector space V.

1. For any polynomial $g(t)$, if $g(T) = T_0$, then $p(t)$ divides $g(t)$. In particular, $p(t)$ divides the characteristic polynomial of T.

2. The minimal polynomial T is unique.

Definition 7.11.
Let $A \in M_{n \times n}(\mathbb{F})$. The **minimal polynomial** $p(t)$ of A is the monic polynomial of least positive degree for which $p(A) = O$.

Theorem 7.13.
Let T be a linear operator on a finite-dimensional vector space V, and let β be an ordered basis for V. Then the minimal polynomial of T is the same as the minimal polynomial of $[T]_\beta$.

Corollary 7.7.
For any $A \in M_{n \times n}(\mathbb{F})$, the minimal polynomial of A is the same as the minimal polynomial of L_A.

Theorem 7.14.
Let T be a linear operator on a finite-dimensional vector space V, and let $p(t)$ be the minimal polynomial of T. A scalar λ is an eigenvalue of T if and only if $p(\lambda) = 0$. Hence the characteristic polynomial and the minimal polynomial of T have the same zeros.

Corollary 7.8.
Let T be a linear operator on a finite-dimensional vector space V with minimal polynomial $p(t)$ and characteristic polynomial $f(T)$. Suppose that $f(t)$ factors as

$$f(t) = (\lambda_1 - f)^{n_1}(\lambda_2 - t)^{n_2} \ldots (\lambda_k - t)^{n_k},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of T. Then there exist integers $m_1, m_2, \ldots, m_k$ such that $1 \le m_i \le n_i$ for all i and

$$p(t) = (t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \ldots (t - \lambda_k)^{m_k}.$$

Theorem 7.15.
Let T be a linear operator on an n-dimensional vector space V such that V is a T-cyclic subspace of itself. Then the characteristic polynomial $f(t)$ and the minimal polynomial $p(t)$ have the same degree, and hence $f(t) = (-1)^n p(t)$.

Theorem 7.16.
Let T be a linear operator on a finite-dimensional vector space V. Then T is diagonalizable if and only if the minimal polynomial of T is of the form

$$p(t) = (t - \lambda_1)(t - \lambda_2) \ldots (t - \lambda_k),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of T.

Definition 7.12.
Let T be a linear operator on a finite-dimensional vector space V, and let x be a nonzero vector in V. The polynomial $p(t)$ is called a T-**annihilator** of x if $p(t)$ is a monic polynomial of lest degree for which $p(T)(x) = 0$.

4   The Rational Canonical Form

Definition 7.13.
Let T be a linear operator on a finite-dimensional vector space V with characteristic polynomial

$$f(t) = (-1)^n (\phi_1(t))^{n_1} (\phi_2(t))^{n_2} \ldots (\phi_k(t))^{n_k},$$

where the $\phi_i(t)$'s ()$1 \leq i \leq k$) are distinct irreducible monic polynomials and the n_i's are positive integers. For $1 \leq i \leq k$, we define the subset K_{ϕ_i} of V by

$$K_{\phi_i} = \{x \in V : (\phi_i(T))^p(x) = 0 \text{ for some positive integer } p\}.$$

Definition 7.14.
Let T be a linear operator on a finite-dimensional vector space V, and let x be a nonzero vector in V. We use the notation C_x for the T-cyclic subspace generated by x. Recall Theorem 5.22, that if $\dim(\mathsf{C}_x) = k$, then the set

$$\{x, T(x), T^2(x), \ldots, T^{k-1}(x)\}$$

is an ordered basis for C_x. To distinguish this basis from all other ordered bases for C_x, we call it the T-**cyclic basis generated by** x and denote it by β_x. Let A be the matrix representation of the restriction of T to C_x relative to the ordered basis β_x. Recall that

$$A = \begin{pmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_{k-1} \end{pmatrix}$$

where

$$a_0 x + a_1 T(x) + \cdots + a_{k-1} T^{k-1}(x) + T^k(x) = 0.$$

furthermore, the characteristic polynomial of A is given by

$$\det(A - tI) = (-1)^k (a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + t^k).$$

The matrix A is called the **companion matrix** of the monic polynomial $h(t) = a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + t^k$. Every monic polynomial has a companion matrix, and the characteristic polynomial of the companion matrix of a monic polynomial $g(t)$ of degree k is equal to $(-1)^k g(T)$. By Theorem 7.15, the monic polynomial $h(t)$ is also the minimal polynomial of A. Since A is the matrix representation of the restriction of T to C_x, $h(t)$ is also the minimal polynomial of this restriction. Note that $h(t)$ is also the T-annihilator of x.

It is the object of this section to prove that for every linear operator T on a finite-dimensional vector space V, there exists an ordered basis β for V such that the matrix representation $[T]_\beta$ is of the form

$$\begin{pmatrix} C_1 & O & \ldots & O \\ O & C_2 & \ldots & O \\ \vdots & \vdots & & \vdots \\ O & O & \ldots & C_r \end{pmatrix},$$

where each C_i is the companion matrix of a polynomial $(\phi(t))^m$ such that $phi(t)$ is a monic irreducible divisor of the characteristic polynomial of T and m is a positive integer. A matrix representation of this kind is called a **rational canonical form** of T. We call the accompanying basis a **rational canonical basis** for T.

Lemma 7.2.
Let T be a linear operator on a finite-dimensional vector space V, let x be a nonzero vector in V, and suppose that the T-annihilator of x is of the form $(\phi(t))^p$ for some irreducible monic polynomial $\phi(t)$. Then $\phi(t)$ divides the minimal polynomial of T, and $x \in K_\phi$.

Theorem 7.17.
Let T be a linear operator on a finite-dimensional vector space V, and let β be an ordered basis for V. Then β is a rational canonical basis for T if and only if β is the disjoint union of T-cyclic bases β_{v_i}, where each v_i lies in K_ϕ for some irreducible monic divisor $\phi(t)$ of the characteristic polynomial of T.

Theorem 7.18.
Let T be a linear operator on a finite-dimensional vector space V, and suppose that

$$p(t) = (\phi_1(t))^{m_1}(\phi_2(t))^{m_2}\ldots(\phi_k(t))^{m_k}$$

is the minimal polynomial of T, where the $\phi_i(t)$'s $(1 \leq i \leq k)$ are the distinct irreducible monic factors of $p(t)$ and the m_i's are the positive integers. Then the following statements are true.

1. K_{ϕ_i} is a nonzero T-invariant subspace of V for each i.

2. If x is a nonzero vector in some K_{ϕ_i}, then the T-annihilator of x is of the form $(\phi_i(t))^p$ for some integer p.

3. $K_{\phi_i} \cap K_{\phi_j} = \{0\}$ for $i \neq j$.

4. K_{ϕ_i} is invariant under $\phi_j(T)$ for $i \neq j$, and the restriction of $\phi_j(T)$ to K_{ϕ_i} is one-to-one and onto.

5. $K_{\phi_i} = N\left((\phi_i(T))^{m_i}\right)$ for each i.

Lemma 7.3.
Let T be a linear operator on a finite-dimensional vector space V, and suppose that

$$p(t) = (\phi_1(t))^{m_1}(\phi_2(t))^{m_2}\ldots(\phi_k(t))^{m_k}$$

is the minimal polynomial of T, where the ϕ_i's $(1 \leq i \leq k)$ are the distinct irreducible monic factors of $p(t)$ and the m_i's are the positive integers. For $1 \leq i \leq k$, let $v_i \in K_{\phi_i}$ be such that

$$v_1 + v_2 + \cdots + v_k = 0.$$

Then $v_i = 0$ for all i.

Theorem 7.19.

Let T be a linear operator on a finite-dimensional vector space V, and suppose that

$$p(t) = (\phi_1(t))^{m_1} (\phi_2(t))^{m_2} \ldots (\phi_k(t))^{m_k}$$

is the minimal polynomial of T, where the ϕ_i's $(1 \leq i \leq k)$ are the distinct irreducible monic factors of $p(t)$ and the m_i's are the positive integers. For $1 \leq i \leq k$, let S_i be a linearly independent subset of K_{ϕ_i}. Then

1. $S_i \cap S_j = \emptyset$ for $i \neq j$.

2. $S_1 \cup S_2 \cup \cdots \cup S_k$ is linearly independent.

Theorem 7.20.

Let $v_1, v_2, \ldots, v_k$ be distinct vectors in K_ϕ such that

$$S_1 = \beta_{v_1} \cup \beta_{v_2} \cup \cdots \cup \beta_{v_k}$$

is linearly independent. For each i, choose $w_i \in V$ such that $\phi(T)(w_i) = v_i$. Then

$$S_2 = \beta_{w_1} \cup \beta_{w_2} \cup \cdots \cup \beta_{w_k}$$

is also linearly independent.

Lemma 7.4.

Let W be a T-invariant subspace of K_ϕ, and let β be a basis for W. Then the following statements are true.

1. Suppose that $x \in N(\phi(T))$, but $x \notin W$. Then $\beta \cup \beta_x$ is linearly independent.

2. For some $w_1, w_2, \ldots, w_s$ in $N(\phi(T))$, β can be extended to the linearly independent set

$$\beta' = \beta \cup \beta_{w_1} \cup \beta_{w_2} \cup \cdots \cup \beta_{w_s},$$

whose span contains $N(\phi(T))$.

Theorem 7.21.

If the minimal polynomial of T is of the form $p(t) = (\phi(t))^m$, then there exists a rational canonical basis for T.

Corollary 7.9.

K_ϕ has a basis consisting of the union of T-cyclic bases.

Theorem 7.22.

Every linear operator on a finite-dimensional vector space has a rational canonical basis and, hence, a rational canonical form.

Theorem 7.23.
Let T be a linear operator on an n-dimensional vector space V with characteristic polynomial

$$f(t) = (-1)^n (\phi_1(t))^{n_1} (\phi_2(t))^{n_2} \dots (\phi_k(t))^{n_k},$$

where the $\phi_i(t)$'s $(1 \leq i \leq k)$ are distinct irreducible monic polynomials and the n_i's are positive integers. Then the following statements are true.

1. $\phi_1(t), \phi_2(t), \dots, \phi_k(t)$ are the irreducible monic factors of the minimal polynomial.

2. For each i, $\dim(K_{\phi_i}) = d_i n_i$, where d_i is the degree of $\phi_i(t)$.

3. If β is a rational canonical basis for T, then $\beta_i = \beta \cap K_{\phi_i}$ is a basis for K_{ϕ_i} for each i.

4. If γ_i is a basis for K_{ϕ_i} for each i, then $\gamma = \gamma_1 \cup \gamma_2 \cup \dots \cup \gamma_k$ is a basis for V. In particular, if each γ_i is a disjoint union of T-cyclic bases, then γ is a rational canonical basis for T.

Definition 7.15.
Let β be a rational canonical basis for T, and $\beta_{v_1}, \beta_{v_2}, \dots, \beta_{v_k}$ be the T-cyclic bases of β that are contained in K_ϕ. Consider these T-cyclic bases β_{v_i}, and suppose again that the T-annihilator of v_j is $(\phi(t))^{p_j}$. Then β_{v_j} consists of dp_j vectors in β, where d is the degree of the polynomial. For $0 \leq i < d$, let γ_i be the cycle of generalized eigenvectors of U corresponding to $\lambda = 0$ with end vector $T^i(v_j)$, where $T^0(v_j) = b_j$. Then

$$\gamma_i = \{(\phi(T))^{p_j-1} T^i(v_j), (\phi(T))^{p_j-2} T^i(v_j), \dots, (\phi(T)) T^i(v_j), T^i(v_j)\}.$$

By Theorem 7.1, γ_i is a linearly independent subset of C_{v_i}. Now let

$$\alpha_j = \gamma_0 \cup \gamma_1 \cup \dots \cup \gamma_{d-1}.$$

Notice that α_j contains $p_j d$ vectors.

Uniqueness of the Rational Canonical Form

Lemma 7.5.
α_j is an ordered basis for C_{v_j}.

Lemma 7.6.
α is a Jordan canonical basis for K_ϕ.

Theorem 7.24.
Let T be a linear operator on a finite-dimensional vector space V, let $\phi(t)$ be an irreducible monic divisor of the characteristic polynomial of T of degree d, and let r_i denote the number of dots in the ith row of the dot diagram for $\phi(t)$ with respect to a rational canonical basis for T. Then

1. $r_1 = \dfrac{1}{d}[\dim(V) - \text{rank}(\phi(T))]$

2. $r_i = \dfrac{1}{d}[\text{rank}((\phi(T))^{i-1}) - \text{rank}((\phi(t))^i)]$ for $i > 1$.

Corollary 7.10.
Under the conventions described earlier, the rational canonical form of a linear operator is unique up to the arrangement of the irreducible monic divisors of the characteristic polynomial.

Definition 7.16.
Since the rational canonical form of a linear operator is unique, the polynomials corresponding to the companion matrices that determine this form are also unique. These polynomials, which are powers of the irreducible monic divisors, are called the **elementary divisors** of the linear operator. Since a companion matrix may occur more than once in a rational canonical form, the same is true for the elementary divisors. We call the number of such occurrences the **multiplicity** of the elementary divisor.

Conversely, the elementary divisors and their multiplicities determine the companion matrices and, therefore, the rational canonical form of a linear operator.

Definition 7.17.
Let $A \in M_{n \times n}(\mathbb{F})$. The **rational canonical form** of A is defined to be the rational canonical form of L_A. Likewise, for A, the **elementary divisors** and their **multiplicities** are the same as those of L_A.

Direct Sums

Theorem 7.25 (Primary Decomposition Theorem).
Let T be a linear operator on an n-dimensional vector space V with characteristic polynomial

$$f(t) = (-1)^n (\phi_1(t))^{n_1} (\phi_2(t))^{n_2} \ldots (\phi_k(t))^{n_k},$$

where the $\phi_i(t)$'s ($1 \leq i \leq k$) are distinct irreducible monic polynomials and the n_i's are positive integers. Then the following statements are true.

1. $V = K_{\phi_i} \oplus K_{\phi_2} \oplus \cdots \oplus K_{\phi_k}$.

2. If T_i ($1 \leq i \leq k$) is the restriction of T to K_{ϕ_i} and C_i is the rational canonical form of T_i, then $C_1 \oplus C_2 \oplus \cdots \oplus C_k$ is the rational canonical form of T.

Theorem 7.26.
Let T be a linear operator on a finite-dimensional vector space V. Then V is a direct sum of T-cyclic subspaces C_{v_i}, where each v_i lies in K_ϕ for some irreducible monic divisor $\phi(t)$ of the characteristic polynomial of T.

Chapter 8

Appendices

A   Sets

Definition 8.1.
A **set** is a collection of objects, called **elements** of the set. If x is an element of the set A, then we write $x \in A$; otherwise, we write $x \notin A$.

Definition 8.2.
Two sets A and B are called **equal**, written $A = B$, if they contain exactly the same elements.

Remark 8.1.
Sets may be described in one of two ways

1. By listing the elements of the set between set braces { }.

2. By describing the elements of the set in terms of some characteristic property.

Definition 8.3.
A set B is called a **subset** of A, written $B \supseteq A$ or $A \supseteq B$, if every element of B is an element of A. If $B \subseteq A$, and $B \neq A$, then B is called a **proper subset** of A.

Definition 8.4.
The **empty set**, denoted by $\emptyset$, is the set containing no elements. The empty set is a subset of every set.

Definition 8.5.
The **union** of two sets A and B, denoted $A \cup B$, is the set of elements that are in A, or B, or both; that is,

$$A \cup B = \{x : x \in A \text{ or } B\}.$$

Definition 8.6.
The **intersection** of two sets A and B, denoted $A \cap B$, is the set of elements that are in both A and B; that is,

$$A \cap B = \{x : x \in A \text{ and } B\}.$$

Definition 8.7.
Two sets are called **disjoint** if their intersection equals the empty set.

Definition 8.8.
A **relation** on a set A is a set S of ordered pairs of elements of A such that $(x, y) \in S$ if and only if x stands in the given relationship to y.

Definition 8.9.
A relation S on a set A is called an **equivalence relation** on A if the following three conditions hold:

1. For each $x \in A$, $x \sim x$ (reflexivity).

2. If $x \sim y$, then $y \sim x$ (symmetry).

3. If $x \sim y$ and $y \sim z$, then $x \sim z$ (transitivity).

B   Functions

Definition 8.10.
If A and B are sets, then a **function** f from A to B, written $f : A \to B$, is a rule that associates to each element of x in A a unique element denoted $f(x)$ in B.

The element $f(x)$ is called the **image** of x (under f), and x is called a **preimage** of $f(x)$ (under f).

If $f : A \to B$, then A is called the **domain** of f, B is called the **codomain** of f, and the set $\{f(x) : x \in A\}$ is called the **range** of f.

Two functions $f : A \to B$ and $g : A \to B$ are **equal**, written $f = g$, if $f(x) = g(x)$ for all $x \in A$.

Definition 8.11.
Functions such that each element of the range has a unique preimage are called **one-to-one**; that is, $f : A \to B$ is one-to-one if $f(x) = f(y)$ implies $x = y$ or, equivalently, if $x \neq y$ implies $f(x) \neq f(y)$.

If $f : A \to B$ is a function with range B, that is, if $f(A) = B$, then f is called **onto**. So f is onto if and only if the range of f equals the codomain of f.

Definition 8.12.
Let $f : A \to B$ be a function and $S \subseteq A$. Then a function $f_S : S \to B$, called the **restriction** of f to S, can be formed by defining $f_S(x) = f(x)$ for each $x \in S$.

Definition 8.13.
A function $f : A \to B$ is said to be **invertible** if there exists a function $g : B \to A$ such that $(f \circ g)(y) = y$ for all $y \in B$ and $(g \circ f)(x) = x$ for all $x \in A$.

If such a function g exists, then it is unique and is called the **inverse** of f. We denote the inverse of f (when it exists) by f^{-1}.

It can be shown that f is invertible if and only if f is both one-to-one and onto.

C   Fields

Definition 8.14. A field $\mathbb{F}$ is a set on which two operations $+$ and $\cdot$ (called **addition** and **multiplication**, respectively) are defined so that, for each pair of elements $x, y \in \mathbb{F}$, there are unique elements $x + y$ and $x \cdot y$ in $\mathbb{F}$ for which the following conditions hold for all elements $a, b, c \in \mathbb{F}$.

(F 1) $a + b = b + a$ and $a \cdot b = b \cdot a$
(commutativity of addition and multiplication)

(F 2) $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
(associativity of addition and multiplication)

(F 3) There exist distinct elements 0 and 1 in $\mathbb{F}$ such that

$$0 + a = a \quad \text{and} \quad 1 \cdot a = a$$

(existence of identity elements for addition and multiplication)

(F 4) For each element a in $\mathbb{F}$ and each nonzero element b in $\mathbb{F}$, there exists elements c and d in $\mathbb{F}$ such that

$$a + c = 0 \quad \text{and} \quad b \cdot d = 1$$

(existence of inverses for addition and multiplication)

(F 5) $a \cdot (b + c) = a \cdot b + a \cdot c$
(distributivity of multiplication over addition)

The elements $x + y$ and $x \cdot y$ are called the **sum** and **product**, respectively, of x and y. The elements 0 (read "**zero**") and 1 (read "**one**") mentioned in (F 3) are called **identity elements** for addition and multiplication, respectively, and the elements c and d referred to in (F 4) are called an **additive inverse** for a and a **multiplicative inverse** for b, respectively.

Theorem 8.1 (Cancellation Laws).
For arbitrary elements a, b, and c in a field, the following statements are true.

  1. If $a + b = c + b$, then $a = c$.

  2. If $a \cdot b = \cdot b$ and $b\ neq 0$, then $a = c$.

Corollary 8.1.
The elements 0 and 1 mentioned in (F 3), and the elements c and d mentioned in (F 4), are unique.

Theorem 8.2.
Let a and b be arbitrary elements of a field. Then each of the following statements are true.

1. $a \cdot 0 = 0$.

2. $(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$.

3. $(-a) \cdot (-b) = a \cdot b$.

Corollary 8.2.
The additive identity of a field has no multiplicative inverse.

Definition 8.15.
In an arbitrary field $\mathbb{F}$, the smallest positive integer p for which a sum of p 1's equals 0 is called the **characteristic** of $\mathbb{F}$; if no such positive integer exists, then $\mathbb{F}$ is said to have **characteristic zero**.

D   Complex Numbers

Definition 8.16.
A **complex number** is an expression of the form $z = a + bi$, where a and b are real numbers called the **real part** and the **imaginary part** of z, respectively.

The **sum** and **product** of two complex numbers $z = a + bi$ and $w = c + di$ (where a, b, c, and d are real numbers) are defined, respectively, as follows:

$$z + w = (a + bi) + (c + di) = (a + c) + (b + d)i$$

and

$$zw = (a + bi)(c + di) = (ac - bd) + (bc + ad)i$$

Definition 8.17.
Any complex number of the form $bi = 0 + bi$, where b is a nonzero real number, is called **imaginary**.

Theorem 8.3.
The set of complex numbers with the operations of addition and multiplication previously defined is a field.

Definition 8.18.
The (**complex**) **conjugate** of a complex number $a + bi$ is the complex number $a - bi$. We denote the conjugate of a complex number z by $\overline{z}$.

Theorem 8.4.
Let z and w be complex numbers. Then the following statements are true.

1. $\overline{\overline{z}} = z$.

2. $\overline{(z + w)} = \overline{z} + \overline{w}$.

3. $\overline{zw} = \overline{z} \cdot \overline{w}$.

4. $\overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}}$ if $w \neq 0$.

5. z is a real number if and only if $\overline{z} = z$.

Definition 8.19.
Let $z = a + bi$, where $a, b \in \mathbb{R}$. The **absolute value** (or **modulus**) of z is the real number $\sqrt{a^2 + b^2}$. We denote the absolute value of z as $|z|$.

Theorem 8.5.
Let z and w denote any two complex numbers. Then the following statements are true.

1. $|zw| = |z| \cdot |w|$.

2. $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$ if $w \neq 0$.

3. $|z + w| \leq |z| + |w|$.

4. $|z| - |w| \leq |z + w|$.

Definition 8.20.
Notice that, as in $\mathbb{R}^2$, there are two axes, the **real axis** and the **imaginary axis**.

Theorem 8.6 (The Fundamental Theorem of Algebra).
Suppose that $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ is a polynomial in $P(\mathbb{C})$ degree $n \geq 1$. Then $p(z)$ has a zero.

Corollary 8.3. If $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ is a polynomial of degree $n \geq 1$ with complex coefficients, then there exists complex numbers $c_1, c_2, \ldots, c_n$ (not necessarily distinct) such that

$$p(z) = a_n(z - c_1)(z - c_2) \ldots (z - c_n).$$

Definition 8.21.
A field is called **algebraically closed** if it has the property that every polynomial of positive degree 1. Thus the preceding corollary asserts that the field of complex numbers is algebraically closed.

E   Polynomials

Definition 8.22.
A polynomial $f(x)$ **divides** a polynomial $g(x)$ if there exists a polynomial $q(x)$ such that $g(x) = f(x)q(x)$.

Theorem 8.7 (The Division Algorithm for Polynomials). Let $f(x)$ be a polynomial of degree n, and let $g(x)$ be a polynomial of degree $m \geq 0$. Then there exists unique polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x)g(x) + r(x),$$

where the degree of $r(x)$ is less than x.

Definition 8.23.
In the context of Theorem 8.7, we call $q(x)$ and $r(x)$ the **quotient** and **remainder**, respectively, for the division of $f(x)$ by $g(x)$.

Corollary 8.4.
Let $f(x)$ be a polynomial of positive degree, and let $a \in \mathbb{F}$. Then $f(a) = 0$ if and only if $x - a$ divides $f(x)$.

Definition 8.24.
For any polynomial $f(x)$ with coefficients from a field $\mathbb{F}$, an element $a \in \mathbb{F}$ is called a **zero** of $f(x)$ if $f(a) = 0$. With this terminology, the preceding corollary states that a is a zero of $f(x)$ if and only if $x - a$ divides $f(x)$.

Corollary 8.5.
Any polynomial of degree $n \geq 1$ has at most n distinct zeros.

Definition 8.25.
Two nonzero polynomials are called **relatively prime** if no polynomial of positive degree divides each of them.

Theorem 8.8.
If $f_1(x)$ and $f_2(x)$ are relatively prime polynomials, there exist polynomials $q_1(x)$ and $q_2(x)$ such that

$$q_1(x)f_1(x) + q_2(x)f_2(x) = 1,$$

where 1 denoted the constant polynomial with value 1.

Definition 8.26.
Let

$$f(x) = a_0 + a_1(x) + \cdots + a_n x^n$$

be a polynomial with coefficients from a field $\mathbb{F}$. If T is a linear operator on a vector space V over $\mathbb{F}$, we define

$$f(T) = a_0 I + a_1 T + \cdots + a_n T^n.$$

Similarly, if A is an $n \times n$ matrix with entries from $\mathbb{F}$, we define

$$f(A) = a_0 I + a_1 A + \cdots + a_n A^n.$$

Theorem 8.9.
Let $f(x)$ be a polynomial with coefficients from a field $\mathbb{F}$, and let T be a linear operator on a vector space V over $\mathbb{F}$. Then the following statements are true.

1. $f(T)$ is a linear operator on V.

2. If β is a finite ordered basis for V and $A = [T]_\beta$, then $[f(T)]_\beta = f(A)$.

Theorem 8.10.
Let T be a linear operator on a vector space V over a field $\mathbb{F}$, and let A be a square matrix with entries from $\mathbb{F}$. Then, for any polynomials $f_1(x)$ and $f_2(x)$ with coefficients $\mathbb{F}$,

1. $f_1(T)f_2(T) = f_2(T)f_1(T)$

2. $f_1(A)f_2(A) = f_2(A)f_1(A)$.

Theorem 8.11.
Let T be a linear operator on a vector space V over a field $\mathbb{F}$, and let A be an $n \times n$ matrix with entries from $\mathbb{F}$. If $f_1(x)$ and $f_2(x)$ are relatively prime polynomials with entries from $\mathbb{F}$, then there exist polynomials $q_1(x)$ and $q_2(x)$ with entries from $\mathbb{F}$ such that

1. $q_1(T)f_1(T) + q_2(T)f_2(T) = I$

2. $q_1(A)f_1(A) + q_2(A)f_2(A) = I$.

Definition 8.27.
A polynomial $f(x)$ with coefficients from a field $\mathbb{F}$ is called **monic** if its leading coefficient is 1. If $f(x)$ has positive degree and cannot be expressed as a product of polynomials with coefficients from $\mathbb{F}$ each having positive degree, then $f(x)$ is called **irreducible**.

Theorem 8.12.
Let $\phi(x)$ and $f(x)$ be polynomials. If $\phi(x)$ is irreducible and $\phi(x)$ does not divide $f(x)$, then $\phi(x)$ and $f(x)$ are relatively prime.

Theorem 8.13.
Any two distinct irreducible monic polynomials are relatively prime.

Theorem 8.14.
Let $f(x)$, $g(x)$, and $\phi(x)$ be polynomials. If $\phi(x)$ is irreducible and divides the product $f(x)g(x)$, then $\phi(x)$ divides $f(x)$ or $\phi(x)$ divides $g(x)$.

Corollary 8.6.
Let $\phi(x), \phi_1(x)\phi_2(x), \ldots, \phi_n(x)$ be irreducible monic polynomials. If $\phi(x)$ divides the product $\phi_1(x)\phi_2(x)\ldots\phi_n(x)$, then $\phi(x) = \phi_i(x)$ for some i ($i = 1, 2, \ldots n$).

Theorem 8.15 (Unique Factorization Theorem for Polynomials).
For any polynomial $f(x)$ of positive degree, there exist a unique constant c; unique distinct irreducible monic polynomials $\phi_1(x), \phi_2(x), \ldots, \phi_n(x)$; and unique positive integers $n_1, n_2, \ldots, n_k$ such that

$$f(x) = c[\phi_1(x)]^{n_1}[\phi_2(x)]^{n_2}\ldots[\phi_k(x)]^{n_k}.$$

Theorem 8.16.
Let $f(x)$ and $g(x)$ be polynomials with coefficients from an infinite field $\mathbb{F}$. If $f(a) = g(a)$ for all $a \in \mathbb{F}$, then $f(x)$ and $g(x)$ are equal.