Linear Algebra:

Concepts and Theories on Abstract Vector Spaces

Third e-Book Edition (Version 3.1)

by

Ma Siu Lun

Department of Mathematics

National University of Singapore

December 4, 2018

Preface

This is my lecture notes of the module MA2101 Linear Algebra II. It is the continuation of the book "Linear Algebra: Concepts and Techniques on Euclidean Spaces, Second Edition, McGraw Hill" which is the textbook of the first year module MA1101R Linear Algebra I.

You shall notice that the first chapter is named Chapter 8. For Chapters 1 to 7, I refer to the chapters of the textbook of MA1101R mentioned above. References and results from the textbook of MA1101R will be quoted directly using the reference codes, e.g. Theorem 3.1.6, Definition 7.1.1, etc.

From linear algebra modules in 1000 level, you basically learnt the properties of Euclidean Spaces $\mathbb{R}^n$ and their subspaces. In MA2101, you shall study the abstract version of vector spaces. Most concepts and results will be generalized to this abstract version of vector spaces. Furthermore, new topics like direct sums, quotient spaces, isomorphisms, Jordan canonical forms, etc., will be introduced. A key difference from linear algebra modules in 1000 level is that there is a greater emphasis on conceptual understanding of theoretical results than on routine computations. Since MA2101 is built upon the background knowledge of MA1101R, you are advised to revise or study the textbook of MA1101R before attending classes of MA2101.

S.L. Ma

Contents

Chapter 8

General Vector Spaces

Section 8.1 Fields

Discussion 8.1.1 In Chapter 3, we have defined the Euclidean n-space $\mathbb{R}^n$ with two operations: the addition and scalar multiplication. The same operations are also defined in Chapter 2 when we study matrices. It is natural to ask if the results of n-vectors studied in Chapter 3 can be applied to matrices as well. In order to unify all these similar mathematical objects, we need a more general framework for vector spaces. But before we study this abstract version of vector spaces, we first introduce an abstract version of real numbers.

Definition 8.1.2 A *field* consists of the following:

(a) a nonempty set $\mathbb{F}$;

(b) an operation of *addition* $a+b$ between every pair of elements $a, b \in \mathbb{F}$; and

(c) an operation of *multiplication* ab between every pair of elements $a, b \in \mathbb{F}$.

Furthermore, the operations satisfy the following axioms:

(F1) **(Closure under Addition)** For all $a, b \in \mathbb{F}$, $a+b \in \mathbb{F}$.

(F2) **(Commutative Law for Addition)** For all $a, b \in \mathbb{F}$, $a+b=b+a$.

(F3) **(Associative Law for Addition)** For all $a, b, c \in \mathbb{F}$, $(a+b)+c=a+(b+c)$.

(F4) **(Existence of the Additive Identity)** There exists an element $0 \in \mathbb{F}$ such that $a+0=a$ for all $a \in \mathbb{F}$. We call 0 the *zero element* of $\mathbb{F}$ and all other elements in $\mathbb{F}$ are called *nonzero elements* of $\mathbb{F}$.
(By Proposition 8.1.5.1, we know that there exists only one such element in $\mathbb{F}$.)

(F5) **(Existence of Additive Inverses)** For every $a \in \mathbb{F}$, there exists $b \in \mathbb{F}$ such that $a + b = 0$. We call b the *additive inverse* of a and denote b by $-a$.
(By Proposition 8.1.5.2, we know that for each a, the additive inverse of a is unique.)

(F6) **(Closure under Multiplication)** For all $a, b \in \mathbb{F}$, $ab \in \mathbb{F}$.

(F7) **(Commutative Law for Multiplication)** For all $a, b \in \mathbb{F}$, $ab = ba$.

(F8) **(Associative Law for Multiplication)** For all $a, b, c \in \mathbb{F}$, $(ab)c = a(bc)$.

(F9) **(Existence of the Multiplicative Identity)** There exists a nonzero element 1 in $\mathbb{F}$ such that $1a = a$ for all $a \in \mathbb{F}$.
(By Proposition 8.1.5.3, we know that there exists only one such element in $\mathbb{F}$.)

(F10) **(Existence of Multiplicative Inverses)** For every nonzero element $a \in \mathbb{F}$, there exists $c \in \mathbb{F}$ such that $ac = 1$. We call c the *multiplicative inverse* of a and denote c by a^{-1}.
(By Proposition 8.1.5.4, we know that for each a, the multiplicative inverse of a is unique.)

(F11) **(Distributive Law)** For all $a, b, c \in \mathbb{F}$, $a(b + c) = (ab) + (ac)$.

Example 8.1.3

1. **(Number Systems)** When we talk about a *number system*, we refer to one of the following sets together with the usual addition and multiplication.

   (a) $\mathbb{N}$ = the set of all natural numbers, i.e. $\mathbb{N} = \{1, 2, 3, \dots\}$.
   (b) $\mathbb{Z}$ = the set of all integers, i.e. $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$.
   (c) $\mathbb{Q}$ = the set of all rational numbers, i.e. $\mathbb{Q} = \left\{ \frac{p}{q} \,\middle|\, p, q \in \mathbb{Z} \text{ and } q \neq 0 \right\}$.
   (d) $\mathbb{R}$ = the set of all real numbers.
   (e) $\mathbb{C}$ = the set of all complex numbers, i.e. $\mathbb{C} = \{a + b\mathrm{i} \mid a, b \in \mathbb{R}\}$ where $\mathrm{i}^2 = -1$ ($\mathrm{i} = \sqrt{-1}$).

   Note that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. The number systems $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ (with the usual addition and multiplication) satisfy all axioms in Definition 8.1.2. So they are fields.

   The number system $\mathbb{N}$ does not satisfy (F4), (F5) and (F10) and hence is not a field.

   The number system $\mathbb{Z}$ does not satisfy (F10) and hence is not a field.

2. Let $\mathbb{F}_2 = \{0, 1\}$. Define the addition and multiplication on $\mathbb{F}_2$ by

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

   It can be checked that $\mathbb{F}_2$ is a field.

A field which has only finitely many elements is called a *finite field*. The field $\mathbb{F}_2$ here is an example of a finite field. It is known that a finite field of q elements exists if and only if $q = p^s$ for some prime number p and positive integer s.

Remark 8.1.4 The approach of defining something by properties (or axioms) is one of the characteristics of modern mathematics. The advantage is that once we prove a theorem based on these properties, the theorem can automatically be applied to all mathematical objects that have these properties.

Proposition 8.1.5 Let $\mathbb{F}$ be a field.

1. **(Uniqueness of the Additive Identity)** If $b, c \in \mathbb{F}$ satisfies the property that $a + b = a + c = a$ for all $a \in \mathbb{F}$, then $b = c$.

2. **(Uniqueness of the Additive Inverse)** For any $a \in \mathbb{F}$, if there exist $b, c \in \mathbb{F}$ such that $a + b = a + c = 0$, then $b = c$.

3. **(Uniqueness of the Multiplicative Identity)** If b, c are nonzero elements in $\mathbb{F}$ satisfying the property $ba = ca = a$ for all $a \in \mathbb{F}$, then $b = c$.

4. **(Uniqueness of the Multiplicative Inverse)** For any $a \in \mathbb{F}$ and $a \neq 0$, if there exist $b, c \in \mathbb{F}$ such that $ab = ac = 1$, then $b = c$.

5. For any $a \in \mathbb{F}$, $a0 = 0$ and $(-1)a = -a$.

6. For any $a, b \in \mathbb{F}$, if $ab = 0$, then $a = 0$ or $b = 0$.

Proof We only show the proof of Part 2:

$$\begin{aligned}
b &= b + 0 && \text{by (F4)} \\
&= b + (a + c) && \text{by the given assumption} \\
&= (b + a) + c && \text{by (F3)} \\
&= (a + b) + c && \text{by (F2)} \\
&= 0 + c && \text{by the given assumption} \\
&= c + 0 && \text{by (F2)} \\
&= c && \text{by (F4).}
\end{aligned}$$

(Proofs of the other parts are left as exercises. See Question 8.5.)

Definition 8.1.6 Let $\mathbb{F}$ be a field. We can define the subtraction and division as follows.

1. For any $a, b \in \mathbb{F}$, the *substraction* of a by b is defined to be $a + (-b)$ and is denoted by $a - b$.

2. For any $a, b \in \mathbb{F}$ where b is nonzero, the *division* of a by b is defined to be ab^{-1}. Unlike the real and complex numbers, we seldom use the notation $a \div b$ and $\frac{a}{b}$ when working with an abstract field.

Discussion 8.1.7 The results we established in Chapter 1 and Chapter 2 can be generalized to any fields. In particular, row-echelon and reduced row-echelon forms, Gaussian and Gauss-Jordan Eliminations, matrix operations, inverses and determinants of square matrices will be used in the following chapters over any fields.

Definition 8.1.8 Let $\mathbb{F}$ be a field.

1. A linear system with all coefficients taken from $\mathbb{F}$ is called a *linear system over* $\mathbb{F}$. In particular, a linear system over $\mathbb{R}$ is called a *real linear system* and a linear system over $\mathbb{C}$ is called a *complex linear system.*

2. A matrix with all entries taken from $\mathbb{F}$ is called a *matrix over* $\mathbb{F}$. In particular, a matrix over $\mathbb{R}$ is called a *real matrix* and a matrix over $\mathbb{C}$ is called a *complex matrix.*

Example 8.1.9

1. Solve the following complex linear system:

$$\begin{cases} x_1 + \mathrm{i}x_2 + 3\mathrm{i}x_3 = 0 \\ \mathrm{i}x_1 + x_2 + x_3 = 0 \\ (1-\mathrm{i})x_1 + (1+\mathrm{i})x_3 = 0. \end{cases}$$

Solution

$$\left(\begin{array}{ccc|c} 1 & \mathrm{i} & 3\mathrm{i} & 0 \\ \mathrm{i} & 1 & 1 & 0 \\ 1-\mathrm{i} & 0 & 1+\mathrm{i} & 0 \end{array}\right) \xrightarrow[R_3-(1-\mathrm{i})R_1]{R_2-\mathrm{i}R_1} \left(\begin{array}{ccc|c} 1 & \mathrm{i} & 3\mathrm{i} & 0 \\ 0 & 2 & 4 & 0 \\ 0 & -1-\mathrm{i} & -2-2\mathrm{i} & 0 \end{array}\right) \xrightarrow{R_3+\frac{1}{2}(1+\mathrm{i})R_2}$$

$$\left(\begin{array}{ccc|c} 1 & \mathrm{i} & 3\mathrm{i} & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\frac{1}{2}R_2} \left(\begin{array}{ccc|c} 1 & \mathrm{i} & 3\mathrm{i} & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_1-\mathrm{i}R_2} \left(\begin{array}{ccc|c} 1 & 0 & \mathrm{i} & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The last augmented matrix corresponds to the complex system

$$\begin{cases} x_1 + \mathrm{i}x_3 = 0 \\ x_2 + 2x_3 = 0 \end{cases}$$

which has a general solution $x_1 = -\mathrm{i}t$, $x_2 = -2t$, $x_3 = t$ for $t \in \mathbb{C}$.

2. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ be a matrix over $\mathbb{F}_2$. Find the inverse of $\boldsymbol{A}$.

Solution Note that in $\mathbb{F}_2$, $1+1=0$.

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_3+R_1} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right) \xrightarrow{R_3+R_2} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[R_2+R_3]{R_1+R_3} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right) \xrightarrow{R_1+R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right)$$

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

3. Find the determinant of the complex matrix $\boldsymbol{B} = \begin{pmatrix} \mathrm{i} & 0 & 0 & 0 \\ 0 & 1+\mathrm{i} & 1 & 0 \\ 0 & 0 & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 1+2\mathrm{i} & 0 \end{pmatrix}$.

   Is $\boldsymbol{B}$ invertible?

   Solution

$$\det(\boldsymbol{B}) = \begin{vmatrix} \mathrm{i} & 0 & 0 & 0 \\ 0 & 1+\mathrm{i} & 1 & 0 \\ 0 & 0 & 2 & \mathrm{i} \\ 1 & -\mathrm{i} & 1+2\mathrm{i} & 0 \end{vmatrix} = \mathrm{i} \begin{vmatrix} 1+\mathrm{i} & 1 & 0 \\ 0 & 2 & \mathrm{i} \\ -\mathrm{i} & 1+2\mathrm{i} & 0 \end{vmatrix} - 0 + 0 - 0$$

$$= \mathrm{i}\left[(1+\mathrm{i}) \begin{vmatrix} 2 & \mathrm{i} \\ 1+2\mathrm{i} & 0 \end{vmatrix} - \begin{vmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{vmatrix} + 0\right] = \mathrm{i}[(1+\mathrm{i})(2-\mathrm{i})+1] = -1+4\mathrm{i}.$$

   Since $\det(\boldsymbol{B}) \neq 0$, $\boldsymbol{B}$ is invertible.

Definition 8.1.10 Let $\mathbb{F}$ be a field and $\boldsymbol{A} = (a_{ij})$ an $n \times n$ matrix over $\mathbb{F}$, i.e.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

The *trace* of $\boldsymbol{A}$, denoted by $\operatorname{tr}(\boldsymbol{A})$, is defined to be the sum of the entries on the diagonal of $\boldsymbol{A}$, i.e.

$$\operatorname{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

Note that $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^{\mathrm{T}})$. (See also Question 2.11.)

Proposition 8.1.11

1. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ matrix over $\mathbb{F}$, then $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$.
2. If $c \in \mathbb{F}$ and $\boldsymbol{A}$ is an $n \times n$ matrix over $\mathbb{F}$, then $\operatorname{tr}(c\boldsymbol{A}) = c\operatorname{tr}(\boldsymbol{A})$.
3. If $\boldsymbol{C}$ and $\boldsymbol{D}$ are $m \times n$ and $n \times m$ matrices, respectively, over $\mathbb{F}$, then $\operatorname{tr}(\boldsymbol{CD}) = \operatorname{tr}(\boldsymbol{DC})$.

Proof The proof is left as exercises. See Question 8.6.

Remark 8.1.12 In general, $\operatorname{tr}(\boldsymbol{XYZ}) \neq \operatorname{tr}(\boldsymbol{YXZ})$ even when the matrices $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ can be multiplied accordingly.

For example, let $\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $\boldsymbol{Y} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\boldsymbol{Z} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then

$$\boldsymbol{XYZ} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{YXZ} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

Hence $\operatorname{tr}(\boldsymbol{XYZ}) = 1 \neq -1 = \operatorname{tr}(\boldsymbol{YXZ})$.

Section 8.2 Vector Spaces

Discussion 8.2.1 In this section, we give an abstract definition of vector spaces. Under this general framework, vector spaces learnt in Chapter 3 are particular examples.

Definition 8.2.2 A *vector space* consists of the following:

(a) a field $\mathbb{F}$, where the elements in $\mathbb{F}$ are called *scalars*;

(b) a nonempty set V, where the elements in V are called *vectors*;

(c) an operation of *vector addition* $\boldsymbol{u} + \boldsymbol{v}$ between every pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in V$; and

(d) an operation of *scalar multiplication* $c\boldsymbol{u}$ between every $c \in \mathbb{F}$ and every vector $\boldsymbol{u} \in V$.

Furthermore, the operations satisfy the following axioms:

(V1) **(Closure under Vector Addition)** For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} + \boldsymbol{v} \in V$.

(V2) **(Commutative Law for Vector Addition)** For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.

(V3) **(Associative Law for Vector Addition)** For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w}) = (\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$.

(V4) **(Existence of the Zero Vector)** There exists a vector $\mathbf{0} \in V$, called the *zero vector*, such that $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$.
(By Proposition 8.2.4.1, we know that there exists only one such vector in V.)

(V5) **(Existence of Additive Inverse)** For every vector $\boldsymbol{u} \in V$, there exists a vector $\boldsymbol{v} \in V$ such that $\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}$. We call $\boldsymbol{v}$ the *negative* of $\boldsymbol{u}$ and denote $\boldsymbol{v}$ by $-\boldsymbol{u}$.
(By Proposition 8.2.4.2, we know that for each $\boldsymbol{u}$, the negative of $\boldsymbol{u}$ is unique.)

By this axiom, for $\boldsymbol{u}, \boldsymbol{v} \in V$, we can define the *subtraction* of $\boldsymbol{u}$ by $\boldsymbol{v}$ to be $\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v})$.

(V6) **(Closure under Scalar Multiplication)** For all $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $c\boldsymbol{u} \in V$.

(V7) For all $b, c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $b(c\boldsymbol{u}) = (bc)\boldsymbol{u}$.

(V8) For all $\boldsymbol{u} \in V$, $1\boldsymbol{u} = \boldsymbol{u}$.

(V9) For all $c \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$, $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$.

(V10) For all $b, c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $(b + c)\boldsymbol{u} = b\boldsymbol{u} + c\boldsymbol{u}$.

(The last two axioms, i.e. (V9) and (V10), are also known as **Distributive Laws**.)

As a convention, we say that V together with the given vector addition and scalar multiplication is a *vector space over* $\mathbb{F}$. If the vector addition and scalar multiplication are known, we simply say that V is a vector space over $\mathbb{F}$. In particular, if $\mathbb{F} = \mathbb{R}$, V is called a *real vector space*; and if $\mathbb{F} = \mathbb{C}$, V is called a *complex vector space*.

Example 8.2.3

1. The Euclidean n-space $\mathbb{R}^n$ in Chapter 3 is a real vector space using the vector addition and scalar multiplication defined in Definition 3.1.3. The axioms (V1) and (V6) are obviously satisfied. The other axioms follow by Theorem 3.1.6.

2. Let $\mathbb{F}$ be a field. The set $\mathbb{F}^n = \{(u_1, u_2, \ldots, u_n) \mid u_1, u_2, \ldots, u_n \in \mathbb{F}\}$ together with the vector addition:

$$(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) = (u_1 + v_1,\ u_2 + v_2,\ \ldots,\ u_n + v_n),$$

and the scalar multiplication

$$c(u_1, u_2, \ldots, u_n) = (cu_1,\ cu_2,\ \ldots,\ cu_n),$$

where $c \in \mathbb{F}$ and $(u_1, u_2, \ldots, u_n)$, $(v_1, v_2, \ldots, v_n) \in \mathbb{F}^n$, is a vector space over $\mathbb{F}$. The zero vector is $(0, 0, \ldots, 0)$.

In particular, $\mathbb{Q}^n$, $\mathbb{C}^n$ and $\mathbb{F}_2^n$ are vector spaces.

3. Let $\mathbb{F}$ be a field and let $\mathcal{M}_{m\times n}(\mathbb{F})$ be the set of all $m\times n$ matrices over $\mathbb{F}$, i.e. $\boldsymbol{A}\in\mathcal{M}_{m\times n}(\mathbb{F})$ if and only if

$$\boldsymbol{A}=(a_{ij})_{m\times n}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{where } a_{ij}\in\mathbb{F} \text{ for all } i,j.$$

We can define the matrix addition and scalar multiplication for matrices over $\mathbb{F}$ as discussed in Discussion 8.1.7, i.e.

$$(a_{ij})+(b_{ij})=(a_{ij}+b_{ij}) \quad \text{for } (a_{ij}),\ (b_{ij})\in\mathcal{M}_{m\times n}(\mathbb{F}),$$

and

$$c(a_{ij})=(ca_{ij}) \quad \text{for } c\in\mathbb{F} \text{ and } (a_{ij})\in\mathcal{M}_{m\times n}(\mathbb{F})$$

(see Notation 2.1.5 for the notation used above). Then $\mathcal{M}_{m\times n}(\mathbb{F})$ is a vector space over $\mathbb{F}$. The zero vector is the zero matrix $\boldsymbol{0}_{\boldsymbol{m\times n}}$.

In particular, $\mathcal{M}_{m\times n}(\mathbb{Q})$, $\mathcal{M}_{m\times n}(\mathbb{R})$, $\mathcal{M}_{m\times n}(\mathbb{C})$ and $\mathcal{M}_{m\times n}(\mathbb{F}_2)$ are vector spaces.

4. Let $\mathbb{F}$ be a field and let $\mathbb{F}^{\mathbb{N}}$ be the set of all infinite sequences $\boldsymbol{a}=(a_n)_{n\in\mathbb{N}}=(a_1,a_2,a_3,\dots)$ with $a_1,a_2,a_3,\dots\in\mathbb{F}$, i.e.

$$\mathbb{F}^{\mathbb{N}}=\{(a_n)_{n\in\mathbb{N}} \mid a_1,a_2,a_3,\dots\in\mathbb{F}\}.$$

For $\boldsymbol{a}=(a_n)_{n\in\mathbb{N}}$, $\boldsymbol{b}=(b_n)_{n\in\mathbb{N}}\in\mathbb{F}^{\mathbb{N}}$, define the addition of sequences by

$$\boldsymbol{a}+\boldsymbol{b}=(a_n+b_n)_{n\in\mathbb{N}}=(a_1+b_1,\ a_2+b_2,\ a_3+b_3,\ \dots).$$

For $c\in\mathbb{F}$ and $\boldsymbol{a}=(a_1,a_2,a_3,\dots)\in\mathbb{F}^{\mathbb{N}}$, define the scalar multiplication by

$$c\boldsymbol{a}=(ca_n)_{n\in\mathbb{N}}=(ca_1,\ ca_2,\ ca_3,\ \dots).$$

Then $\mathbb{F}^{\mathbb{N}}$ is a vector space over $\mathbb{F}$. The zero vector is the *zero sequence* $(0,0,0,\dots)$.

In particular, $\mathbb{Q}^{\mathbb{N}}$, $\mathbb{R}^{\mathbb{N}}$, $\mathbb{C}^{\mathbb{N}}$ and $\mathbb{F}_2^{\mathbb{N}}$ are vector spaces.

5. Let $\mathbb{F}$ be a field. A polynomial $p(x)=a_0+a_1x+\cdots+a_mx^m$, with $a_0,a_1,\dots,a_m\in\mathbb{F}$, is called a *polynomial over* $\mathbb{F}$. In particular, if $\mathbb{F}=\mathbb{R}$, $p(x)$ is called a *real polynomial*; and if $\mathbb{F}=\mathbb{C}$, $p(x)$ is called a *complex polynomial.*

As a convention, terms in a polynomial with zero coefficients can be dropped, e.g. $1+0x+2x^2+x^3+0x^4$ can be written as $1+2x+x^3$. Two polynomials are the same if all terms having nonzero coefficients are the same, e.g. $1+0x+2x^2+x^3+0x^4=1+2x+x^3$.

Let $\mathcal{P}(\mathbb{F})$ be the set of all polynomials over $\mathbb{F}$, i.e.

$$\mathcal{P}(\mathbb{F})=\{a_0+a_1x+\cdots+a_mx^m \mid m \text{ is a nonnegative integer and } a_0,a_1,\dots,a_m\in\mathbb{F}\}.$$

For $a_0 + a_1x + \cdots + a_mx^m$, $b_0 + b_1x + \cdots + b_nx^n \in \mathcal{P}(\mathbb{F})$ where $m \le n$, define the addition of the polynomials by

$$\begin{aligned}
&(a_0 + a_1x + \cdots + a_mx^m) + (b_0 + b_1x + \cdots + b_nx^n) \\
&= (b_0 + b_1x + \cdots + b_nx^n) + (a_0 + a_1x + \cdots + a_mx^m) \\
&= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_m + b_m)x^m + b_{m+1}x^{m+1} + \cdots + b_nx^n.
\end{aligned}$$

For $c \in \mathbb{F}$ and $a_0 + a_1x + \cdots + a_mx^m \in \mathcal{P}(\mathbb{F})$, define the scalar multiplication by

$$c(a_0 + a_1x + \cdots + a_mx^m) = ca_0 + ca_1x + \cdots + ca_mx^m.$$

Then $\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$. The zero vector is the *zero polynomial* 0.

In particular, $\mathcal{P}(\mathbb{Q})$, $\mathcal{P}(\mathbb{R})$, $\mathcal{P}(\mathbb{C})$ and $\mathcal{P}(\mathbb{F}_2)$ are vector spaces.

6. Let A be a nonempty set and $\mathbb{F}$ a field. Let $\mathcal{F}(A, \mathbb{F})$ be the set of all functions $f : A \to \mathbb{F}$. For $f, g \in \mathcal{F}(A, \mathbb{F})$, define the function $f + g : A \to \mathbb{F}$ by

$$(f + g)(a) = f(a) + g(a) \quad \text{for } a \in A.$$

For $c \in \mathbb{F}$ and $f \in \mathcal{F}(A, \mathbb{F})$, define the function $cf : A \to \mathbb{F}$ by

$$(cf)(a) = cf(a) \quad \text{for } a \in A.$$

Then $\mathcal{F}(A, \mathbb{F})$ is a vector space over $\mathbb{F}$. The zero vector is the *zero function* $\mathit{0} : A \to \mathbb{F}$ define by $\mathit{0}(a) = 0$ for $a \in A$.

7. Let $\mathbb{F}$ be a field and $V = \{\mathbf{0}\}$. We define

$$\mathbf{0} + \mathbf{0} = \mathbf{0} \qquad \text{and} \qquad c\mathbf{0} = \mathbf{0} \quad \text{for } c \in \mathbb{F}.$$

Then V is a vector space over $\mathbb{F}$ which is called a *zero space.*

8. $\mathbb{C} = \{a + b\mathrm{i} \mid a, b \in \mathbb{R}\}$ is a vector space over $\mathbb{R}$ using the usual addition of complex numbers as the vector addition and the usual multiplication of real numbers to complex numbers as the scalar multiplication, i.e.

$$(a + b\mathrm{i}) + (c + d\mathrm{i}) = (a + c) + (b + d)\mathrm{i} \quad \text{for } a + b\mathrm{i},\ c + d\mathrm{i} \in \mathbb{C} \text{ with } a, b, c, d \in \mathbb{R}$$

and

$$c(a + b\mathrm{i}) = (ca) + (cb)\mathrm{i} \quad \text{for } c \in \mathbb{R} \text{ and } a + b\mathrm{i} \in \mathbb{C} \text{ with } a, b \in \mathbb{R}.$$

9. Let V be the set of all positive real numbers, i.e. $V = \{a \in \mathbb{R} \mid a > 0\}$.

   (a) V is not a vector space over $\mathbb{R}$ using the usual addition of real numbers as the vector addition and the usual multiplication of real numbers as the scalar multiplication. It does not satisfied axioms (V5) and (V6) in Definition 8.2.2.

(b) Define the vector addition $\dagger$ by

$$a \dagger b = ab \quad \text{for } a, b \in V$$

and define the scalar multiplication $*$ by

$$m * a = a^m \quad \text{for } m \in \mathbb{R} \text{ and } a \in V.$$

Then V is a vector space over $\mathbb{R}$ using these two operations. (We leave the verification as exercise. See Question 8.9.)

Proposition 8.2.4 Let V be a vector space over a field $\mathbb{F}$.

1. **(Uniqueness of the Zero Vector)** If $\boldsymbol{v}, \boldsymbol{w} \in V$ satisfy the property that $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$, then $\boldsymbol{v} = \boldsymbol{w}$.
2. **(Uniqueness of the Additive Inverse)** For any $\boldsymbol{u} \in V$, if there exist $\boldsymbol{v}, \boldsymbol{w} \in V$ such that $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} = \mathbf{0}$, then $\boldsymbol{v} = \boldsymbol{w}$.
3. For all $\boldsymbol{u} \in V$, $0\boldsymbol{u} = \mathbf{0}$ and $(-1)\boldsymbol{u} = -\boldsymbol{u}$.
4. For all $c \in \mathbb{F}$, $c\mathbf{0} = \mathbf{0}$.
5. If $c\boldsymbol{u} = \mathbf{0}$ where $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, then $c = 0$ or $\boldsymbol{u} = \mathbf{0}$.

Proof We only show the proof of $0\boldsymbol{u} = \mathbf{0}$ in Part 3: By (F4), $0 + 0 = 0$. Thus

$$0\boldsymbol{u} = (0+0)\boldsymbol{u} = 0\boldsymbol{u} + 0\boldsymbol{u}. \qquad \text{by (V10)}$$

By (V5), the vector $0\boldsymbol{u}$ has the negative $-0\boldsymbol{u}$. Adding $-0\boldsymbol{u}$ to both sides of the equation above yields

$$\begin{aligned} & & 0\boldsymbol{u} + (-0\boldsymbol{u}) &= [0\boldsymbol{u} + 0\boldsymbol{u}] + (-0\boldsymbol{u}) & \\ &\Rightarrow & 0\boldsymbol{u} + (-0\boldsymbol{u}) &= 0\boldsymbol{u} + [0\boldsymbol{u} + (-0\boldsymbol{u})] & \text{by (V3)} \\ &\Rightarrow & \mathbf{0} &= 0\boldsymbol{u} + \mathbf{0} & \text{by (V5)} \\ &\Rightarrow & \mathbf{0} &= 0\boldsymbol{u}. & \text{by (V4)} \end{aligned}$$

(Proofs of the other parts are left as exercises. See Question 8.12.)

Section 8.3 Subspaces

Discussion 8.3.1 Given an arbitrary subset W of a vector space V, under the same vector addition and scalar multiplication as in V, W automatically satisfies axioms (V2), (V3) and

(V7)-(V10) in Definition 8.2.2 (whenever they make sense). In case W also satisfies (V1), (V4), (V5) and (V6), it forms a vector space sitting inside the larger vector space V. For example, in $\mathbb{R}^3$, the xy-plane is itself a vector space.

For most of the applications of linear algebra, we need to work with smaller vector spaces sitting inside a big vector space.

Definition 8.3.2 A subset W of a vector space V is called a *subspace* of V if W is itself a vector space using the same vector addition and scalar multiplication as in V.

In Definition 3.3.2, subspaces of $\mathbb{R}^n$ are defined differently. It can be shown that when we apply the definition of subspaces here to $\mathbb{R}^n$, we get the same kind of subspaces in Chapter 3 (see Section 8.4).

Example 8.3.3

1. Let V be a vector space and $\mathbf{0}$ the zero vector.

   (a) Since $\{\mathbf{0}\}$ is a subset of V and $\{\mathbf{0}\}$ is a vector space (a zero space), $\{\mathbf{0}\}$ is a subspace of V.

   (b) Since V is a subset of V and V is a vector space, V is a subspace of V.

   These two subspaces, $\{\mathbf{0}\}$ and V, are called *trivial subspaces* of V. Other subspaces of V are called *proper subspaces* of V.

2. Let $\mathbb{F}$ be a field. Let $V = \mathbb{F}^2$ and $W = \{(a, a) \mid a \in \mathbb{F}\} \subseteq V$.

   (V1) Take any two vectors $(a, a), (b, b) \in W$. The sum $(a, a) + (b, b) = (a + b,\ a + b)$ is again a vector in W. Hence W is closed under the vector addition.

   (V4) The zero vector $(0, 0)$ of V is also contained in W.

   (V5) Take any vector $(a, a) \in W$. The negative of (a, a) is $(-a, -a)$ which is again a vector in W. So the negative of every vector in W is also contained in W.

   (V6) Take any vector $(a, a) \in W$ and any scalar $c \in \mathbb{F}$. The scalar multiple $c(a, a) = (ca, ca)$ is again a vector in W. Hence W is closed under the scalar multiplication.

   By Discussion 8.3.1, W is a subspace of V. (Actually, by Theorem 8.3.4 below, we do not need to check (V5).)

Theorem 8.3.4 Let V be a vector space over a field $\mathbb{F}$. A subset W of V is a subspace of V if and only if

(S1) **(Containing the Zero Vector)** $\mathbf{0} \in W$;

(S2) **(Closure under the Vector Addition)** for all $\boldsymbol{u}, \boldsymbol{v} \in W$, $\boldsymbol{u} + \boldsymbol{v} \in W$; and

(S3) **(Closure under the Scalar Multiplication)** for all $c \in \mathbb{F}$ and $\boldsymbol{u} \in W$, $c\boldsymbol{u} \in W$.

Proof

(⇒) (S1), (S2) and (S3) follow directly from the definition of vector spaces.

(⇐) Suppose W satisfies (S1), (S2) and (S3). We only need to show that W satisfies (V5). (Why?)

(V5) Take any vector $\boldsymbol{u} \in W$. By Proposition 8.2.4.3, $-\boldsymbol{u} = (-1)\boldsymbol{u}$ and, by (S3), it is contained in W.

Remark 8.3.5 Theorem 8.3.4 can be simplified further: Let W be a nonempty subset of a vector space V over a field $\mathbb{F}$. Then W is a subspace of V if and only if

$$\text{for all } a, b \in \mathbb{F} \text{ and } \boldsymbol{u}, \boldsymbol{v} \in W,\ a\boldsymbol{u} + b\boldsymbol{v} \in W. \tag{8.1}$$

(The proof is left as exercise. See Question 8.14.)

Comparing with Theorem 8.3.4, (S1) is replaced by the "nonempty" condition while (S2) and (S3) are combined as (8.1).

Example 8.3.6

1. Let $\mathbb{F}$ be a field and $\boldsymbol{W} = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} \middle|\, a, b \in \mathbb{F} \right\} \subseteq \mathcal{M}_{2\times 2}(\mathbb{F})$.

   (S1) $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in W$ (because the matrix is of the form $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ with $a = 0$, $b = 0 \in \mathbb{F}$).

   (S2) For any $\begin{pmatrix} a_1 & b_1 \\ b_1 & a_1 \end{pmatrix}, \begin{pmatrix} a_2 & b_2 \\ b_2 & a_2 \end{pmatrix} \in W$ (where $a_1, a_2, b_1, b_2 \in \mathbb{F}$),

$$\begin{pmatrix} a_1 & b_1 \\ b_1 & a_1 \end{pmatrix} + \begin{pmatrix} a_2 & b_2 \\ b_2 & a_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ b_1 + b_2 & a_1 + a_2 \end{pmatrix} \in W$$

   (because the matrix is of the form $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ with $a = a_1 + a_2$, $b = b_1 + b_2 \in \mathbb{F}$).

   (S3) For any $c \in \mathbb{F}$ and $\begin{pmatrix} a_0 & b_0 \\ b_0 & a_0 \end{pmatrix} \in W$ (where $a_0, b_0 \in \mathbb{F}$),

$$c\begin{pmatrix} a_0 & b_0 \\ b_0 & a_0 \end{pmatrix} = \begin{pmatrix} ca_0 & cb_0 \\ cb_0 & ca_0 \end{pmatrix} \in W$$

   (because the matrix is of the form $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ with $a = ca_0$, $b = cb_0 \in \mathbb{F}$).

As W is a subset of $\mathcal{M}_{2\times 2}(\mathbb{F})$ satisfying (S1), (S2) and (S3), it is a subspace of $\mathcal{M}_{2\times 2}(\mathbb{F})$.

2. (In this example, vectors in $\mathbb{F}^n$ are written as column vectors.) Let $\mathbb{F}$ be a field and $\boldsymbol{A}$ an $m \times n$ matrix over $\mathbb{F}$. Then the solution set W of the homogeneous linear system $\boldsymbol{Ax} = \boldsymbol{0}$ is a subspace of $\mathbb{F}^n$. The subspace W is called the *solution space* of $\boldsymbol{Ax} = \boldsymbol{0}$ or the *nullspace* of $\boldsymbol{A}$.

   Proof Note that $W = \{\boldsymbol{u} \in \mathbb{F}^n \mid \boldsymbol{Au} = \boldsymbol{0}\} \subseteq \mathbb{F}^n$.

   (S1) Since $\boldsymbol{A0} = \boldsymbol{0}$, $\boldsymbol{0} \in W$.

   (S2) Take any $\boldsymbol{u}, \boldsymbol{v} \in W$, i.e. $\boldsymbol{Au} = \boldsymbol{0}$ and $\boldsymbol{Av} = \boldsymbol{0}$. Since
   $$\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{Au} + \boldsymbol{Av} = \boldsymbol{0} + \boldsymbol{0} = \boldsymbol{0},$$
   $\boldsymbol{u} + \boldsymbol{v} \in W$.

   (S3) Take any $c \in \mathbb{F}$ and take any $\boldsymbol{u} \in W$, i.e. $\boldsymbol{Au} = \boldsymbol{0}$. Since
   $$\boldsymbol{A}(c\boldsymbol{u}) = c\boldsymbol{Au} = c\boldsymbol{0} = \boldsymbol{0},$$
   $c\boldsymbol{u} \in W$.

   As W is a subset of $\mathbb{F}^n$ satisfying (S1), (S2) and (S3), it is a subspace of $\mathbb{F}^n$.

3. Consider the vector space $\mathbb{R}^{\mathbb{N}}$ of infinite sequences over $\mathbb{R}$. Define
   $$W = \left\{ (a_1, a_2, a_3, \dots) \in \mathbb{R}^{\mathbb{N}} \,\middle|\, \lim_{n\to\infty} a_n = 0 \right\} \subseteq \mathbb{R}^{\mathbb{N}}.$$

   (S1) Since the limit of the zero sequence $(0, 0, 0, \dots)$ is 0, $(0, 0, 0, \dots) \in W$.

   (S2) Take any $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}}$, $\boldsymbol{b} = (a_n)_{n\in\mathbb{N}} \in W$, i.e. $\lim_{n\to\infty} a_n = 0$ and $\lim_{n\to\infty} b_n = 0$. Since
   $$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n = 0 + 0 = 0,$$
   $\boldsymbol{a} + \boldsymbol{b} = (a_n + b_n)_{n\in\mathbb{N}} \in W$.

   (S3) Take any $c \in \mathbb{R}$ and take any $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}} \in W$, i.e. $\lim_{n\to\infty} a_n = 0$. Since
   $$\lim_{n\to\infty} ca_n = c \lim_{n\to\infty} a_n = c0 = 0$$
   $c\boldsymbol{a} = (ca_n)_{n\in\mathbb{N}} \in W$.

   As W is a subset of $\mathbb{R}^{\mathbb{N}}$ satisfying (S1), (S2) and (S3), it is a subspace of $\mathbb{R}^{\mathbb{N}}$.

4. Let $\mathbb{F}$ be a field and n a positive integer. Define $\mathcal{P}_n(\mathbb{F})$ to be the set of all polynomial over $\mathbb{F}$ with degree at most n. Note that
   $$\mathcal{P}_n(\mathbb{F}) = \{a_0 + a_1x + \cdots + a_nx^n \mid a_0, a_1, \dots, a_n \in \mathbb{F}\} \subseteq \mathcal{P}(\mathbb{F}).$$
   Then $\mathcal{P}_n(\mathbb{F})$ is a subspace of $\mathcal{P}(\mathbb{F})$. If $m \leq n$, $\mathcal{P}_m(\mathbb{F})$ is a subspace of $\mathcal{P}_n(\mathbb{F})$.

5. For $a, b \in \mathbb{R}$ with $a < b$, let $[a, b] = \{c \in \mathbb{R} \mid a \le c \le b\}$ be a closed interval on the real line. Consider the set $\mathcal{F}([a, b], \mathbb{R})$ of all functions $f : [a, b] \to \mathbb{R}$. Using the addition and scalar multiplication defined in Example 8.2.3.6, $\mathcal{F}([a, b], \mathbb{R})$ forms a real vector space. Define

   (a) $C([a, b]) = \{f \in \mathcal{F}([a, b], \mathbb{R}) \mid f$ is continuous on $[a, b]\}$;

   (b) for $n \in \mathbb{N}$, $C^n([a, b]) = \{f \in \mathcal{F}([a, b], \mathbb{R}) \mid f$ is n times differentiable on $[a, b]\}$; and

   (c) $C^\infty([a, b]) = \{f \in \mathcal{F}([a, b], \mathbb{R}) \mid f$ is infinitely differentiable on $[a, b]\}$.

   Then $C([a, b])$, $C^n([a, b])$ and $C^\infty([a, b])$ are subspaces of $\mathcal{F}([a, b], \mathbb{R})$.

   Note that $C^\infty([a, b]) \subseteq C^n([a, b]) \subseteq C([a, b]) \subseteq \mathcal{F}([a, b], \mathbb{R})$ and $C^n([a, b]) \subseteq C^m([a, b])$ whenever $n \ge m$. Thus $C^\infty([a, b])$ is a subspace of $C([a, b])$ and $C^n([a, b])$ for all $n \in \mathbb{N}$; and for each $n \in \mathbb{N}$, $C^n([a, b])$ is a subspace of $C([a, b])$ and $C^m([a, b])$ if $m \le n$.

Remark 8.3.7 Since a real polynomial can be regarded as a real-valued infinitely differentiable function, $\mathcal{P}(\mathbb{R})$ and $\mathcal{P}_n(\mathbb{R})$, for $n \in \mathbb{N}$, can be considered as subspaces of $C([a, b])$, $C^m([a, b])$, for all $m \in \mathbb{N}$, and $C^\infty([a, b])$ for any closed interval $[a, b]$ on the real line.

Theorem 8.3.8 If W_1 and W_2 are two subspaces of a vector space V, then the intersection of W_1 and W_2 is also a subspace of V. (Remind that for sets A and B, the intersection of A and B is the set $A \cap B = \{x \mid x \in A$ and $x \in B\}$.)

Proof

(S1) Since both W_1 and W_2 contain the zero vector, the zero vector is contained in $W_1 \cap W_2$.

(S2) Take any $\boldsymbol{u}, \boldsymbol{v} \in W_1 \cap W_2$. Since W_1 is a subspace and $\boldsymbol{u}, \boldsymbol{v} \in W_1$, we have $\boldsymbol{u} + \boldsymbol{v} \in W_1$. Similarly, $\boldsymbol{u} + \boldsymbol{v} \in W_2$. Thus $\boldsymbol{u} + \boldsymbol{v} \in W_1 \cap W_2$.

(S3) Take any scalar c and any $\boldsymbol{u} \in W_1 \cap W_2$. Since W_1 is a subspace and $\boldsymbol{u} \in W_1$, we have $c\boldsymbol{u} \in W_1$. Similarly, $c\boldsymbol{u} \in W_2$. Thus $c\boldsymbol{u} \in W_1 \cap W_2$.

As $W_1 \cap W_2$ is a subset of V satisfying (S1), (S2) and (S3), it is a subspace of V.

Example 8.3.9 Let $\mathbb{F}$ be a field and n a positive integer. Define

$$W_1 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F}) \mid \boldsymbol{A} \text{ is an upper triangular matrix}\}$$

and

$$W_2 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F}) \mid \boldsymbol{A} \text{ is a lower triangular matrix}\}.$$

Both W_1 and W_2 are subspaces of $\mathcal{M}_{n\times n}(\mathbb{F})$ (check it). Note that $W_1 \cap W_2$ is the set of all $n \times n$ diagonal matrices over $\mathbb{F}$. It is also a subspace of $\mathcal{M}_{n\times n}(\mathbb{F})$.

Remark 8.3.10

1. If $W_1, W_2, \ldots, W_n$ are subspaces of a vector space V, then $W_1 \cap W_2 \cap \cdots \cap W_n$ is also a subspace of V.

2. The union of two subspaces of a vector space may not be a vector space. (Reminded that for sets A and B, the union of A and B is the set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.)

   For example, let $\mathbb{F}$ be a field, $W_1 = \{(x, 0) \mid x \in \mathbb{F}\}$ and $W_2 = \{(0, y) \mid y \in \mathbb{F}\}$. It is easy to check that W_1 and W_2 are subspaces of $\mathbb{F}^2$. Since $(1,0) \in W_1$ and $(0,1) \in W_2$, both $(1,0)$ and $(0,1)$ are elements of $W_1 \cup W_2$. However, $(1,0) + (0,1) = (1,1)$ is neither contained in W_1 nor W_2 and hence it is not an element of $W_1 \cup W_2$. This shows that $W_1 \cup W_2$ is not a subspace of $\mathbb{F}^2$.

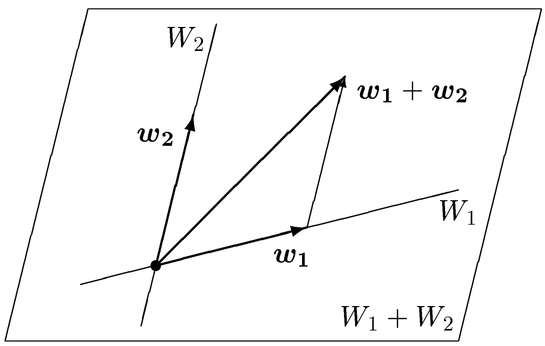
Definition 8.3.11 Let W_1 and W_2 be subspaces of a vector space V. The *sum* of W_1 and W_2 is defined to be the set $W_1 + W_2 = \{\boldsymbol{w_1} + \boldsymbol{w_2} \mid \boldsymbol{w_1} \in W_1 \text{ and } \boldsymbol{w_2} \in W_2\}$.

Theorem 8.3.12 If W_1 and W_2 are subspaces of a vector space V, then $W_1 + W_2$ is a subspace of V.

Proof The proof is left as an exercise. See Question 8.15(a).

Example 8.3.13

1. Let W_1 and W_2 be two nonparallel lines in $\mathbb{R}^3$ such that both lines pass thought the origin. Then $W_1 + W_2$ is the plane that contains both lines. It is obvious that W_1, W_2 and $W_1 + W_2$ are subspaces of $\mathbb{R}^3$.



   In particular, if W_1 is the x-axis in $\mathbb{R}^3$ and W_2 is the y-axis in $\mathbb{R}^3$, then $W_1 + W_2$ is the xy-plane in $\mathbb{R}^3$.

2. Let $\mathbb{F}$ be a field. Define

$$W_1 = \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \middle| \, a \in \mathbb{F} \right\} \quad \text{and} \quad W_2 = \left\{ \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix} \middle| \, a \in \mathbb{F} \right\}.$$

   Then $W_1 + W_2 = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} \middle| \, a, b \in \mathbb{F} \right\}$.

   W_1, W_2 and $W_1 + W_2$ are subspaces of $\mathcal{M}_{2\times 2}(\mathbb{F})$.

Remark 8.3.14 Let W_1 and W_2 be subspaces of a vector space V. Then $W_1 + W_2$ is the smallest subspace of V that contains both W_1 and W_2. Precisely, if U is a subspace of V such that $W_1 \subseteq U$ and $W_2 \subseteq U$, then $W_1 + W_2 \subseteq U$. (See Question 8.15(b).)

Section 8.4 Linear Spans and Linear Independence

Discussion 8.4.1 Most of the discussions about $\mathbb{R}^n$ in Chapter 3 can be rephrased using abstract vector spaces. In the following two sections, we shall study the concepts of linear spans, linear independence and bases using the framework of abstract vector spaces.

Definition 8.4.2 Let V be a vector space and $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m} \in V$. For any scalars $c_1, c_2, \dots, c_m$, the vector

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_m\boldsymbol{v_m}$$

is called a *linear combination* of $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, ..., $\boldsymbol{v_m}$. (Note that we only accept linear combinations using finite number of vectors.)

Theorem 8.4.3 Let V be a vector space over a field $\mathbb{F}$ and let B be a nonempty subset of V. The set of all linear combinations of vectors taken from B,

$$\begin{aligned} W &= \{\boldsymbol{u} \in V \mid \boldsymbol{u} \text{ is a linear combination of some vectors from } B\} \\ &= \{c_1\boldsymbol{v_1} + \cdots + c_m\boldsymbol{v_m} \mid m \in \mathbb{N},\ c_1, \dots, c_m \in \mathbb{F} \text{ and } \boldsymbol{v_1}, \dots, \boldsymbol{v_m} \in B\}, \end{aligned}$$

is a subspace of V.

Proof

(S1) Take any $\boldsymbol{u} \in B$. Since $\boldsymbol{0} = 0\boldsymbol{u}$, $\boldsymbol{0} \in W$.

(S2) Take any $\boldsymbol{u}, \boldsymbol{u'} \in W$, i.e. $\boldsymbol{u} = a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v_m}$ and $\boldsymbol{u'} = b_1\boldsymbol{v'_1} + \cdots + b_n\boldsymbol{v'_n}$ for $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$ and $\boldsymbol{v_1}, \dots, \boldsymbol{v_m}, \boldsymbol{v'_1}, \dots, \boldsymbol{v'_n} \in B$. Then

$$\boldsymbol{u} + \boldsymbol{u'} = a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v_m} + b_1\boldsymbol{v'_1} + \cdots + b_n\boldsymbol{v'_n}$$

which is also a linear combination of vectors from B. So $\boldsymbol{u} + \boldsymbol{u'} \in W$.

(S3) Take any $c \in \mathbb{F}$ and any $\boldsymbol{u} \in W$, i.e. $\boldsymbol{u} = a_1\boldsymbol{v_1} + \cdots + a_m\boldsymbol{v_m}$ for $a_1, \dots, a_m \in \mathbb{F}$ and $\boldsymbol{v_1}, \dots, \boldsymbol{v_m} \in B$. Then

$$c\boldsymbol{u} = ca_1\boldsymbol{v_1} + \cdots + ca_m\boldsymbol{v_m}$$

which is also a linear combination of vectors from B. So $c\boldsymbol{u} \in W$.

As W is a subset of V satisfying (S1), (S2) and (S3), it is a subspace of V.

Definition 8.4.4 Let V be a vector space over a field $\mathbb{F}$ and let B be a nonempty subset of V. The subspace

$$W = \{\boldsymbol{u} \in V \mid \boldsymbol{u} \text{ is a linear combination of some vectors from } B\}$$

in Theorem 8.4.3 is called the *subspace of V spanned* by B and we write $W = \operatorname{span}_{\mathbb{F}}(B)$ or simply $W = \operatorname{span}(B)$ if the field $\mathbb{F}$ is known. Sometimes, we also say that W is a *linear span* of B and B *spans* W. Note that $B \subseteq W$.

In particular if $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots\}$, then

$$W = \{c_1\boldsymbol{v_1} + \cdots + c_m\boldsymbol{v_m} \mid m \in \mathbb{N} \text{ and } c_1, \dots, c_m \in \mathbb{F}\}$$

and we write $W = \operatorname{span}_{\mathbb{F}}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots\}$ or simply $W = \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots\}$ and say that W is the *subspace of V spanned* by the vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots$; W is a *linear span* of the vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots$; and the vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \dots$ *span* W.

Remark 8.4.5 In Definition 8.4.4, if B is finite, say $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}\}$, then

$$\operatorname{span}_{\mathbb{F}}(B) = \{c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} \mid c_1, c_2, \dots, c_k \in \mathbb{F}\}.$$

(Compare with Definition 3.2.3.)

Example 8.4.6

1. Let $\boldsymbol{A_1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{A_2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{A_3} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ be real matrices. Determine whether $\boldsymbol{B}$ is a linear combination of $\boldsymbol{A_1}$, $\boldsymbol{A_2}$ and $\boldsymbol{A_3}$.

   Solution Consider the equation

$$c_1\boldsymbol{A_1} + c_2\boldsymbol{A_2} + c_3\boldsymbol{A_3} = \boldsymbol{B}.$$

   Since

$$c_1\boldsymbol{A_1} + c_2\boldsymbol{A_2} + c_3\boldsymbol{A_3} = \begin{pmatrix} c_1 + c_3 & c_1 + c_2 + c_3 \\ c_1 + c_2 - c_3 & c_1 - c_3 \end{pmatrix},$$

   we have

$$\begin{cases} c_1 + c_3 = 1 \\ c_1 + c_2 + c_3 = 2 \\ c_1 + c_2 - c_3 = 3 \\ c_1 - c_3 = 4. \end{cases}$$

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & -1 & 3 \\ 1 & 0 & -1 & 4 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 2 \end{array}\right)$$

   Since the system is inconsistent, $\boldsymbol{B}$ is not a linear combination of $\boldsymbol{A_1}$, $\boldsymbol{A_2}$ and $\boldsymbol{A_3}$.

2. Let $W_1 = \left\{ \begin{pmatrix} a-b & a+b+2c \\ 2(b+c) & 3(a+c) \end{pmatrix} \middle| \, a, b, c \in \mathbb{R} \right\} \subseteq \mathcal{M}_{2\times 2}(\mathbb{R})$. Since

$$\begin{pmatrix} a-b & a+b+2c \\ 2(b+c) & 3(a+c) \end{pmatrix} = a\begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} + b\begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix},$$

$W_1 = \operatorname{span}\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} \right\}$ and by Theorem 8.4.3, W_1 is a subspace of $\mathcal{M}_{2\times 2}(\mathbb{R})$.

3. Let $W_2 = \{p(x) \in \mathcal{P}_3(\mathbb{R}) \mid p(-1) = 0 \text{ and } p(1) = 0\} \subseteq \mathcal{P}_3(\mathbb{R})$. For any polynomial $p(x) = a + bx + cx^2 + dx^3 \in \mathcal{P}_3(\mathbb{R})$,

$$p(x) \in W_2 \quad \Leftrightarrow \quad \begin{cases} p(-1) = 0 \\ p(1) = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} a-b+c-d = 0 \\ a+b+c+d = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} a = -s \\ b = -t \\ c = s \\ d = t \end{cases} \quad \text{for } s, t \in \mathbb{R},$$

i.e. $p(x) \in W_2$ if and only if $p(x) = -s - tx + sx^2 + tx^3 = s(-1+x^2) + t(-x+x^3)$ for some $s, t \in \mathbb{R}$. Thus $W_2 = \operatorname{span}\{-1 + x^2, \, -x + x^3\}$ and by Theorem 8.4.3, W_2 is a subspace of $\mathcal{P}_3(\mathbb{R})$.

4. Let $p_1(x) = 1 + x + x^2$, $p_2(x) = x + 2x^2$ and $p_3(x) = 2 - x - x^2$ be real polynomials. Prove that $\mathcal{P}_2(\mathbb{R}) = \operatorname{span}\{p_1(x), \, p_2(x), \, p_3(x)\}$.

Solution Since $p_1(x), \, p_2(x), \, p_3(x) \in \mathcal{P}_2(\mathbb{R})$, $\operatorname{span}\{p_1(x), \, p_2(x), \, p_3(x)\} \subseteq \mathcal{P}_2(\mathbb{R})$.

To show $\mathcal{P}_2(\mathbb{R}) \subseteq \operatorname{span}\{p_1(x), \, p_2(x), \, p_3(x)\}$, we only need to show that any polynomial $q(x) = a_1 + a_2 x + a_3 x^2 \in \mathcal{P}_2(\mathbb{R})$ is a linear combination of $p_1(x)$, $p_2(x)$ and $p_3(x)$. Consider the equation

$$c_1 p_1(x) + c_2 p_2(x) + c_3 p_3(x) = q(x).$$

Since

$$\begin{aligned} c_1 p_1(x) + c_2 p_2(x) + c_3 p_3(x) &= c_1(1 + x + x^2) + c_2(x + 2x^2) + c_3(2 - x - x^2) \\ &= (c_1 + 2c_3) + (c_1 + c_2 - c_3)x + (c_1 + 2c_2 - c_3)x^2, \end{aligned}$$

we have

$$\begin{cases} c_1 \phantom{{}+2c_2} + 2c_3 = a_1 \\ c_1 + c_2 - c_3 = a_2 \\ c_1 + 2c_2 - c_3 = a_3. \end{cases}$$

$$\left(\begin{array}{ccc|c} 1 & 0 & 2 & a_1 \\ 1 & 1 & -1 & a_2 \\ 1 & 2 & -1 & a_3 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{3}a_1 + \frac{4}{3}a_2 - \frac{2}{3}a_3 \\ 0 & 1 & 0 & -a_2 + a_3 \\ 0 & 0 & 1 & \frac{1}{3}a_1 - \frac{2}{3}a_2 + \frac{1}{3}a_3 \end{array}\right)$$

The system has a solution $c_1 = \frac{1}{3}a_1 + \frac{4}{3}a_2 - \frac{2}{3}a_3$, $c_2 = -a_2 + a_3$ and $c_3 = \frac{1}{3}a_1 - \frac{2}{3}a_2 + \frac{1}{3}a_3$. It means

$$q(x) = \left(\tfrac{1}{3}a_1 + \tfrac{4}{3}a_2 - \tfrac{2}{3}a_3\right) p_1(x) + (-a_2 + a_3)\, p_2(x) + \left(\tfrac{1}{3}a_1 - \tfrac{2}{3}a_2 + \tfrac{1}{3}a_3\right) p_3(x).$$

As every polynomial in $\mathcal{P}(\mathbb{R})$ is a linear combination of $p_1(x)$, $p_2(x)$ and $p_3(x)$, we have $\mathcal{P}_2(\mathbb{R}) \subseteq \operatorname{span}\{p_1(x),\, p_2(x),\, p_3(x)\}$ and hence $\mathcal{P}_2(\mathbb{R}) = \operatorname{span}\{p_1(x),\, p_2(x),\, p_3(x)\}$.

5. Let $\mathbb{F}$ be a field. For $1 \le i \le n$, let $\boldsymbol{e_i}$ be the vector in $\mathbb{F}^n$ such that its ith coordinate is 1 and all other coordinates are 0, i.e.

$$\boldsymbol{e_i} = (0, 0, \dots, 0, \underset{\substack{\uparrow \\ \text{the } i\text{th coordinate}}}{1}, 0, \dots, 0).$$

For any $\boldsymbol{u} = (u_1, u_2, \dots, u_n) \in \mathbb{F}^n$,

$$\boldsymbol{u} = u_1\boldsymbol{e_1} + u_2\boldsymbol{e_2} + \cdots + u_n\boldsymbol{e_n}.$$

So $\mathbb{F}^n = \operatorname{span}_{\mathbb{F}}\{\boldsymbol{e_1},\, \boldsymbol{e_2},\, \dots,\, \boldsymbol{e_n}\}$.

6. Since every complex number in $\mathbb{C}$ can be written as $a + b\mathrm{i}$ for $a, b \in \mathbb{R}$, $\mathbb{C} = \operatorname{span}_{\mathbb{R}}\{1,\, \mathrm{i}\}$.

Furthermore, using the vectors $\boldsymbol{e_1},\, \boldsymbol{e_2},\, \dots,\, \boldsymbol{e_n}$ in Part 5,

$$\begin{aligned} \mathbb{C}^n &= \operatorname{span}_{\mathbb{C}}\{\boldsymbol{e_1},\, \boldsymbol{e_2},\, \dots,\, \boldsymbol{e_n}\} \\ &= \operatorname{span}_{\mathbb{R}}\{\boldsymbol{e_1},\, \boldsymbol{e_2},\, \dots,\, \boldsymbol{e_n},\, \mathrm{i}\boldsymbol{e_1},\, \mathrm{i}\boldsymbol{e_2},\, \dots,\, \mathrm{i}\boldsymbol{e_n}\}. \end{aligned}$$

7. Let $\mathbb{F}$ be a field. For $1 \le i \le m$ and $1 \le j \le n$, let $\boldsymbol{E_{ij}}$ be the $m \times n$ matrix over $\mathbb{F}$ such that its (i, j)-entry is 1 and all other entries are 0. For any $\boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$,

$$\boldsymbol{A} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\boldsymbol{E_{ij}}.$$

So $\mathcal{M}_{m\times n}(\mathbb{F}) = \operatorname{span}_{\mathbb{F}}\{\boldsymbol{E_{ij}} \mid 1 \le i \le m \text{ and } 1 \le j \le n\}$.

8. Let $\mathbb{F}$ be a field. We have

$$\mathcal{P}(\mathbb{F}) = \{a_0 + a_1x + \cdots + a_mx^m \mid m \in \mathbb{N} \text{ and } a_0, a_1, \dots, a_m \in \mathbb{F}\} = \operatorname{span}_{\mathbb{F}}\{1,\, x,\, x^2,\, \dots\}$$

and

$$\mathcal{P}_n(\mathbb{F}) = \{a_0 + a_1x + \cdots + a_nx^n \mid a_0, a_1, \dots, a_n \in \mathbb{F}\} = \operatorname{span}_{\mathbb{F}}\{1,\, x,\, \dots,\, x^n\}.$$

Remark 8.4.7 In Definition 8.4.2, Theorem 8.4.3 and Definition 8.4.4, we only accept linear combinations using finite number of vectors. For example, the power series

$$\sum_{i=0}^{\infty} x^i = 1 + x + x^2 + \cdots$$

is not contained in $\operatorname{span}\{1,\, x,\, x^2,\, \dots\}$.

Definition 8.4.8 Let V be a vector space over a field $\mathbb{F}$.

1. Let $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k} \in V$.

   (a) The vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}$ are called *linearly independent* if the vector equation

   $$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} = \boldsymbol{0}$$

   has only the trivial solution $c_1 = 0, c_2 = 0, \ldots, c_k = 0$.

   (b) The vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}$ are called *linearly dependent* if they are not linearly independent, i.e. there exists $a_1, a_2, \ldots, a_k \in \mathbb{F}$, not all a_i's are zero, such that $a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_k\boldsymbol{v_k} = \boldsymbol{0}$.

   (Compare this definition with Definition 3.4.2.)

2. Let B be a subset of V.

   (a) B is called *linearly independent* if for every finite subset $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ of B, $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}$ are linearly independent.

   (b) B is called *linearly dependent* if there exists a finite subset $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ of B such that $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}$ are linearly dependent.

(For convenience, whenever we write a set in the form $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$, we always assume that (i) $n \geq 1$; and (ii) $\boldsymbol{u_i} \neq \boldsymbol{u_j}$ for $i \neq j$.)

Remark 8.4.9 Same as the discussion in Section 3.4, linear independence is used to determine whether there are redundant vectors in a set. (See Theorem 3.4.4 and Remark 3.4.5.)

1. If a set B is linearly dependent, at least one vector $\boldsymbol{u} \in B$ can be expressed as a linear combination of other vectors in B and hence $\boldsymbol{u}$ is a redundant vector, i.e. $\mathrm{span}(B - \{\boldsymbol{u}\}) = \mathrm{span}(B)$.

2. If a set B is linearly independent, no vector in the set can be expressed as a linear combination of other vectors in B and hence the set has no redundant vector, i.e. for all $\boldsymbol{u} \in B$, $\mathrm{span}(B - \{\boldsymbol{u}\}) \subsetneqq \mathrm{span}(B)$.

Example 8.4.10

1. Determine whether the subset $\{(1,\ \mathrm{i},\ 1 - \mathrm{i}),\ (\mathrm{i},\ 1,\ 0),\ (3\mathrm{i},\ 1,\ 1 + \mathrm{i})\}$ of $\mathbb{C}^3$ is linearly independent.

Solution To answer the question, we need to solve the vector equation $c_1(1,\ \mathrm{i},\ 1-\mathrm{i}) + c_2(\mathrm{i},\ 1,\ 0) + c_3(3\mathrm{i},\ 1,\ 1+\mathrm{i}) = (0,\ 0,\ 0)$.

$$
\begin{aligned}
& c_1(1,\ \mathrm{i},\ 1-\mathrm{i}) + c_2(\mathrm{i},\ 1,\ 0) + c_3(3\mathrm{i},\ 1,\ 1+\mathrm{i}) = (0,\ 0,\ 0) \\
\Leftrightarrow\quad & (c_1 + \mathrm{i}c_2 + 3\mathrm{i}c_3,\ \mathrm{i}c_1 + c_2 + c_3,\ (1-\mathrm{i})c_1 + (1+\mathrm{i})c_3) = (0,\ 0,\ 0) \\
\Leftrightarrow\quad & \left\{\begin{aligned} c_1 + \mathrm{i}c_2 + 3\mathrm{i}c_3 &= 0 \\ \mathrm{i}c_1 + c_2 + c_3 &= 0 \\ (1-\mathrm{i})c_1 + (1+\mathrm{i})c_3 &= 0 \end{aligned}\right. \\
\Leftrightarrow\quad & \left\{\begin{aligned} c_1 &= -\mathrm{i}t \\ c_2 &= -2t \\ c_3 &= t \end{aligned}\right. \qquad \text{for } t \in \mathbb{C}.
\end{aligned}
$$

Since we have nontrivial solutions, $\{(1,\ \mathrm{i},\ 1-\mathrm{i}),\ (\mathrm{i},\ 1,\ 0),\ (3\mathrm{i},\ 1,\ 1+\mathrm{i})\}$ is linearly dependent.

2. Determine whether the subset $\{1,\ x+x^2,\ 2-x^2,\ x+3x^3\}$ of $\mathcal{P}_3(\mathbb{R})$ is linearly independent.

Solution We need to solve the polynomial equation $c_1 + c_2(x+x^2) + c_3(2-x^2) + c_4(x+3x^3) = 0$ where 0 is the zero polynomial.

$$
\begin{aligned}
& c_1 + c_2(x+x^2) + c_3(2-x^2) + c_4(x+3x^3) = 0 \\
\Leftrightarrow\quad & (c_1 + 2c_3) + (c_2 + c_4)x + (c_2 - c_3)x^2 + 3c_4x^3 = 0 + 0x + 0x^2 + 0x^3 \\
\Leftrightarrow\quad & \left\{\begin{aligned} c_1 + 2c_3 &= 0 \\ c_2 + c_4 &= 0 \\ c_2 - c_3 &= 0 \\ 3c_4 &= 0 \end{aligned}\right. \\
\Leftrightarrow\quad & c_1 = 0,\ c_2 = 0,\ c_3 = 0,\ c_4 = 0.
\end{aligned}
$$

Since we only have the trivial solution, $\{1,\ x+x^2,\ 2-x^2,\ x+3x^3\}$ is linearly independent.

3. Let $f_1,\ f_2,\ f_3 \in C^\infty([-1,1])$ where $f_1(x) = e^x$, $f_2(x) = xe^x$ and $f_3(x) = x$ for $x \in [-1,1]$. Determine whether $f_1,\ f_2,\ f_3$ are linearly independent.

Solution Consider the function equation $c_1f_1 + c_2f_2 + c_3f_3 = \mathit{0}$ where $\mathit{0}$ is the zero function. The equation means

$$c_1e^x + c_2xe^x + c_3x = c_1f_1(x) + c_2f_2(x) + c_3f_3(x) = \mathit{0}(x) = 0 \quad \text{for all } x \in [-1,1].$$

In particular, substituting $x=-1$, $x=0$, $x=1$ into the equation above, we have

$$\left\{\begin{aligned} \tfrac{1}{e}c_1 - \tfrac{1}{e}c_2 - c_3 &= 0 \\ c_1 &= 0 \\ ec_1 + ec_2 + c_3 &= 0 \end{aligned}\right. \quad\Leftrightarrow\quad c_1 = 0,\ c_2 = 0,\ c_3 = 0.$$

Since we only have the trivial solution, $f_1,\ f_2,\ f_3$ are linearly independent.

Section 8.5 Bases and Dimensions

Definition 8.5.1 A subset B of a vector space V is called a *basis* for V if B is linearly independent and B spans V. (See Section 3.5.)

A vector space V is called *finite dimensional* if it has a basis consisting of finitely many vectors; otherwise, V is called *infinite dimensional.*

Remark 8.5.2

1. For convenience, the empty set $\emptyset$ is defined to be the basis for a zero space.
2. Every vector space has a basis. The proof of this requires a fundamental result in set theory called *Zorn's Lemma.*
3. In some infinite dimensional (topological) vector spaces, bases defined in Definition 8.5.1 are called *algebraic bases* or *Hamel bases* in order to distinguish it from other kinds of "bases" where infinite sums are allowed.

Example 8.5.3

1. Let $W_1 = \left\{ \begin{pmatrix} a-b & a+b+2c \\ 2(b+c) & 3(a+c) \end{pmatrix} \middle| \, a, b, c \in \mathbb{R} \right\} \subseteq \mathcal{M}_{2\times 2}(\mathbb{R})$. By Example 8.4.6.2, we have $W_1 = \operatorname{span}\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} \right\}$. Note that
$$c_1 \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} + c_2 \begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} c_1 = -t \\ c_2 = -t \\ c_3 = t \end{cases} \quad \text{for } t \in \mathbb{R}.$$
So $\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} \right\}$ is linearly dependent and hence is not a basis for W_1.
2. Let $W_2 = \{p(x) \in \mathcal{P}_3(\mathbb{R}) \mid p(-1) = 0 \text{ and } p(1) = 0\} \subseteq \mathcal{P}_3(\mathbb{R})$. By Example 8.4.6.3, we have $W_2 = \operatorname{span}\{-1 + x^2, \, -x + x^3\}$. Note that
$$c_1(-1 + x^2) + c_2(-x + x^3) = 0 \quad \Leftrightarrow \quad c_1 = 0, \; c_2 = 0.$$
So $\{-1 + x^2, \, -x + x^3\}$ is linearly independent and hence is a basis for W_2.
3. By Example 8.4.6.5, we have $\mathbb{F}^n = \operatorname{span}_{\mathbb{F}}\{\boldsymbol{e_1}, \, \boldsymbol{e_2}, \, \ldots, \, \boldsymbol{e_n}\}$. Since
$$\begin{aligned} c_1\boldsymbol{e_1} + c_2\boldsymbol{e_2} + \cdots + \boldsymbol{e_n} = \boldsymbol{0} \quad &\Leftrightarrow \quad (c_1, c_2, \ldots, c_n) = (0, 0, \ldots, 0) \\ &\Leftrightarrow \quad c_1 = 0, \; c_2 = 0, \; \cdots, \; c_n = 0, \end{aligned}$$
$\{\boldsymbol{e_1}, \, \boldsymbol{e_2}, \, \ldots, \, \boldsymbol{e_n}\}$ is linearly independent and hence is a basis for $\mathbb{F}^n$. This basis is called the *standard basis for* $\mathbb{F}^n$.

4. By Example 8.4.6.6, $\mathbb{C}^n = \text{span}\{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}, \text{i}\boldsymbol{e_1}, \text{i}\boldsymbol{e_2}, \dots, \text{i}\boldsymbol{e_n}\}$. Note that the set

$$B = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}, \text{i}\boldsymbol{e_1}, \text{i}\boldsymbol{e_2}, \dots, \text{i}\boldsymbol{e_n}\}$$

is linearly dependent over $\mathbb{C}$ but linearly independent over $\mathbb{R}$. So we have the following conclusion:

   (a) B is not a basis for $\mathbb{C}^n$ if $\mathbb{C}^n$ is regarded as a vector space over $\mathbb{C}$.

   (b) B is a basis for $\mathbb{C}^n$ if $\mathbb{C}^n$ is regarded as a vector space over $\mathbb{R}$.

   (We always assume that $\mathbb{C}^n$ is a vector space over $\mathbb{C}$ unless we specify the otherwise.)

5. By Example 8.4.6.7, we have $\mathcal{M}_{m\times n}(\mathbb{F}) = \text{span}_{\mathbb{F}}\{\boldsymbol{E_{ij}} \mid 1 \le i \le m \text{ and } 1 \le j \le n\}$. The set $\{\boldsymbol{E_{ij}} \mid 1 \le i \le m \text{ and } 1 \le j \le n\}$ is linearly independent and hence is a basis for $\mathcal{M}_{m\times n}(\mathbb{F})$. This basis is called the *standard basis for* $\mathcal{M}_{m\times n}(\mathbb{F})$.

6. By Example 8.4.6.8, we have $\mathcal{P}(\mathbb{F}) = \text{span}_{\mathbb{F}}\{1, x, x^2, \dots\}$. The set $\{1, x, x^2, \dots\}$ is linearly independent and hence is a basis for $\mathcal{P}(\mathbb{F})$. This basis is called the *standard basis for* $\mathcal{P}(\mathbb{F})$.

   Also by Example 8.4.6.8, we have $\mathcal{P}_n(\mathbb{F}) = \text{span}_{\mathbb{F}}\{1, x, \dots, x^n\}$. The set $\{1, x, \dots, x^n\}$ is linearly independent and hence is a basis for $\mathcal{P}_n(\mathbb{F})$. This basis is called the *standard basis for* $\mathcal{P}_n(\mathbb{F})$.

Remark 8.5.4

1. The vectors spaces $\mathbb{F}^n$, $\mathcal{M}_{m\times n}(\mathbb{F})$ and $\mathcal{P}_n(\mathbb{F})$ are finite dimensional while $\mathcal{P}(\mathbb{F})$ is infinite dimensional.

2. The vector space $\mathbb{F}^{\mathbb{N}}$ is infinite dimensional. The vector space $\mathcal{F}(A, \mathbb{F})$ is finite dimensional if A is a finite set; and $\mathcal{F}(A, \mathbb{F})$ is infinite dimensional if A is an infinite set.

Lemma 8.5.5 Let V be a finite dimensional vector space and $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ a basis for V. Any vector $\boldsymbol{u} \in V$ can be expressed uniquely as a linear combination

$$\boldsymbol{u} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n}$$

where $c_1, c_2, \dots, c_n$ are scalars.

(See Question 8.28 for the infinite dimensional version of the lemma.)

Proof The proof follows the same argument as the proof for Theorem 3.5.7.

Definition 8.5.6 Let V be a finite dimensional vector space over a field $\mathbb{F}$ where V is not a zero space.

1. A basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ for V is called an *ordered basis* if the vectors in B have a fixed order such that $\boldsymbol{v_1}$ is the first vector, $\boldsymbol{v_2}$ is the second vector, etc.

2. Let $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ be an ordered basis for V and let $\boldsymbol{u} \in V$. If
$$\boldsymbol{u} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n} \quad \text{for } c_1, c_2, \ldots, c_n \in \mathbb{F},$$
then the coefficients $c_1, c_2, \ldots, c_n$ are called the *coordinates* of $\boldsymbol{u}$ relative to the basis B. The vector
$$(\boldsymbol{u})_B = (c_1, c_2, \ldots, c_n) \quad \text{or} \quad [\boldsymbol{u}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$
in $\mathbb{F}^n$ is called the *coordinate vector* of $\boldsymbol{u}$ relative to the basis B.

Lemma 8.5.7 Let V be a finite dimensional vector space over a field $\mathbb{F}$, where V is not a zero space, and let B be an ordered basis for V.

1. For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} = \boldsymbol{v}$ if and only if $(\boldsymbol{u})_B = (\boldsymbol{v})_B$.

2. For any $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r} \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{F}$,
$$(c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_B = c_1(\boldsymbol{v_1})_B + c_2(\boldsymbol{v_2})_B + \cdots + c_r(\boldsymbol{v_r})_B.$$

Proof The proof is left as exercise. See Question 8.24

Example 8.5.8

1. Let $B_1 = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ where $\boldsymbol{v_1} = (1, 1, 0)$, $\boldsymbol{v_2} = (0, 1, 1)$ and $\boldsymbol{v_3} = (1, 1, 1)$ are vectors in $\mathbb{F}_2^3$. Note that B_1 is a basis for $\mathbb{F}_2^3$. Using B_1 as an ordered basis, find the coordinate vector of $\boldsymbol{u} = (a, b, c) \in \mathbb{F}_2^3$ relative to B_1.

   Solution (Recall that in $\mathbb{F}_2$, $1 + 1 = 0$.)
$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + c_3\boldsymbol{v_3} = \boldsymbol{u} \quad \Leftrightarrow \quad \begin{cases} c_1 \phantom{{}+c_2} + c_3 = a \\ c_1 + c_2 + c_3 = b \\ \phantom{c_1 +{}} c_2 + c_3 = c \end{cases} \quad \Leftrightarrow \quad \begin{cases} c_1 = b + c \\ c_2 = a + b \\ c_3 = a + b + c. \end{cases}$$
   Thus $(\boldsymbol{u})_{B_1} = (b + c,\ a + b,\ a + b + c)$.

2. Let $B_2 = \{\boldsymbol{A_1}, \boldsymbol{A_2}, \boldsymbol{A_3}, \boldsymbol{A_4}\}$ where $\boldsymbol{A_1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{A_2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{A_3} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ and $\boldsymbol{A_4} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are real matrices. Note that B_2 is a basis for $\mathcal{M}_{2\times 2}(\mathbb{R})$. Using B_2 as an ordered basis, find the coordinate vector of $\boldsymbol{C} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$ relative to B_2.

Solution

$$c_1\boldsymbol{A_1} + c_2\boldsymbol{A_2} + c_3\boldsymbol{A_3} + c_4\boldsymbol{A_4} = \boldsymbol{C} \quad \Leftrightarrow \quad \begin{cases} c_1 \phantom{{}+c_2} + c_3 + c_4 = 1 \\ c_1 + c_2 + c_3 \phantom{{}+c_4} = 2 \\ c_1 + c_2 - c_3 \phantom{{}+c_4} = 4 \\ c_1 \phantom{{}+c_2} - c_3 - c_4 = 3 \end{cases} \quad \Leftrightarrow \quad \begin{cases} c_1 = 2 \\ c_2 = 1 \\ c_3 = -1 \\ c_4 = 0. \end{cases}$$

Thus $(\boldsymbol{C})_{B_2} = (2,\ 1,\ -1,\ 0)$.

Remark 8.5.9 Let V be a finite dimensional vector space over a field $\mathbb{F}$ such that V has a basis B with n vectors. Using the coordinate system relative to B, we can translate all vectors in V to vectors in $\mathbb{F}^n$. Thus all problems about V can be solved by using theorems and methods worked for $\mathbb{F}^n$.

By this way, most of the results in Chapter 3 and Chapter 4 for Euclidean n-spaces can be applied to other finite dimensional vector spaces.

Theorem 8.5.10 Let V be a vector space which has a basis with n vectors. Then

1. any subset of V with more than n vectors is always linearly dependent; and
2. any subset of V with less than n vectors cannot span V.

Hence every basis for V has n vectors.

Proof The proof follows the same argument as the proof for Theorem 3.6.1.

Definition 8.5.11 The *dimension* of a finite dimensional vector space V over a field $\mathbb{F}$, denoted by $\dim_{\mathbb{F}}(V)$ or simply $\dim(V)$, is defined to be the number of vectors in a basis for V. In addition, we define the dimension of a zero space to be zero. (See also Section 3.6.)

Example 8.5.12

1. $\dim_{\mathbb{F}}(\mathbb{F}^n) = n$.
2. $\dim_{\mathbb{C}}(\mathbb{C}^n) = n$ and $\dim_{\mathbb{R}}(\mathbb{C}^n) = 2n$.
3. $\dim_{\mathbb{F}}(\mathcal{M}_{m\times n}(\mathbb{F})) = mn$.
4. $\dim_{\mathbb{F}}(\mathcal{P}_n(\mathbb{F})) = n + 1$.
5. In Example 8.3.9, $\dim(W_1) = \dim(W_2) = \frac{1}{2}n(n+1)$ and $\dim(W_1 \cap W_2) = n$. (Why?)

Theorem 8.5.13 Let V be a finite dimensional vector space and B a subset of V. The following are equivalent:

1. B is a basis for V.

2. B is linearly independent and $|B| = \dim(V)$.

3. B spans V and $|B| = \dim(V)$.

Proof The proof follows the same argument as the proof for Theorem 3.6.7.

Example 8.5.14 In Example 8.4.6.4, we have $\mathcal{P}_2(\mathbb{R}) = \operatorname{span}\{p_1(x), p_2(x), p_3(x)\}$ where $p_1(x) = 1 + x + x^2$, $p_2(x) = x + 2x^2$ and $p_3(x) = 2 - x - x^2$. Since $\dim(\mathcal{P}_2(\mathbb{R})) = 3$, by Theorem 8.5.13, $\{p_1(x), p_2(x), p_3(x)\}$ is a basis for $\mathcal{P}_2(\mathbb{R})$.

Theorem 8.5.15 Let W be a subspace of a finite dimensional vector space V. Then

1. $\dim(W) \leq \dim(V)$; and

2. if $\dim(W) = \dim(V)$, then $W = V$.

Proof The proof follows the same argument as the proof for Theorem 3.6.9.

Example 8.5.16 For this example, we modify the algorithms in Example 4.1.14 and use them to find bases for finite dimensional vector spaces:

Find a basis for the subspace W of $\mathcal{M}_{2\times 2}(\mathbb{R})$ spanned by

$$\boldsymbol{A_1} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \boldsymbol{A_2} = \begin{pmatrix} 3 & 6 \\ 6 & 3 \end{pmatrix}, \quad \boldsymbol{A_3} = \begin{pmatrix} 4 & 9 \\ 9 & 5 \end{pmatrix},$$
$$\boldsymbol{A_4} = \begin{pmatrix} -2 & -1 \\ -1 & 1 \end{pmatrix}, \quad \boldsymbol{A_5} = \begin{pmatrix} 5 & 8 \\ 9 & 4 \end{pmatrix}, \quad \boldsymbol{A_6} = \begin{pmatrix} 4 & 2 \\ 7 & 3 \end{pmatrix}.$$

Solution Use the (ordered) standard basis $E = \{\boldsymbol{E_{11}}, \boldsymbol{E_{12}}, \boldsymbol{E_{21}}, \boldsymbol{E_{22}}\}$ for $\mathcal{M}_{2\times 2}(\mathbb{R})$ where

$$\boldsymbol{E_{11}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_{12}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_{21}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{E_{22}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$(\boldsymbol{A_1})_E = (1, 2, 2, 1), \quad (\boldsymbol{A_2})_E = (3, 6, 6, 3), \quad (\boldsymbol{A_3})_E = (4, 9, 9, 5),$$
$$(\boldsymbol{A_4})_E = (-2, -1, -1, 1), \quad (\boldsymbol{A_5})_E = (5, 8, 9, 4), \quad (\boldsymbol{A_6})_E = (4, 2, 7, 3).$$

Use Method 1 of Example 4.1.14.1:

$$\begin{pmatrix} 1 & 2 & 2 & 1 \\ 3 & 6 & 6 & 3 \\ 4 & 9 & 9 & 5 \\ -2 & -1 & -1 & 1 \\ 5 & 8 & 9 & 4 \\ 4 & 2 & 7 & 3 \end{pmatrix} \xrightarrow[\text{Elimination}]{\text{Gaussian}} \begin{pmatrix} 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So $\{(1,2,2,1),\ (0,1,1,1),\ (0,0,1,1)\}$ is a basis for the subspace of $\mathbb{R}^4$ spanned by $(\boldsymbol{A_1})_E$, $(\boldsymbol{A_2})_E$, $(\boldsymbol{A_3})_E$, $(\boldsymbol{A_4})_E$, $(\boldsymbol{A_5})_E$, $(\boldsymbol{A_6})_E$. Let $\boldsymbol{B_1}$, $\boldsymbol{B_2}$, $\boldsymbol{B_3}$ be 2×2 real matrices such that

$$(\boldsymbol{B_1})_E = (1,2,2,1) \quad \Rightarrow \quad \boldsymbol{B_1} = \boldsymbol{E_{11}} + 2\boldsymbol{E_{12}} + 2\boldsymbol{E_{21}} + \boldsymbol{E_{22}} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},$$

$$(\boldsymbol{B_2})_E = (0,1,1,1) \quad \Rightarrow \quad \boldsymbol{B_2} = 0\boldsymbol{E_{11}} + \boldsymbol{E_{12}} + \boldsymbol{E_{21}} + \boldsymbol{E_{22}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

$$(\boldsymbol{B_3})_E = (0,0,1,1) \quad \Rightarrow \quad \boldsymbol{B_3} = 0\boldsymbol{E_{11}} + 0\boldsymbol{E_{12}} + \boldsymbol{E_{21}} + \boldsymbol{E_{22}} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then $\{\boldsymbol{B_1},\ \boldsymbol{B_2},\ \boldsymbol{B_3}\} = \left\{ \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \right\}$ is a basis for W.

Use Method 2 of Example 4.1.14.1:

$$\begin{pmatrix} 1 & 3 & 4 & -2 & 5 & 4 \\ 2 & 6 & 9 & -1 & 8 & 2 \\ 2 & 6 & 9 & -1 & 9 & 7 \\ 1 & 3 & 5 & 1 & 4 & 3 \end{pmatrix} \xrightarrow[\text{Elimination}]{\text{Gaussian}} \begin{pmatrix} 1 & 3 & 4 & -2 & 5 & 4 \\ 0 & 0 & 1 & 3 & -2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since the 1st, 2nd and 5th columns are pivot columns of the row echelon form, $\{\,[\boldsymbol{A_1}]_E,\ [\boldsymbol{A_3}]_E,\ [\boldsymbol{A_5}]_E\}$ is a basis for the subspace of $\mathbb{R}^4$ spanned by $[\boldsymbol{A_1}]_E$, $[\boldsymbol{A_2}]_E$, $[\boldsymbol{A_3}]_E$, $[\boldsymbol{A_4}]_E$, $[\boldsymbol{A_5}]_E$, $[\boldsymbol{A_6}]_E$. Thus $\{\boldsymbol{A_1},\ \boldsymbol{A_3},\ \boldsymbol{A_5}\}$ is a basis for W.

Theorem 8.5.17 Let V be a finite dimensional vector space. Suppose C is a linearly independent subset of V. Then there exists a basis B for V such that $C \subseteq B$.

Proof If $\operatorname{span}(C) = V$, then $B = C$ is a basis for V. Suppose $\operatorname{span}(C) \subsetneqq V$. There exists $\boldsymbol{u} \in V$ but $\boldsymbol{u} \notin \operatorname{span}(C)$. Let $C_1 = C \cup \{\boldsymbol{u}\}$. Note that C_1 is linearly independent (check it). If $\operatorname{span}(C_1) = V$, then $B = C_1$ is a basis for V. If not, we repeat the process above to find a new vector in V but not in $\operatorname{span}(C_1)$. Since V is finite dimensional, by Theorem 8.5.13, we shall eventually get enough vectors to form a basis B for V.

Example 8.5.18 Let

$$C = \{1 + 4x - 2x^2 + 5x^3 + x^4,\ \ 2 + 9x - x^2 + 8x^3 + 2x^4,\ \ 2 + 9x - x^2 + 9x^3 + 3x^4\}$$

which is a linearly independent set in $\mathcal{P}_4(\mathbb{R})$. Extend C to a basis for $\mathcal{P}_4(\mathbb{R})$.

Solution Use the standard basis $E = \{1, x, x^2, x^3, x^4\}$ for $\mathcal{P}_4(\mathbb{R})$. Then

$$\begin{aligned} (1 + 4x - 2x^2 + 5x^3 + x^4)_E &= (1, 4, -2, 5, 1), \\ (2 + 9x - x^2 + 8x^3 + 2x^4)_E &= (2, 9, -1, 8, 2), \\ (2 + 9x - x^2 + 9x^3 + 3x^4)_E &= (2, 9, -1, 9, 3). \end{aligned}$$

We use the algorithm in Example 4.1.14.2:

$$\begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 2 & 9 & -1 & 8 & 2 \\ 2 & 9 & -1 & 9 & 3 \end{pmatrix} \xrightarrow[\text{Elimination}]{\text{Gaussian}} \begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 0 & 1 & 3 & -2 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Since the third and fifth columns are non-pivot columns of the row-echelon form on the right,

$$\{(1,4,-2,5,1),\ (2,9,-1,8,2),\ (2,9,-1,9,3),\ (0,0,1,0,0),\ (0,0,0,0,1)\}$$

is a basis for $\mathbb{R}^5$. Thus

$$B = \{1+4x-2x^2+5x^3+x^4,\ \ 2+9x-x^2+8x^3+2x^4,\ \ 2+9x-x^2+9x^3+3x^4,\ \ x^2,\ \ x^4\}$$

is a basis for $\mathcal{P}_4(\mathbb{R})$.

Section 8.6 Direct Sums of Subspaces

Discussion 8.6.1 Let W_1 and W_2 be subspaces of a vector space V. In Theorem 8.3.12, we learn that the sum of W_1 and W_2,

$$W_1 + W_2 = \{\boldsymbol{w_1} + \boldsymbol{w_2} \mid \boldsymbol{w_1} \in W_1 \text{ and } \boldsymbol{w_2} \in W_2\},$$

is a subspace of V. Sometimes, in order to study the behavior of a large vector space, it is more convenient to decompose the space into sums of smaller subspaces. (For example, see Chapter 11.) To do so, we need to make sure each vector in $W_1 + W_2$ is expressed uniquely as $\boldsymbol{w_1} + \boldsymbol{w_2}$ with $\boldsymbol{w_1} \in W_1$ and $\boldsymbol{w_2} \in W_2$.

Example 8.6.2 Let W_1 be the xy-plane and W_2 the yz-plane. Then $W_1 + W_2 = \mathbb{R}^3$. Take $(1,2,3) \in \mathbb{R}^3$. We have

$$(1,2,3) = (1,1,0) + (0,1,3) = (1,2,0) + (0,0,3)$$

where $(1,1,0)$, $(1,2,0) \in W_1$ and $(0,1,3)$, $(0,0,3) \in W_2$. So there are more than one way to write $(1,2,3)$ as $\boldsymbol{w_1} + \boldsymbol{w_2}$ with $\boldsymbol{w_1} \in W_1$ and $\boldsymbol{w_2} \in W_2$.

Let W_3 be the z-axis. If we replace W_2 by W_3, we still have $W_1 + W_3 = \mathbb{R}^3$. Now, every vector $(a,b,c) \in \mathbb{R}^3$ is written uniquely as

$$(a,b,c) = (a,b,0) + (0,0,c)$$

with $(a,b,0) \in W_1$ and $(0,0,c) \in W_3$.

Definition 8.6.3 Let W_1 and W_2 be subspaces of a vector space V. We say that the subspace $W_1 + W_2$ is a *direct sum* of W_1 and W_2 if every vector $\boldsymbol{u} \in W_1 + W_2$ can be expressed uniquely as

$$\boldsymbol{u} = \boldsymbol{w_1} + \boldsymbol{w_2} \quad \text{where } \boldsymbol{w_1} \in W_1 \text{ and } \boldsymbol{w_2} \in W_2.$$

In this case, we denote $W_1 + W_2$ by $W_1 \oplus W_2$.

Note that as a set, $W_1 \oplus W_2 = W_1 + W_2 = \{\boldsymbol{w_1} + \boldsymbol{w_2} \mid \boldsymbol{w_1} \in W_1 \text{ and } \boldsymbol{w_2} \in W_2\}$. The "circle" added to "+" can be regarded as a remark saying that the sum $W_1 + W_2$ is a direct sum.

Example 8.6.4 In Example 8.6.2, $\mathbb{R}^3 = W_1 + W_2$ but it is not a direct sum. By replacing W_2 by W_3, we have $\mathbb{R}^3 = W_1 \oplus W_3$, i.e. $\mathbb{R}^3$ is a direct sum of W_1 and W_3.

Theorem 8.6.5 Let W_1 and W_2 be subspaces of a vector space V. Then $W_1 + W_2$ is a direct sum if and only if $W_1 \cap W_2 = \{\boldsymbol{0}\}$.

Proof

($\Rightarrow$) Take any $\boldsymbol{w} \in W_1 \cap W_2$. Note that $\boldsymbol{0} = \boldsymbol{w} + (-\boldsymbol{w})$ where $\boldsymbol{w} \in W_1 \cap W_2 \subseteq W_1$ and $-\boldsymbol{w} \in W_1 \cap W_2 \subseteq W_2$.

On the other hand, $\boldsymbol{0} = \boldsymbol{0} + \boldsymbol{0}$ where $\boldsymbol{0} \in W_1$ and $\boldsymbol{0} \in W_2$. As $W_1 + W_2$ is a direct sum, $\boldsymbol{0}$ can only be written uniquely as $\boldsymbol{w_1} + \boldsymbol{w_2}$ for $\boldsymbol{w_1} \in W_1$ and $\boldsymbol{w_2} \in W_2$. So $\boldsymbol{w} = \boldsymbol{0}$.

Thus $W_1 \cap W_2 = \{\boldsymbol{0}\}$.

($\Leftarrow$) Suppose a vector $\boldsymbol{u} \in W_1 + W_2$ can be written as

$$\boldsymbol{u} = \boldsymbol{w_1} + \boldsymbol{w_2} \quad \text{and} \quad \boldsymbol{u} = \boldsymbol{w_1'} + \boldsymbol{w_2'}$$

where $\boldsymbol{w_1}, \boldsymbol{w_1'} \in W_1$ and $\boldsymbol{w_2}, \boldsymbol{w_2'} \in W_2$. Then $\boldsymbol{w_1} + \boldsymbol{w_2} = \boldsymbol{w_1'} + \boldsymbol{w_2'}$ implies

$$\boldsymbol{w_1} - \boldsymbol{w_1'} = \boldsymbol{w_2'} - \boldsymbol{w_2}.$$

Since W_1 is a subspace and $\boldsymbol{w_1}, \boldsymbol{w_1'} \in W_1$, $\boldsymbol{w_1} - \boldsymbol{w_1'} \in W_1$. Similarly, $\boldsymbol{w_2'} - \boldsymbol{w_2} \in W_2$. It follows that $\boldsymbol{w_1} - \boldsymbol{w_1'} = \boldsymbol{w_2'} - \boldsymbol{w_2} \in W_1 \cap W_2$. As $W_1 \cap W_2 = \{\boldsymbol{0}\}$, we conclude that

$$\boldsymbol{w_1} - \boldsymbol{w_1'} = \boldsymbol{w_2'} - \boldsymbol{w_2} = \boldsymbol{0}$$

and hence $\boldsymbol{w_1} = \boldsymbol{w_1'}$ and $\boldsymbol{w_2} = \boldsymbol{w_2'}$. So the sum is unique.

We have shown that $W_1 + W_2$ is a direct sum.

Example 8.6.6

1. In Example 8.6.2, $W_1 \cap W_2$ is the y-axis while $W_1 \cap W_3 = \{(0, 0, 0)\}$. By Theorem 8.6.5, $W_1 + W_2$ is not a direct sum and $W_1 + W_3$ is a direct sum.

2. Let

$$W_1 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\} \quad \text{and} \quad W_2 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{A}\}.$$

Both W_1 and W_2 are subspaces of $\mathcal{M}_{n\times n}(\mathbb{R})$. So $W_1 + W_2 \subseteq \mathcal{M}_{n\times n}(\mathbb{R})$.

For any matrix $\boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$,

$$\boldsymbol{B} = \tfrac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) + \tfrac{1}{2}(\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}})$$

where $\frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) \in W_1$ and $\frac{1}{2}(\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}) \in W_2$. This means $\boldsymbol{B} \in W_1 + W_2$ for all $\boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$. So $\mathcal{M}_{n\times n}(\mathbb{R}) \subseteq W_1 + W_2$.

Thus we have shown $W_1 + W_2 = \mathcal{M}_{n\times n}(\mathbb{R})$.

Furthermore, $W_1 \cap W_2 = \{\boldsymbol{0}\}$. By Theorem 8.6.5, $W_1 + W_2$ is a direct sum, i.e. $\mathcal{M}_{n\times n}(\mathbb{R}) = W_1 \oplus W_2$.

In this example, elements of W_1 are symmetric matrices while elements of W_2 are called *skew symmetric matrices* or *anti-symmetric matrices*.

3. Let W_1 be the subspace of $C([0, 2\pi])$ spanned by $g \in C([0, 2\pi])$ where $g(x) = \sin(x)$ for $x \in [0, 2\pi]$. Define

$$W_2 = \left\{ f \in C([0, 2\pi]) \;\middle|\; \int_0^{2\pi} f(t)\sin(t)\,dt = 0 \right\}.$$

Then W_2 is a subspace of $C([0, 2\pi])$ and $C([0, 2\pi]) = W_1 \oplus W_2$. (We leave the verification of the results as exercise. See Question 8.34.)

Theorem 8.6.7 Let W_1 and W_2 be subspaces of a vector space V. Suppose $W_1 + W_2$ is a direct sum, i.e. $W_1 \cap W_2 = \{\boldsymbol{0}\}$.

1. If B_1 and B_2 are bases for W_1 and W_2 respectively, then $B_1 \cup B_2$ is a basis for $W_1 \oplus W_2$.
2. If both W_1 and W_2 are finite dimensional, then

$$\dim(W_1 \oplus W_2) = \dim(W_1) + \dim(W_2).$$

(See Question 8.35 for a formula of $\dim(W_1 + W_2)$ when $W_1 + W_2$ is not a direct sum.)

Proof

1. (i) It is obvious that $\mathrm{span}(B_1 \cup B_2) \subseteq W_1 \oplus W_2$.

Take any $\boldsymbol{u} \in W_1 \oplus W_2$, i.e. $\boldsymbol{u} = \boldsymbol{w_1} + \boldsymbol{w_2}$ where $\boldsymbol{w_1} \in W_1$ and $\boldsymbol{w_2} \in W_2$. Since B_1 and B_2 span W_1 and W_2 respectively, there exists, $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \in B_1$ and $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m} \in B_2$ such that

$$\boldsymbol{w_1} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} \quad \text{and} \quad \boldsymbol{w_2} = b_1\boldsymbol{v_1} + b_2\boldsymbol{v_2} + \cdots + b_m\boldsymbol{v_m}$$

for some scalars $a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_m$. Then

$$\boldsymbol{u} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} + b_1\boldsymbol{v_1} + b_2\boldsymbol{v_2} + \cdots + b_m\boldsymbol{v_m}$$

is a linear combination of vectors from $B_1 \cup B_2$. So $W_1 \oplus W_2 \subseteq \text{span}(B_1 \cup B_2)$.

Thus we have shown that $\text{span}(B_1 \cup B_2) = W_1 \oplus W_2$.

(ii) By Definition 8.4.8, to show $B_1 \cup B_2$ is linearly independent, we need to show that every finite subset of $B_1 \cup B_2$ is linearly independent.

Take any finite set $C = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}, \boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}\}$ where $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \in B_1$ and $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m} \in B_2$. Consider the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} + d_1\boldsymbol{v_1} + d_2\boldsymbol{v_2} + \cdots + d_m\boldsymbol{v_m} = \boldsymbol{0}. \qquad (8.2)$$

Set $\boldsymbol{w} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$. Note that $\boldsymbol{w} \in \text{span}(B_1) = W_1$. By (8.2),

$$\boldsymbol{w} = -d_1\boldsymbol{v_1} - d_2\boldsymbol{v_2} - \cdots - d_m\boldsymbol{v_m} \in \text{span}(B_2) = W_2$$

and hence $\boldsymbol{w} \in W_1 \cap W_2$. Since $W_1 \cap W_2 = \{\boldsymbol{0}\}$, $\boldsymbol{w} = \boldsymbol{0}$. This means

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{0} \quad \text{and} \quad -d_1\boldsymbol{v_1} - d_2\boldsymbol{v_2} - \cdots - d_m\boldsymbol{v_m} = \boldsymbol{0}.$$

As B_1 and B_2 are linearly independent, the two equations above have only the trivial solutions $c_1 = 0$, $c_2 = 0$, $\ldots$, $c_k = 0$, $d_1 = 0$, $d_2 = 0$, $\ldots$, $d_m = 0$. Thus the equation (8.2) has only the trivial solution and C is linearly independent.

We have shown that $B_1 \cup B_2$ is linearly independent.

By (i) and (ii), $B_1 \cup B_2$ is a basis for $W_1 \oplus W_2$.

2. Since $W_1 \cap W_2 = \{\boldsymbol{0}\}$, $B_1 \cap B_2 = \emptyset$. So by Part 1, $\dim(W_1 \oplus W_2) = |B_1 \cup B_2| = |B_1| + |B_2| = \dim(W_1) + \dim(W_2)$.

Remark 8.6.8 In Theorem 8.6.7.1, suppose $W_1 + W_2$ is not a direct sum, i.e. $W_1 \cap W_2 \neq \{\boldsymbol{0}\}$. It is still true that $\text{span}(B_1 \cup B_2) = W_1 + W_2$ but $B_1 \cup B_2$ may not be linearly independent and hence $B_1 \cup B_2$ may not be a basis for $W_1 + W_2$.

For example, let W_1 be the xy-plane and W_2 the yz-plane (see Example 8.6.2). Take bases $B_1 = \{(1, 0, 0),\ (0, 1, 0)\}$ and $B_2 = \{(0, 1, 1),\ (0, 0, 1)\}$ for W_1 and W_2 respectively. It is obvious that $\text{span}(B_1 \cup B_2) = W_1 + W_2$ but $B_1 \cup B_2$ is linearly dependent and hence $B_1 \cup B_2$ is not be a basis for $W_1 + W_2$.

Definition 8.6.9 Let V be a vector space and W_1, W_2 ..., W_k subspaces of V.

1. The *sum* of W_1, W_2 ..., W_k is defined to be

$$W_1 + W_2 + \cdots + W_k = \{\boldsymbol{w_1} + \boldsymbol{w_2} + \cdots + \boldsymbol{w_k} \mid \boldsymbol{w_i} \in W_i \text{ for } i = 1, 2, \ldots, k\},$$

   which is a subspace of V.

2. The subspace $W_1 + W_2 + \cdots + W_k$ is said to be a *direct sum* of W_1, W_2 ..., W_k if every vector $\boldsymbol{u} \in W_1 + W_2 + \cdots + W_k$ can be expressed uniquely as

$$\boldsymbol{u} = \boldsymbol{w_1} + \boldsymbol{w_2} + \cdots + \boldsymbol{w_k} \quad \text{where } \boldsymbol{w_i} \in W_i \text{ for } i = 1, 2, \ldots, k.$$

   In this case, we shall write the sum $W_1 + W_2 + \cdots + W_k$ as $W_1 \oplus W_2 \oplus \cdots \oplus W_k$.

Remark 8.6.10 We can use Theorem 8.6.5 repeatedly to determine whether the sum $W_1 + W_2 + \cdots + W_k$ is a direct sum. For example, check the following one by one:

$$W_1 \cap W_2 = \{\mathbf{0}\}, \quad (W_1 + W_2) \cap W_3 = \{\mathbf{0}\}, \quad \ldots, \quad (W_1 + \cdots + W_{k-1}) \cap W_k = \{\mathbf{0}\}.$$

Example 8.6.11

1. Let W_1, W_2 and W_3 be the x-axis, the y-axis and the z-axis in $\mathbb{R}^3$ respectively. Since every $(a, b, c) \in \mathbb{R}^3$ can be expressed uniquely as

$$(a, b, c) = (a, 0, 0) + (0, b, 0) + (0, 0, c)$$

   where $(a, 0, 0) \in W_1$, $(0, b, 0) \in W_2$ and $(0, 0, c) \in W_3$, $\mathbb{R}^3$ is a direct sum of W_1, W_2, W_3, i.e. $\mathbb{R}^3 = W_1 \oplus W_2 \oplus W_3$.

   Alternatively, we can use the method discussed in Remark 8.6.10. It is obvious that $W_1 \cap W_2 = \{(0, 0, 0)\}$. Since $W_1 + W_2$ is the xy-plane, $(W_1 + W_2) \cap W_3 = \{(0, 0, 0)\}$. Thus $W_1 + W_2 + W_3$ is a direct sum.

2. Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ be a basis for V. For each $i = 1, 2, \ldots, n$, define $W_i = \operatorname{span}\{\boldsymbol{v_i}\}$. Since each $\boldsymbol{u} \in V$ can be expressed uniquely as $\boldsymbol{u} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n}$ for $c_1, c_2, \ldots, c_n \in \mathbb{F}$,

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_n.$$

   The example in Part 1 is a particular case of this example.

Section 8.7 Cosets and Quotient Spaces

Definition 8.7.1 Let W be a subspace of a vector space V. For $\boldsymbol{u} \in V$, the set

$$W + \boldsymbol{u} = \{\boldsymbol{w} + \boldsymbol{u} \mid \boldsymbol{w} \in W\}$$

is called the *coset of W containing $\boldsymbol{u}$*.

Example 8.7.2

1. Let W be the subspace of $\mathbb{F}_2^3$ spanned by $(1, 0, 1)$ and $(0, 1, 1)$. Then

   $$W = \{a(1, 0, 1) + b(0, 1, 1) \mid a, b \in \mathbb{F}_2\} = \{(0, 0, 0),\ (1, 0, 1),\ (0, 1, 1), (1, 1, 0)\}.$$

   The following are all the cosets of W:

   $$\begin{aligned} W + (0, 0, 0) &= \{(0, 0, 0),\ (1, 0, 1),\ (0, 1, 1), (1, 1, 0)\}, \\ W + (0, 0, 1) &= \{(0, 0, 1),\ (1, 0, 0),\ (0, 1, 0), (1, 1, 1)\}, \\ W + (0, 1, 0) &= \{(0, 1, 0),\ (1, 1, 1),\ (0, 0, 1), (1, 0, 0)\}, \\ W + (0, 1, 1) &= \{(0, 1, 1),\ (1, 1, 0),\ (0, 0, 0), (1, 0, 1)\}, \\ W + (1, 0, 0) &= \{(1, 0, 0),\ (0, 0, 1),\ (1, 1, 1), (0, 1, 0)\}, \\ W + (1, 0, 1) &= \{(1, 0, 1),\ (0, 0, 0),\ (1, 1, 0), (0, 1, 1)\}, \\ W + (1, 1, 0) &= \{(1, 1, 0),\ (0, 1, 1),\ (1, 0, 1), (0, 0, 0)\}, \\ W + (1, 1, 1) &= \{(1, 1, 1),\ (0, 1, 0),\ (1, 0, 0), (0, 0, 1)\}. \end{aligned}$$

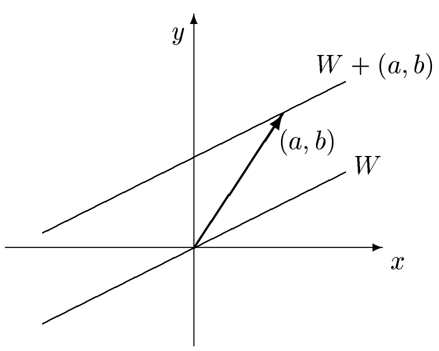
   Note that $W + (0, 0, 0) = W + (0, 1, 1) = W + (1, 0, 1) = W + (1, 1, 0) = W$ and $W + (0, 0, 1) = W + (0, 1, 0) = W + (1, 0, 0) = W + (1, 1, 1) = \mathbb{F}_2^3 - W$.

2. Let $W = \{(x, y) \in \mathbb{R}^2 \mid x - 2y = 0\}$. It is the line in $\mathbb{R}^2$ represented by the homogeneous linear equation $x - 2y = 0$ and is a subspace of $\mathbb{R}^2$.

   Take any $(a, b) \in \mathbb{R}^2$. The coset of W containing (a, b) is

   $$\begin{aligned} &W + (a, b) \\ &= \{(x, y) + (a, b) \mid (x, y) \in \mathbb{R}^2 \text{ and } x - 2y = 0\} \\ &= \{(x', y') \in \mathbb{R}^2 \mid x' - 2y' = a - 2b\} \end{aligned}$$

   which is the line in $\mathbb{R}^2$ represented by the linear equation $x - 2y = c$ where $c = a - 2b$. Note that the line $x - 2y = c$ is parallel to $x - 2y = 0$.

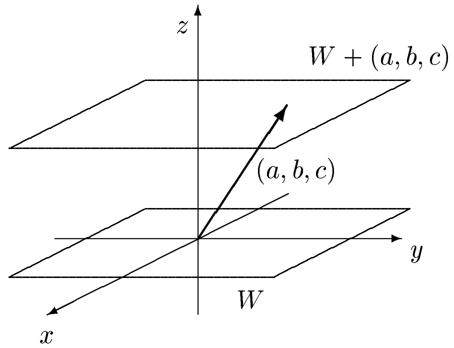
   In general, let W be a line in $\mathbb{R}^2$ (or $\mathbb{R}^3$) that passes through the origin. The cosets of W are lines in $\mathbb{R}^2$ (or $\mathbb{R}^3$) parallel to W. (See also Discussion 3.2.15.1.)

3. Let $W = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$. It is the xy-plane in $\mathbb{R}^3$ represented by the homogeneous linear equation $z = 0$ and is a subspace of $\mathbb{R}^2$.

   Take any $(a, b, c) \in \mathbb{R}^3$. Same as the example in Part 2, the coset of W containing (a, b, c) is

   $$W + (a, b, c) = \{(x', y', z') \in \mathbb{R}^3 \mid z' = c\}$$

   which is the plane in $\mathbb{R}^3$ represented by the linear equation $z = c$. Note that the plane $z = c$ is parallel to the xy-plane.



   In general, let W be a plane in $\mathbb{R}^3$ that contains the origin. The cosets of W are planes in $\mathbb{R}^3$ parallel to W. (See also Discussion 3.2.15.2.)

4. (In this example, vectors in $\mathbb{F}^n$ are written as column vectors.) Let $\boldsymbol{A}$ be an $m \times n$ matrix over a field $\mathbb{F}$ and let W be the nullspace of $\boldsymbol{A}$, i.e. $W = \{\boldsymbol{u} \in \mathbb{F}^n \mid \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0}\}$. Take any $\boldsymbol{v} \in \mathbb{F}^n$. Set $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{v}$. The coset of W containing $\boldsymbol{v}$ is

   $$W + \boldsymbol{v} = \{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \in \mathbb{F}^n \text{ and } \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0}\} = \{\boldsymbol{w} \in \mathbb{F}^n \mid \boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}\}$$

   which is the solution set of the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. (See also Theorem 4.3.6.)

Theorem 8.7.3 Let W be a subspace of a vector space V.

1. For any $\boldsymbol{v}, \boldsymbol{w} \in V$, the following are equivalent:

   (a) $\boldsymbol{v} \in W + \boldsymbol{w}$;

   (b) $\boldsymbol{w} \in W + \boldsymbol{v}$;

   (c) $\boldsymbol{v} - \boldsymbol{w} \in W$;

   (d) $W + \boldsymbol{v} = W + \boldsymbol{w}$.

2. For any $\boldsymbol{v}, \boldsymbol{w} \in V$, either $W + \boldsymbol{v} = W + \boldsymbol{w}$ or $(W + \boldsymbol{v}) \cap (W + \boldsymbol{w}) = \emptyset$.

Proof

1. ((a)⇔(b)) Since W is a subspace, $\boldsymbol{u} \in W$ if and only if $-\boldsymbol{u} \in W$. Thus

   $$\begin{aligned} \boldsymbol{v} \in W + \boldsymbol{w} \quad &\Leftrightarrow \quad \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} \text{ for some } \boldsymbol{u} \in W \\ &\Leftrightarrow \quad \boldsymbol{w} = (-\boldsymbol{u}) + \boldsymbol{v} \text{ for some } \boldsymbol{u} \in W \\ &\Leftrightarrow \quad \boldsymbol{w} \in W + \boldsymbol{v}. \end{aligned}$$

   ((a)⇔(c)) $$\begin{aligned} \boldsymbol{v} \in W + \boldsymbol{w} \quad &\Leftrightarrow \quad \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} \text{ for some } \boldsymbol{u} \in W \\ &\Leftrightarrow \quad \boldsymbol{v} - \boldsymbol{w} = \boldsymbol{u} \text{ for some } \boldsymbol{u} \in W \end{aligned}$$

$$\Leftrightarrow \quad \boldsymbol{v} - \boldsymbol{w} \in W.$$

((a)$\Leftrightarrow$(d)) ($\Rightarrow$) Suppose $\boldsymbol{v} \in W + \boldsymbol{w}$. Then $\boldsymbol{v} - \boldsymbol{w} \in W$. Hence

$$\begin{aligned} \boldsymbol{u} \in W + \boldsymbol{v} \ \Leftrightarrow \ \boldsymbol{u} - \boldsymbol{v} \in W \ &\Leftrightarrow \ (\boldsymbol{u} - \boldsymbol{v}) + (\boldsymbol{v} - \boldsymbol{w}) \in W \\ &\Leftrightarrow \ \boldsymbol{u} - \boldsymbol{w} \in W \ \Leftrightarrow \ \boldsymbol{u} \in W + \boldsymbol{w}. \end{aligned}$$

So $W + \boldsymbol{v} = W + \boldsymbol{w}$.

($\Leftarrow$) Since $\boldsymbol{v} \in W + \boldsymbol{v}$, if $W + \boldsymbol{v} = W + \boldsymbol{w}$, then $\boldsymbol{v} \in W + \boldsymbol{w}$.

2. Given $\boldsymbol{u}, \boldsymbol{v} \in V$, assume $(W + \boldsymbol{v}) \cap (W + \boldsymbol{w}) \neq \emptyset$. We need to show that $W + \boldsymbol{v} = W + \boldsymbol{w}$.

   Take any $\boldsymbol{u} \in (W + \boldsymbol{v}) \cap (W + \boldsymbol{w})$, i.e. $\boldsymbol{u} \in W + \boldsymbol{v}$ and $\boldsymbol{u} \in W + \boldsymbol{w}$. By Part 1, we have $W + \boldsymbol{v} = W + \boldsymbol{u} = W + \boldsymbol{w}$.

Example 8.7.4

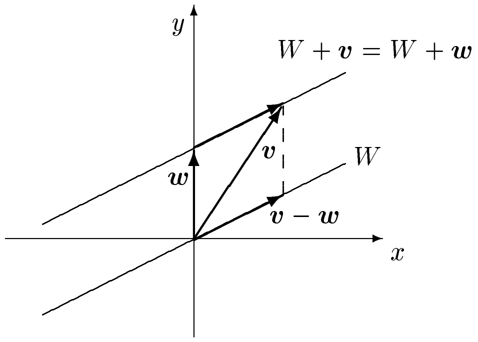
1. Following Example 8.7.2.2, let

   $$W = \{(x, y) \in \mathbb{R}^2 \mid x - 2y = 0\}.$$

   Let $\boldsymbol{v} = (2, 3)$ and $\boldsymbol{w} = (0, 2)$. Since

   $$\boldsymbol{v} - \boldsymbol{w} = (2, 1) \in W,$$

   by Theorem 8.7.3.1, $W + \boldsymbol{v} = W + \boldsymbol{w}$.



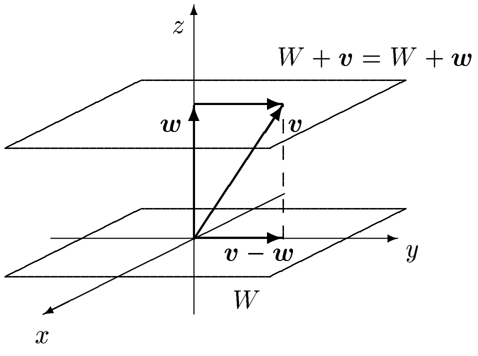
2. Following Example 8.7.2.3, let

   $$W = \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}.$$

   Let $\boldsymbol{w} = (0, 1, 2)$ and $\boldsymbol{v} = (0, 0, 2)$. Since

   $$\boldsymbol{v} - \boldsymbol{w} = (0, 1, 0) \in W,$$

   by Theorem 8.7.3.1, $W + \boldsymbol{v} = W + \boldsymbol{w}$.



Lemma 8.7.5 Let V be a vector space over a field $\mathbb{F}$ and let W be a subspace of V.

1. Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{v_1}, \boldsymbol{v_2} \in V$ such that $W + \boldsymbol{u_1} = W + \boldsymbol{u_2}$ and $W + \boldsymbol{v_1} = W + \boldsymbol{v_2}$. Then $W + (\boldsymbol{u_1} + \boldsymbol{v_1}) = W + (\boldsymbol{u_2} + \boldsymbol{v_2})$.

2. Suppose $\boldsymbol{u_1}, \boldsymbol{u_2} \in V$ such that $W + \boldsymbol{u_1} = W + \boldsymbol{u_2}$. Then $W + c\boldsymbol{u_1} = W + c\boldsymbol{u_2}$ for all $c \in \mathbb{F}$.

Proof The proof is left as exercise. See Question 8.39.

Definition 8.7.6 Let V be a vector space over a field $\mathbb{F}$ and let W be a subspace of V. We define the *addition* of two cosets by

$$(W+\boldsymbol{u})+(W+\boldsymbol{v}) = W+(\boldsymbol{u}+\boldsymbol{v}) \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in V. \tag{8.3}$$

Let A and B be two cosets of W. Since A and B can be represented as $W+\boldsymbol{u}$ and $W+\boldsymbol{v}$, respectively, by many different choices of $\boldsymbol{u}$ and $\boldsymbol{v}$, we need to make sure that our definition of $A+B$ will always give us the same answer in despite of the choices of $\boldsymbol{u}$ and $\boldsymbol{v}$. Luckily, by Lemma 8.7.5.1, the addition defined in (8.3) is well-defined.

Similarly, we define the *scalar multiplication* of a coset by

$$c(W+\boldsymbol{u}) = W+c\boldsymbol{u} \quad \text{for } c \in \mathbb{F} \text{ and } \boldsymbol{u} \in V. \tag{8.4}$$

By Lemma 8.7.5.2, the scalar multiplication defined in (8.4) is well-defined.

Example 8.7.7 Following Example 8.7.2.2, let $W = \{(x,y) \in \mathbb{R}^2 \mid x-2y=0\}$. Let $\boldsymbol{u} = (1,1)$ and $\boldsymbol{v} = (-2,1)$. Note that

$$W+\boldsymbol{u} = W+(1,1) = \{(x,y) \mid x-2y=-1\} \text{ and } W+\boldsymbol{v} = W+(-2,1) = \{(x,y) \mid x-2y=-4\}.$$

Then

$$(W+\boldsymbol{u})+(W+\boldsymbol{v}) = W+(\boldsymbol{u}+\boldsymbol{v}) = W+(-1,2) = \{(x,y) \mid x-2y=-5\}$$

and

$$3(W+\boldsymbol{u}) = W+3\boldsymbol{u} = W+(3,3) = \{(x,y) \mid x-2y=-3\}.$$

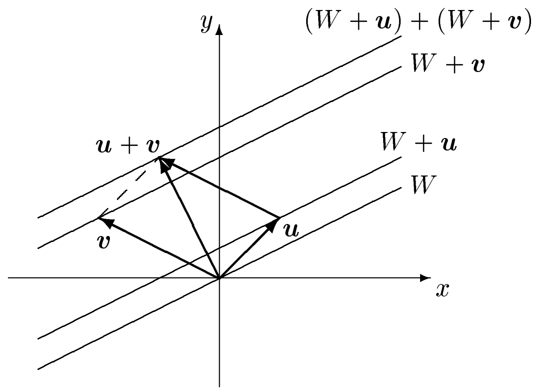


Figure A

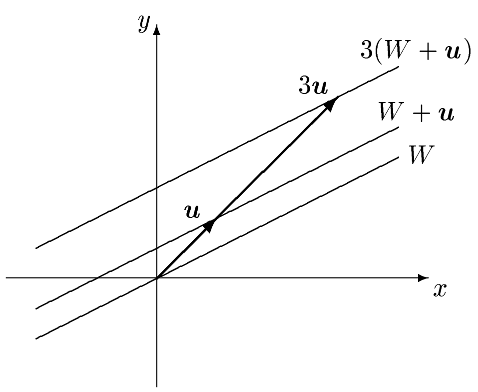


Figure B

Let $\boldsymbol{u}' = (0, \frac{1}{2})$ and $\boldsymbol{v}' = (0,2)$. Note that

$$W+\boldsymbol{u}' = W+(0,\tfrac{1}{2}) = \{(x,y) \in \mathbb{R}^2 \mid x-2y=-1\} = W+\boldsymbol{u}$$

and

$$W + \boldsymbol{v}' = W + (0, 2) = \{(x, y) \in \mathbb{R}^2 \mid x - 2y = -4\} = W + \boldsymbol{v}.$$

Also

$$(W+\boldsymbol{u}')+(W+\boldsymbol{v}') = W+(\boldsymbol{u}'+\boldsymbol{v}') = W+(0, \tfrac{5}{2}) = \{(x, y) \mid x-2y = -5\} = (W+\boldsymbol{u})+(W+\boldsymbol{v})$$

and

$$3(W + \boldsymbol{u}') = W + 3\boldsymbol{u}' = W + (0, \tfrac{3}{2}) = \{(x, y) \mid x - 2y = -3\} = 3(W + \boldsymbol{u}).$$

(As an exercise, draw the vectors $\boldsymbol{u}'$, $\boldsymbol{v}'$, $\boldsymbol{u}' + \boldsymbol{v}'$ on Figure A above and also draw the vectors $\boldsymbol{u}'$, $3\boldsymbol{u}'$ on Figure B.)

Theorem 8.7.8 Let V be a vector space over a field $\mathbb{F}$ and W a subspace of V. Denote the set of all cosets of W in V by V/W, i.e.

$$V/W = \{W + \boldsymbol{u} \mid \boldsymbol{u} \in V\}.$$

Then V/W is a vector space over $\mathbb{F}$ using the addition and scalar multiplication defined in (8.3) and (8.4).

Proof (V1) and (V6) follow from the definitions of the addition and scalar multiplication. (V2)-(V3) and (V7)-(V10) follow directly from the properties of V.

For example, for all $W + \boldsymbol{u}$, $W + \boldsymbol{v} \in V/W$,

$$\begin{aligned}(W + \boldsymbol{u}) + (W + \boldsymbol{v}) &= W + (\boldsymbol{u} + \boldsymbol{v}) \\ &= W + (\boldsymbol{v} + \boldsymbol{u}) \qquad \text{(by (V2) of } V\text{)} \\ &= (W + \boldsymbol{v}) + (W + \boldsymbol{u})\end{aligned}$$

and hence (V2) is satisfied.

Finally, for (V4), the zero vector is W $(= W + \mathbf{0})$; and for (V5), the negative of $W + \boldsymbol{u} \in V/W$ is $W + (-\boldsymbol{u})$ (which we usually written as $W - \boldsymbol{u}$).

Definition 8.7.9 The vector space V/W in Theorem 8.7.8 is called the *quotient space of V modulo W*.

Remark 8.7.10 In abstract algebra, "quotient" is used to define modulo arithmetics for algebraic structures.

For example, let $n\mathbb{Z} = \{0, \pm n, \pm 2n, \dots\} \subseteq \mathbb{Z}$. Define

$$\mathbb{Z}/n\mathbb{Z} = \{n\mathbb{Z} + a \mid a \in \mathbb{Z}\} \quad \text{where for } a \in \mathbb{Z},\ n\mathbb{Z} + a = \{a, \pm n + a, \pm 2n + a, \dots\}.$$

For $a, b \in \mathbb{Z}$, $n\mathbb{Z} + a = n\mathbb{Z} + b$ if and only if $a \equiv b \pmod{n}$. The operations of addition and multiplication defined by

$$(n\mathbb{Z}+a)+(n\mathbb{Z}+b) = n\mathbb{Z}+(a+b) \quad \text{and} \quad (n\mathbb{Z}+a)(n\mathbb{Z}+b) = n\mathbb{Z}+ab \quad \text{for } n\mathbb{Z}+a,\ n\mathbb{Z}+b \in \mathbb{Z}/n\mathbb{Z}$$

resemble the arithmetic of integer addition and multiplication modulo n.

Theorem 8.7.11 Let V be a finite dimensional vector space and W a subspace of V. Let $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_m}\}$ be a basis for W.

1. For $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k} \in V$, $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}, \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_m}\}$ is a basis for V if and only if $\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}\}$ is a basis for V/W.

2. $\dim(V/W) = \dim(V) - \dim(W)$.

Proof In the following, we only prove that if $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}, \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_m}\}$ is a basis for V, then $\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}\}$ is a basis for V/W:

(i) Take any $W + \boldsymbol{u} \in V/W$. As $\boldsymbol{u} \in V$,

$$\boldsymbol{u} = a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_k\boldsymbol{v_k} + b_1\boldsymbol{w_1} + b_2\boldsymbol{w_2} + \cdots + b_m\boldsymbol{w_m}$$

for some scalars $a_1, a_2 \ldots, a_k, b_1, b_2, \ldots, b_m$. Since $b_1\boldsymbol{w_1} + b_2\boldsymbol{w_2} + \cdots + b_m\boldsymbol{w_m} \in W$, $\boldsymbol{u} - (a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_k\boldsymbol{v_k}) \in W$. By Theorem 8.7.3.1,

$$\begin{aligned} W + \boldsymbol{u} &= W + (a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + \boldsymbol{v_k}) \\ &= a_1(W + \boldsymbol{v_1}) + a_2(W + \boldsymbol{v_2}) + \cdots + a_k(W + \boldsymbol{v_k}) \\ &\in \operatorname{span}\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}\}. \end{aligned}$$

Hence $V/W = \operatorname{span}\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}\}$.

(ii) Consider the equation

$$c_1(W + \boldsymbol{v_1}) + c_2(W + \boldsymbol{v_2}) + \cdots + c_k(W + \boldsymbol{v_k}) = W. \tag{8.5}$$

As $c_1(W + \boldsymbol{v_1}) + c_2(W + \boldsymbol{v_2}) + \cdots + c_k(W + \boldsymbol{v_k}) = W + (c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k})$, the equation (8.5) implies $c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} \in W$. So

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} = d_1\boldsymbol{w_1} + d_2\boldsymbol{w_2} + \cdots + d_m\boldsymbol{w_m},$$

for some scalar $d_1, d_2, \ldots, d_m$, and hence

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} - d_1\boldsymbol{w_1} - d_2\boldsymbol{w_2} - \cdots - d_m\boldsymbol{w_m} = \boldsymbol{0} \tag{8.6}$$

Since $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}, \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_m}$ are linearly independent, all the coefficients of (8.6) must be zero. In particular, we have $c_1 = 0$, $c_2 = 0$, $\ldots$, $c_k = 0$, i.e. (8.5) has only the trivial solution.

So $W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}$ are linearly independent.

By (i) and (ii), $\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_k}\}$ is a basis for V/W.

(Proofs of the other parts are left as exercises. See Question 8.44.)

Example 8.7.12

1. Following Example 8.7.2.2, let $W = \{(x, y) \in \mathbb{R}^2 \mid x - 2y = 0\}$. It is easy to check that $W = \operatorname{span}\{(2, 1)\}$ and hence $\{(2, 1)\}$ is a basis for W and $\dim(W) = 1$. We extend

$\{(2,1)\}$ to a basis $\{(2,1),\ (0,1)\}$ for $\mathbb{R}^2$. By Theorem 8.7.11.1, $\{W+(0,1)\}$ is a basis for the quotient space $\mathbb{R}^2/W$.

Note that $\dim(\mathbb{R}^2/W) = 1 = 2-1 = \dim(\mathbb{R}^2) - \dim(W)$.

2. Let $W = \mathrm{span}\{(2,2,-1,0,1),\ (-2,-2,4,-6,2),\ (0,0,1,1,1),\ (1,1,-2,0,-1)\}$ be a subspace of $\mathbb{R}^5$.

$$\begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -2 & -2 & 4 & -6 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix} \xrightarrow[\text{Elimination}]{\text{Gaussian}} \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

So $C = \{(2,2,-1,0,1),\ (0,0,3,-6,3),\ (0,0,0,3,0)\}$ is a basis for W. Following the algorithm in Example 4.1.14.2 (see also Example 8.5.18), we extend C to a basis for $\mathbb{R}^5$:

$$\{(2,2,-1,0,1),\ (0,0,3,-6,3,\ (0,0,0,3,0),\ (0,1,0,0,0),\ (0,0,0,0,1)\}.$$

By Theorem 8.7.11.1, $\{W+(0,1,0,0,0),\ W+(0,0,0,0,1)\}$ is a basis for the quotient space $\mathbb{R}^5/W$.

Note that $\dim(\mathbb{R}^5/W) = 2 = 5-3 = \dim(\mathbb{R}^5) - \dim(W)$.

Exercise 8

Question 8.1 to Question 8.6 are exercises for Section 8.1.

1. (a) Let $A = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \,\middle|\, a, b \in \mathbb{R} \right\}$.

   Prove that A is a field using the usual matrix addition and multiplication.
   (Hint: The axioms (F2), (F3), (F8) and (F11) are also properties of matrices. You only need to check the remaining axioms.)

   (b) Let $B = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} \,\middle|\, a, b \in \mathbb{R} \right\}$.

   Is B a field using the usual matrix addition and multiplication?

2. Let $K = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$. Show that K is a field using the usual addition and multiplication of real numbers.
   (Hint: (F2), (F3), (F7), (F8) and (F11) are properties of real numbers. You only need to check the remaining axioms.)

3. The finite field $\mathbb{F}_4$ is defined as follows: Let $\mathbb{F}_4 = \{0,\ 1,\ x,\ 1+x\}$. The addition and multiplication are the same as the polynomial addition and multiplication except $1+1=0$, $x+x=0$ and $x^2 = 1+x$.

(a) Write down the addition and multiplication tables for $\mathbb{F}_4$ (as in Example 8.1.3.2).

(b) For every element a of $\mathbb{F}_4$, find $-a$.

(c) For every nonzero element b of $\mathbb{F}_4$, find b^{-1}.

(d) Find the inverse of the square matrix $\begin{pmatrix} 1 & 1 & 1 \\ 0 & x & 1 \\ 0 & 0 & 1 \end{pmatrix}$ over $\mathbb{F}_4$.

4. Solve the following linear system over $\mathbb{F}$ when (i) $\mathbb{F} = \mathbb{R}$ and (ii) $\mathbb{F} = \mathbb{F}_2$.

$$\begin{cases} x_1 + x_2 + x_3 \qquad + x_5 = 0 \\ x_1 \qquad + x_3 + x_4 \qquad = 1 \\ \qquad x_2 \qquad + x_4 + x_5 = 1 \end{cases}$$

5. Complete the proof of Proposition 8.1.5:

   Let $\mathbb{F}$ be a field.

   (a) If $b, c \in \mathbb{F}$ satisfies the property that $a + b = a + c = a$ for all $a \in \mathbb{F}$, show that $b = c$.

   (b) If b, c are nonzero elements in $\mathbb{F}$ satisfying the property $ba = ca = a$ for all $a \in \mathbb{F}$, show that $b = c$.

   (c) For any $a \in \mathbb{F}$ and $a \neq 0$, if there exist $b, c \in \mathbb{F}$ such that $ab = ac = 1$, show that $b = c$.

   (d) For any $a \in \mathbb{F}$, show that $a0 = 0$.

   (e) For any $a \in \mathbb{F}$, show that $(-1)a = -a$.

   (f) For any $a, b \in \mathbb{F}$, prove that if $ab = 0$, then $a = 0$ or $b = 0$.

6. Prove Proposition 8.1.11:

   (a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ matrix over $\mathbb{F}$, show that $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$.

   (b) If $c \in \mathbb{F}$ and $\boldsymbol{A}$ is an $n \times n$ matrix over $\mathbb{F}$, show that $\mathrm{tr}(c\boldsymbol{A}) = c\,\mathrm{tr}(\boldsymbol{A})$.

   (c) If $\boldsymbol{C}$ and $\boldsymbol{D}$ are $m \times n$ and $n \times m$ matrices, respectively, over $\mathbb{F}$, show that $\mathrm{tr}(\boldsymbol{CD}) = \mathrm{tr}(\boldsymbol{DC})$.

Question 8.7 to Question 8.12 are exercises for Section 8.2.

7. For each of the following, list all axioms in the definition of vector space, i.e. Definition 8.2.2, which are not satisfied by the given vector addition and scalar multiplication defined on the vector set V over $\mathbb{R}$.

(a) $V = \mathbb{R}^2$, $(x, y) + (x', y') = (x + x', x + x' + y + y')$ and $c(x, y) = (cx, cy)$ for $(x, y), (x', y') \in V$ and $c \in \mathbb{R}$.

(b) $V = \{(x, y) \in \mathbb{R}^2 \mid y \neq 0\}$, $(x, y) + (x', y') = (x + x', yy')$ and $c(x, y) = (cx, y)$ for $(x, y), (x', y') \in V$ and $c \in \mathbb{R}$.

8. Prove that $\mathbb{R}^2$ is a vector space over $\mathbb{R}$ using the following vector addition and scalar multiplication:

$$(x, y) + (x', y') = (x + x' + 1,\ y + y' - 2) \quad \text{and} \quad c(x, y) = (cx + c - 1,\ cy - 2c + 2)$$

for $(x, y), (x', y') \in \mathbb{R}^2$ and $c \in \mathbb{R}$.

9. Verify Example 8.2.3.9(b):

Let V be the set of all positive real numbers, i.e. $V = \{a \in \mathbb{R} \mid a > 0\}$. Define the vector addition $\dagger$ by

$$a \dagger b = ab \quad \text{for } a, b \in V$$

and define the scalar multiplication $*$ by

$$m * a = a^m \quad \text{for } m \in \mathbb{R} \text{ and } a \in V.$$

Prove that V is a vector space over $\mathbb{R}$ using these two operations.

10. Let $A = \{0, 1\}$. Let $f, g, h \in \mathcal{F}(A, \mathbb{R})$ such that $f(t) = 1 + t^2$, $g(t) = \sin(\frac{1}{2}t\pi) + \cos(\frac{1}{2}t\pi)$ and $h(t) = t$ for $t \in A$. Show that $f = g + h$.

11. Let U and V be vector spaces over a field $\mathbb{F}$. Let

$$U \times V = \{(\boldsymbol{u},\ \boldsymbol{v}) \mid \boldsymbol{u} \in U \text{ and } \boldsymbol{v} \in V\}.$$

Define the vector addition and scalar multiplication as follows:

$$(\boldsymbol{u},\ \boldsymbol{v}) + (\boldsymbol{u}',\ \boldsymbol{v}') = (\boldsymbol{u} + \boldsymbol{u}',\ \boldsymbol{v} + \boldsymbol{v}') \quad \text{and} \quad c(\boldsymbol{u},\ \boldsymbol{v}) = (c\boldsymbol{u},\ c\boldsymbol{v})$$

for $(\boldsymbol{u},\ \boldsymbol{v})$, $(\boldsymbol{u}',\ \boldsymbol{v}') \in U \times V$ and $c \in \mathbb{F}$. Prove that $U \times V$ is a vector space.

12. Complete the proof of Proposition 8.2.4:

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

(a) If $\boldsymbol{v}, \boldsymbol{w} \in V$ satisfies the property that $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} = \boldsymbol{u}$ for all $\boldsymbol{u} \in V$, show that $\boldsymbol{v} = \boldsymbol{w}$.

(b) For any $\boldsymbol{u} \in V$, if there exist $\boldsymbol{v}, \boldsymbol{w} \in V$ such that $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} = \boldsymbol{0}$, show that $\boldsymbol{v} = \boldsymbol{w}$.

(c) For all $\boldsymbol{u} \in V$, show that $(-1)\boldsymbol{u} = -\boldsymbol{u}$.

(d) For all $c \in \mathbb{F}$, show that $c\mathbf{0} = \mathbf{0}$.

(e) If $c\boldsymbol{u} = \mathbf{0}$ where $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, prove that $c = 0$ or $\boldsymbol{u} = \mathbf{0}$.

Question 8.13 to Question 8.16 are exercises for Section 8.3.

13. For each of the following subsets W of the vector space V, determine whether W is a subspace of V.

(a) $V = \mathbb{F}_2^3$ and $W = \{(0,0,0),\ (1,1,0),\ (0,0,1)\}$.

(b) $V = \mathbb{F}_2^3$ and $W = \{(0,0,0),\ (1,1,0),\ (0,0,1),\ (1,1,1)\}$.

(c) $V = \mathcal{M}_{n\times n}(\mathbb{C})$ and $W = \{\boldsymbol{A} \in V \mid \boldsymbol{AB} = \mathbf{0}\}$ where $\boldsymbol{B}$ is a given $n \times n$ real matrix.

(d) $V = \mathcal{M}_{n\times n}(\mathbb{C})$ and $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is invertible}\}$.

(e) $V = \mathcal{M}_{n\times n}(\mathbb{C})$ and $W = \{\boldsymbol{A} \in V \mid \det(\boldsymbol{A}) = 0\}$.

(f) $V = \mathbb{R}^{\mathbb{N}}$ and $W = \{(a_n)_{n\in\mathbb{N}} \in V \mid a_{n+2} = a_n + a_{n+1} \text{ for } n = 1,2,3,\dots\}$.

(g) $V = \mathbb{R}^{\mathbb{N}}$ and $W = \{(a_n)_{n\in\mathbb{N}} \in V \mid a_{n+2} = a_n a_{n+1} \text{ for } n = 1,2,3,\dots\}$.

(h) $V = \mathbb{R}^{\mathbb{N}}$ and W is the set of all convergent sequences over $\mathbb{R}$.

(i) $V = \mathbb{R}^{\mathbb{N}}$ and W is the set of all divergent sequences over $\mathbb{R}$.

(j) $V = \mathcal{P}(\mathbb{R})$ and $W = \{a + bx + cx^2 \mid a, b, c \in \mathbb{Z}\}$.

(k) $V = \mathcal{P}(\mathbb{R})$ and $W = \{(a+b) + (a-b)x \mid a, b \in \mathbb{R}\}$.

(l) $V = \mathcal{P}(\mathbb{R})$ and $W = \{p(x) \in V \mid p(1) = 0\}$.

(m) $V = \mathcal{P}(\mathbb{R})$ and $W = \{p(x) \in V \mid p(1) = 0 \text{ and } p(2) = 0\}$.

(n) $V = \mathcal{P}(\mathbb{R})$ and $W = \{p(x) \in V \mid p(1) = 0 \text{ or } p(2) = 0\}$.

(o) $V = \mathcal{P}(\mathbb{R})$ and $W = \{p(x) \in V \mid p(2) = 1\}$.

(p) $V = \mathcal{P}(\mathbb{R})$ and $W = \{p(x) \in V \mid p(1) \geq 0\}$.

(q) $V = \mathcal{P}(\mathbb{F})$ and $W = \{p(x) \in V \mid \text{the degree of } p(x) \text{ is equal to } n\}$ where $\mathbb{F}$ is a field and n is a positive integer.

(r) $V = C^2([a,b])$, where $[a,b]$, with $a < b$, is a closed interval on the real line, and
$$W = \left\{ f \in V \;\middle|\; \frac{d^2 f(x)}{dx^2} - 3\frac{df(x)}{dx} + 2f(x) = 0 \text{ for } x \in [a,b] \right\}.$$

(s) $V = C^2([a,b])$, where $[a,b]$, with $a < b$, is a closed interval on the real line, and
$$W = \left\{ f \in V \;\middle|\; \frac{d^2 f(x)}{dx^2} - 3\frac{df(x)}{dx} + 2f(x) = x \text{ for } x \in [a,b] \right\}.$$

14. Prove Remark 8.3.5:

    Let W be a nonempty subset of a vector space V over a field $\mathbb{F}$. Prove that W is a subspace of V if and only if for all $a, b \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in W$, $a\boldsymbol{u} + b\boldsymbol{v} \in W$.

15. Let W_1 and W_2 be subspaces of a vector space V.

    (a) Prove Theorem 8.3.12:

    Show that $W_1 + W_2 = \{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \in W_1 \text{ and } \boldsymbol{v} \in W_2\}$ is a subspace of V.

    (b) Prove Remark 8.3.14:

    Suppose U is a subspace of V that contains both W_1 and W_2, i.e. $W_1 \subseteq U$ and $W_2 \subseteq U$. Prove that $W_1 + W_2 \subseteq U$.

    (c) Let $W_1 = \{(a, a, 0) \mid a \in \mathbb{F}\}$ and $W_2 = \{(0, a, a) \mid a \in \mathbb{F}\}$ where $\mathbb{F}$ is a field. Write down the subspace $W_1 + W_2$ of V explicitly.

16. Let V be a vector space over a field $\mathbb{F}$.

    (a) Let W_1 and W_2 be proper subspaces of V. Prove that $W_1 \cup W_2 \neq V$.

    (b) Let W_1, W_2 and W_3 be proper subspaces of V.

    (i) Prove that if $\mathbb{F}$ has at least three elements, $W_1 \cup W_2 \cup W_3 \neq V$.

    (ii) Let $\mathbb{F} = \mathbb{F}_2$. Give an example of proper subspaces W_1, W_2, W_3 of a vector space V over $\mathbb{F}$ such that $W_1 \cup W_2 \cup W_3 = V$.

    (See Example 8.3.3.1 for the definition of proper subspaces.)

Question 8.17 to Question 8.20 are exercises for Section 8.4.

17. For each of the sets B_1 to B_6, determine whether the set (i) spans $\mathcal{P}_2(\mathbb{R})$ and (ii) are linearly independent.

    $B_1 = \{1 + x - x^2,\ -2 + 2x + x^2\}$.

    $B_2 = \{1 + x - x^2,\ -2 - 2x + 2x^2\}$.

    $B_3 = \{1 + x - x^2,\ -2 + 2x + x^2,\ 1 + 5x - 2x^2\}$.

    $B_4 = \{1 + x - x^2,\ -2 + 2x + x^2,\ 4 + 3x^2\}$.

    $B_5 = \{1 + x - x^2,\ -2 + 2x + x^2,\ 1 + 5x - 2x^2,\ 8x - 2x^2\}$.

    $B_6 = \{1 + x - x^2,\ -2 + 2x + x^2,\ 4 + 3x^2,\ 2 + 6x - 3x^2\}$.

18. Let V be a vector space over a field $\mathbb{F}$.

    (a) Let $W_1 = \operatorname{span}\{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \boldsymbol{v_3}\}$ and $W_2 = \operatorname{span}\{\boldsymbol{v_1} + \boldsymbol{v_2},\ \boldsymbol{v_2} + \boldsymbol{v_3},\ \boldsymbol{v_1} + \boldsymbol{v_3}\}$ where $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3} \in V$.

(i) Prove that $W_1 = W_2$ if $\mathbb{F} = \mathbb{R}$.

(ii) Suppose $\mathbb{F} = \mathbb{F}_2$. Give one example that $W_1 = W_2$ and one example that $W_1 \neq W_2$.

(b) Let $\boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. Take any $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n} \in V$. Define

$$W_1 = \operatorname{span}\{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \dots,\ \boldsymbol{v_n}\} \quad \text{and} \quad W_2 = \operatorname{span}\{\boldsymbol{w_1},\ \boldsymbol{w_2},\ \dots,\ \boldsymbol{w_n}\}$$

where $\boldsymbol{w_j} = a_{1j}\boldsymbol{v_1} + a_{2j}\boldsymbol{v_2} + \cdots + a_{nj}\boldsymbol{v_n}$ for $j = 1, 2, \dots, n$. Prove that if $\boldsymbol{A}$ is invertible, then $W_1 = W_2$.

19. Let $f_1, f_2, \dots, f_n \in C^{n-1}([a,b])$ where $[a,b]$,with $a < b$, is a closed interval on the real line. The *Wronskian* $W(f_1, f_2, \dots, f_n) : [a,b] \to \mathbb{R}$ is the function defined by

$$W(f_1, f_2, \dots, f_n)(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ \dfrac{df_1(x)}{dx} & \dfrac{df_2(x)}{dx} & \cdots & \dfrac{df_n(x)}{dx} \\ \vdots & \vdots & & \vdots \\ \dfrac{d^{n-1}f_1(x)}{dx^{n-1}} & \dfrac{d^{n-1}f_2(x)}{dx^{n-1}} & \cdots & \dfrac{d^{n-1}f_n(x)}{dx^{n-1}} \end{vmatrix} \quad \text{for } x \in [a,b].$$

(a) Let $f_1, f_2 \in C^1([-1,1])$ such that $f_1(x) = e^x$ and $f_2(x) = xe^x$ for $x \in [-1,1]$. Compute $W(f_1, f_2)$.

(b) Let $f_1, f_2, \dots, f_n \in C^{n-1}([a,b])$. Prove that if $W(f_1, f_2, \dots, f_n)(x_0) \neq 0$ for some $x_0 \in [a,b]$, then $f_1, f_2, \dots, f_n$ are linearly independent.

(c) Let $f_1, f_2, \dots, f_n \in C^{n-1}([a,b])$. If $W(f_1, f_2, \dots, f_n)(x) = 0$ for all $x \in [a,b]$, is it true that $f_1, f_2, \dots, f_n$ must be linearly dependent?

20. Let $V = \mathbb{F}^{\mathbb{N}}$ be the vector space of infinite sequence over the field $\mathbb{F}$. For $i = 1, 2, 3, \dots$, define $\boldsymbol{e_i} \in V$ to be the infinite sequence such that the ith term of the sequence is 1 and all other terms are 0. Let $B = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \dots\}$.

(a) Is B linearly independent?

(b) Is $V = \operatorname{span}(B)$?

Question 8.21 to Question 8.31 are exercises for Section 8.5.

21. Recall that $\mathbb{C}^2$ forms a complex vector space. If we restrict the scalars to real numbers, then it is also a real vector space. (See Example 8.4.6.6 and Example 8.5.3.4.) Let $W = \{(z, \overline{z}) \in \mathbb{C}^2 \mid z \in \mathbb{C}\}$. Here $\overline{z}$ is the complex conjugate of z (see Notation 12.1.2).

(a) Show that W is not a subspace of the complex vector space $\mathbb{C}^2$.

(b) Prove that W is a subspace of the real vector space $\mathbb{C}^2$ and find a basis for W.

22. For each of the following subset B of $V = \mathcal{P}_n(\mathbb{R})$, determine whether B is a basis for V.

(a) $B = \{1,\ 1+x,\ 1+x+x^2,\ \ldots,\ 1+x+x^2+\cdots+x^n\}$.

(b) $B = \{1+x,\ x+x^2,\ x^2+x^3,\ \ldots,\ x^{n-1}+x^n,\ x^n+1\}$.

23. For each of the following subspace W of the vector space V, (i) find a basis for W; and (ii) determine the dimension of W.

(a) $V = \mathbb{F}_2^3$ and $W = \{(0,0,0),\ (1,1,0),\ (0,1,1),\ (1,0,1)\}$.

(b) $V = \mathbb{C}^4$ and $W = \{\boldsymbol{A}\boldsymbol{u} \mid \boldsymbol{u} \in \mathbb{C}^3\}$ where $\boldsymbol{A} = \begin{pmatrix} 1 & -1 & \mathrm{i} \\ \mathrm{i} & 0 & 1 \\ \mathrm{i} & 0 & 1 \\ 1 & 0 & -\mathrm{i} \end{pmatrix}$.

(In here, vectors in $\mathbb{C}^3$ and $\mathbb{C}^4$ are written as column vectors.)

(c) $V = \mathcal{M}_{2\times 2}(\mathbb{R})$ and $W = \operatorname{span}\left\{\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ -1 & -3 \end{pmatrix}, \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}\right\}$.

(d) $V = \mathcal{M}_{n\times n}(\mathbb{R})$ and $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is diagonal}\}$.

(e) $V = \mathcal{M}_{n\times n}(\mathbb{R})$ and $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is symmetric}\}$.

(f) $V = \mathcal{M}_{n\times n}(\mathbb{R})$ and $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is skew symmetric}\}$.

(g) $V = \mathcal{M}_{n\times n}(\mathbb{R})$ and $W = \{\boldsymbol{A} \in V \mid \operatorname{tr}(\boldsymbol{A}) = \boldsymbol{0}\}$.

(h) $V = \mathcal{P}_4(\mathbb{C})$ and $W = \operatorname{span}\{1+x+x^2,\ \mathrm{i}x - x^4,\ 1+\mathrm{i}x - x^2 - \mathrm{i}x^3,\ 2+x-\mathrm{i}x^3+x^4,\ (1-\mathrm{i})x + 2x^2 + \mathrm{i}x^3\}$.

(i) $V = \mathcal{P}_n(\mathbb{C})$ and $W = \{p(x) \in V \mid p(z) = 0\}$ where $z \in \mathbb{C}$ is a constant.

(j) $V = \mathbb{R}^{\mathbb{N}}$ and $W = \{(a_n)_{n\in\mathbb{N}} \in V \mid a_{n+3} = 2a_n \text{ for } n = 1,2,3,\ldots\}$.

(k) $V = C^\infty([0,2\pi])$ and $W = \operatorname{span}\{f_1, f_2, f_3, f_4\}$ where $f_1(x) = \sin(x)$, $f_2(x) = \cos(x)$, $f_3(x) = \sin(2x)$ and $f_4(x) = \cos(2x)\}$ for $x \in [0, 2\pi]$.

24. In(a) and (b) below, you are asked to prove Lemma 8.5.7:

Let V be a finite dimensional vector space over a field $\mathbb{F}$, where $\dim(V) = n \geq 1$, and let B be an ordered basis for V.

(a) For any $\boldsymbol{u}, \boldsymbol{v} \in V$, prove that $\boldsymbol{u} = \boldsymbol{v}$ if and only if $(\boldsymbol{u})_B = (\boldsymbol{v})_B$.

(b) For any $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r} \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{F}$, show that

$$(c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_B = c_1(\boldsymbol{v_1})_B + c_2(\boldsymbol{v_2})_B + \cdots + c_r(\boldsymbol{v_r})_B.$$

(c) For $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r} \in V$, prove that $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r}$ are linearly dependent (respectively, independent) vectors in V if and only if $(\boldsymbol{v_1})_B,\ (\boldsymbol{v_2})_B,\ \dots,\ (\boldsymbol{v_r})_B$ are linearly dependent (respectively, independent) (respectively, independent) vectors in $\mathbb{F}^n$.

(d) Let W be a subspace of V and define $W_B = \{(\boldsymbol{u})_B \mid \boldsymbol{u} \in W\} \subseteq \mathbb{F}^n$.

(i) Is W_B a subspace of $\mathbb{F}^n$?

(ii) For $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r} \in V$, prove that $\operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r}\} = W$ if and only if $\operatorname{span}\{(\boldsymbol{v_1})_B,\ (\boldsymbol{v_2})_B,\ \dots,\ (\boldsymbol{v_r})_B\} = W_B$.

(iii) Prove that $\dim(W) = \dim(W_B)$.

25. Let W be a subspace of $V = \mathcal{M}_{2\times 3}(\mathbb{R})$ with an ordered basis

$$B = \left\{ \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 3 & 1 & 1 \\ -1 & -3 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -2 & 0 \\ 2 & 0 & 1 \end{pmatrix} \right\}.$$

For each of the following matrices $\boldsymbol{A}$, (i) determine whether $\boldsymbol{A} \in W$; (ii) if so, compute the coordinate vector of $\boldsymbol{A}$ relative to B.

(a) $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. (b) $\boldsymbol{A} = \begin{pmatrix} 0 & 4 & -1 \\ -4 & 0 & -2 \end{pmatrix}$. (c) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

(d) $\boldsymbol{A} = \begin{pmatrix} -3 & -6 & -1 \\ 6 & 3 & 3 \end{pmatrix}$. (e) $\boldsymbol{A} = \begin{pmatrix} \mathrm{i} & 0 & 0 \\ 0 & -\mathrm{i} & 0 \end{pmatrix}$.

26. Let W be a subspace of $V = \mathcal{P}_4(\mathbb{R})$ with an ordered basis

$$B = \{1 - x,\ \ 2 + x - x^4,\ \ 1 + x + x^2 + x^3 + x^4\}.$$

For each of the following polynomials $p(x)$, (i) determine whether $p(x) \in W$; (ii) if so, compute the coordinate vector of $p(x)$ relative to B.

(a) $p(x) = 5x + x^2 + x^3$. (b) $p(x) = 1 + x - x^2 - x^3 + x^4$.

(c) $p(x) = -2 + 2x^2 + 2x^3 + 4x^4$.

27. For each of the set B in Question 8.25 and Question 8.26, extend it to a basis for V.

28. Let V be a vector space over a field $\mathbb{F}$ and let B be a basis for V. (In here, V can be infinite dimensional.) Prove that every nonzero vector $\boldsymbol{u} \in V$ can be expressed uniquely as a linear combination

$$\boldsymbol{u} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_m\boldsymbol{v_m}$$

for some $m \in \mathbb{N}$, $c_1, c_2, \dots, c_m \in \mathbb{F}$ and $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m} \in B$ such that $c_i \neq 0$ for all i and $\boldsymbol{v_i} \neq \boldsymbol{v_j}$ whenever $i \neq j$.

29. Let $\mathbb{F}$ be a field, $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $W = \{\boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{F}) \mid \boldsymbol{AB} = \boldsymbol{BA}\}$. Suppose there exists a column vector $\boldsymbol{v} \in \mathbb{F}^n$ such that $\{\boldsymbol{v},\ \boldsymbol{Av},\ \boldsymbol{A}^2\boldsymbol{v},\ \ldots,\ \boldsymbol{A}^{n-1}\boldsymbol{v}\}$ is a basis for $\mathbb{F}^n$.

    (a) Prove that W is a subspace of $\mathcal{M}_{n\times n}(\mathbb{F})$.

    (b) Prove that $\boldsymbol{I},\ \boldsymbol{A},\ \boldsymbol{A}^2,\ \ldots,\ \boldsymbol{A}^{n-1}$ are linearly independent vectors contained in W.

    (c) Prove that $\{\boldsymbol{I},\ \boldsymbol{A},\ \boldsymbol{A}^2,\ \ldots,\ \boldsymbol{A}^{n-1}\}$ is a basis for W.

30. Let V be a finite dimensional vector space over $\mathbb{C}$ such that $\dim_{\mathbb{C}}(V) = n$. By restricting the scalars to real numbers in the scalar multiplication, V can be regarded as a vector space over $\mathbb{R}$. (See Example 8.5.3.4.) Prove that $\dim_{\mathbb{R}}(V) = 2n$.

31. Let W be a vector space over $\mathbb{R}$. Define $W' = \{(\boldsymbol{u}, \boldsymbol{v}) \mid \boldsymbol{u}, \boldsymbol{v} \in W\}$ with the addition and scalar multiplication

$$(\boldsymbol{u}, \boldsymbol{v}) + (\boldsymbol{u}', \boldsymbol{v}') = (\boldsymbol{u} + \boldsymbol{u}',\ \boldsymbol{v} + \boldsymbol{v}') \quad \text{and} \quad c(\boldsymbol{u}, \boldsymbol{v}) = (a\boldsymbol{u} - b\boldsymbol{v},\ b\boldsymbol{u} + a\boldsymbol{v})$$

    where $(\boldsymbol{u}, \boldsymbol{v}),\ (\boldsymbol{u}', \boldsymbol{v}') \in W'$ and $c = a + b\mathrm{i} \in \mathbb{C}$ with $a, b \in \mathbb{R}$.

    (a) Prove that W' is a vector space over $\mathbb{C}$.

    (b) Suppose W is finite dimensional such that $\dim_{\mathbb{R}}(W) = n$. Find $\dim_{\mathbb{C}}(W')$.

Question 8.32 to Question 8.38 are exercises for Section 8.6.

32. For each of the following subspaces W_1 and W_2 of the vector space V,

    (i) find the dimensions of W_1, W_2, $W_1 \cap W_2$ and $W_1 + W_2$;

    (ii) determine if $W_1 + W_2$ is a direct sum; and

    (iii) determine if $V = W_1 + W_2$.

    (a) $V = \mathbb{F}^4$, $W_1 = \{(a, a, a, a) \mid a \in \mathbb{F}\}$ and $W_2 = \{(a, b, c, d) \in \mathbb{F}^4 \mid a + d = 0\}$ where $\mathbb{F}$ is a field.

    (b) $V = \mathcal{P}_3(\mathbb{R})$, $W_1 = \{a + bx + bx^2 + ax^3 \mid a, b \in \mathbb{R}\}$ and $W_2 = \{p(x) \in V \mid p(1) = 0\}$.

    (c) $V = \mathcal{P}_3(\mathbb{R})$, $W_1 = \operatorname{span}\{1+x^2,\ 1+2x^2,\ 1+3x^2\}$ and $W_2 = \operatorname{span}\{1+x,\ 1+2x,\ 1+3x\}$.

    (d) $V = \mathcal{M}_{2\times 2}(\mathbb{C})$, $W_1 = \left\{ \boldsymbol{X} \in V \;\middle|\; \begin{pmatrix} 0 & 1 \\ \mathrm{i} & 0 \end{pmatrix} \boldsymbol{X} = \boldsymbol{X} \begin{pmatrix} 0 & 1 \\ \mathrm{i} & 0 \end{pmatrix} \right\}$ and $W_2 = \operatorname{span}\left\{ \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix} \right\}$.

    (e) V has a basis $\{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \boldsymbol{v_3},\ \boldsymbol{v_4}\}$, $W_1 = \operatorname{span}\{\boldsymbol{v_1} + \boldsymbol{v_2},\ \boldsymbol{v_2} + \boldsymbol{v_3},\ \boldsymbol{v_3} + \boldsymbol{v_4},\ \boldsymbol{v_1} + \boldsymbol{v_4}\}$ and $W_2 = \operatorname{span}\{\boldsymbol{v_2}\}$.

33. For each of the following subspaces W_1 and W_2 of $V = \mathcal{M}_{n\times n}(\mathbb{R})$, determine whether (i) $V = W_1 + W_2$; and (ii) $V = W_1 \oplus W_2$.

    (a) $W_1 = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is upper triangular}\}$ and $W_2 = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is lower triangular}\}$.

    (b) $W_1 = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is upper triangular}\}$ and $W_2 = \{\boldsymbol{A} \in V \mid \boldsymbol{A} \text{ is skew symmetric}\}$.

34. Verify Example 8.6.6.3:

    Let W_1 be the subspace of $C([0, 2\pi])$ spanned by $g \in C([0, 2\pi])$ where $g(x) = \sin(x)$ for $x \in [0, 2\pi]$. Define

    $$W_2 = \left\{ f \in C([0, 2\pi]) \;\middle|\; \int_0^{2\pi} f(t)\sin(t)dt = 0 \right\}.$$

    (a) Show that W_2 is a subspace of $C([0, 2\pi])$.

    (b) Prove that $C([0, 2\pi]) = W_1 \oplus W_2$.

35. Let W_1 and W_2 be finite dimensional subspaces of a vector space. Prove that

    $$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

    (Hint: Start with a basis B for $W_1 \cap W_2$ and use Theorem 8.5.17 to extend B to bases B_1 and B_2 for W_1 and W_2, respectively. Then show that $B_1 \cup B_2$ is a basis for $W_1 + W_2$.)

36. Let W_1, W_2 and W_3 be subspaces of vector space.

    (a) Suppose $W_1 \oplus W_2 = W_1 \oplus W_3$. Is it true that $W_2 = W_3$?

    (b) Suppose W_1, W_2 and W_3 are finite dimensional.

        (i) Prove that if $\dim(W_1 + W_2 + W_3) = \dim(W_1) + \dim(W_2) + \dim(W_3)$, then $W_1 \cap W_2 = W_1 \cap W_3 = W_2 \cap W_3 = \{\mathbf{0}\}$.

        (ii) Give an example such that $W_1 \cap W_2 = W_1 \cap W_3 = W_2 \cap W_3 = \{\mathbf{0}\}$ but $\dim(W_1 + W_2 + W_3) \neq \dim(W_1) + \dim(W_2) + \dim(W_3)$.

37. Let U and V be vector spaces over a field $\mathbb{F}$ and let $U \times V = \{(\boldsymbol{u}, \boldsymbol{v}) \mid \boldsymbol{u} \in U \text{ and } \boldsymbol{v} \in V\}$ be the vector space over $\mathbb{F}$ as defined in Question 8.11. Define $U' = \{(\boldsymbol{u}, \mathbf{0}_V) \mid \boldsymbol{u} \in U\}$ and $V' = \{(\mathbf{0}_U, \boldsymbol{v}) \mid \boldsymbol{v} \in V\}$ where $\mathbf{0}_U$ and $\mathbf{0}_V$ are the zero vectors of U and V respectively.

    (a) Show that U' and V' are subspaces of $U \times V$.

    (b) Prove that $U \times V = U' \oplus V'$.

    (c) If U and V are finite dimensional, find $\dim(U')$, $\dim(V')$ and $\dim(U \times V)$ in terms of $\dim(U)$ and $\dim(V)$.

    (The vector space $U \times V$ is called the *external direct sum* of U and V.)

38. Let U, V and W be subspaces of a vector space.

    (a) Prove that $(U \cap V) + (U \cap W) \subseteq U \cap (V + W)$ and $(U + V) \cap (U + W) \supseteq U + (V \cap W)$.

    (b) Is it true that $(U \cap V) + (U \cap W) = U \cap (V + W)$? Is it true that $(U + V) \cap (U + W) = U + (V \cap W)$?

    (c) Prove that $(U \cap V) + (U \cap W) = U \cap (V + (U \cap W))$ and $(U + V) \cap (U + W) = U + (V \cap (U + W))$.

Question 8.39 to Question 8.46 are exercises for Section 8.7.

39. Prove Lemma 8.7.5:

    Let V be a vector space over a field $\mathbb{F}$ and let W be a subspace of V.

    (a) Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{v_1}, \boldsymbol{v_2} \in V$ such that $W + \boldsymbol{u_1} = W + \boldsymbol{u_2}$ and $W + \boldsymbol{v_1} = W + \boldsymbol{v_2}$. Prove that $W + (\boldsymbol{u_1} + \boldsymbol{v_1}) = W + (\boldsymbol{u_2} + \boldsymbol{v_2})$.

    (b) Suppose $\boldsymbol{u_1}, \boldsymbol{u_2} \in V$ such that $W + \boldsymbol{u_1} = W + \boldsymbol{u_2}$. Prove that $W + c\boldsymbol{u_1} = W + c\boldsymbol{u_2}$ for all $c \in \mathbb{F}$.

40. Let $V = \mathbb{F}_2^4$ and $W = \text{span}\{(1, 1, 0, 1),\ (1, 0, 1, 1)\}$.

    (a) What is the dimension of W? What is the dimension of V/W?

    (b) How many distinct cosets of W are there?

41. Let $[a, b]$, with $a < b$, be a closed interval on the real line and let

    $$W = \left\{ f \in C^2([a, b]) \,\middle|\, \frac{d^2 f(x)}{dx^2} - 3\frac{df(x)}{dx} + 2f(x) = 0 \text{ for } x \in [a, b] \right\}$$

    which is a subspace of $C^2([a, b])$. Show that each coset of W is a solution set of a differential equation

    $$\frac{d^2 f(x)}{dx^2} - 3\frac{df(x)}{dx} + 2f(x) = g(x) \quad \text{for } x \in [a, b]$$

    for some $g \in C([a, b])$.

42. Let $W = \{(a_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid a_{n+3} - 2a_n = 0 \text{ for } n = 1, 2, 3, \dots\}$ which is a subspace of $\mathbb{R}^{\mathbb{N}}$. Give an interpretation of the cosets of W in $\mathbb{R}^{\mathbb{N}}$ similar to that of Question 8.41.

43. For each of parts (a)-(e) of Question 8.32, write down a basis for V/W_1 and a basis for V/W_2.

44. Complete the proof of Theorem 8.7.11:

    Let V be a finite dimensional vector space and W a subspace of V. Let $\{\boldsymbol{w_1},\ \boldsymbol{w_2},\ \ldots,\ \boldsymbol{w_m}\}$ be a basis for W.

    (a) For $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k} \in V$, prove that if $\{W + \boldsymbol{v_1},\ W + \boldsymbol{v_2},\ \ldots,\ W + \boldsymbol{v_k}\}$ is a basis for V/W, then $\{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \ldots,\ \boldsymbol{v_k},\ \boldsymbol{w_1},\ \boldsymbol{w_2},\ \ldots,\ \boldsymbol{w_m}\}$ is a basis for V.

    (b) Prove that $\dim(V/W) = \dim(V) - \dim(W)$.

45. Let U and W be subspaces of a vector space V such that $V = U \oplus W$.

    (a) Suppose U is finite dimensional and $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ is a basis for U. Prove that $\{W + \boldsymbol{v_1},\ W + \boldsymbol{v_2},\ \ldots,\ W + \boldsymbol{v_k}\}$ is a basis for V/W.

    (b) If U is infinite dimensional and B is a basis for U, is $\{W + \boldsymbol{v} \mid \boldsymbol{v} \in B\}$ a basis for V/W?

46. Let W be a subspace of a vector space V and $\mathcal{U}$ a subspace of V/W. Define

$$U = \{\boldsymbol{u} \in V \mid W + \boldsymbol{u} \in \mathcal{U}\}.$$

    (a) Show that U is a subspace of V.

    (b) Suppose W and $\mathcal{U}$ are finite dimensional, say, $\dim(W) = k$ and $\dim(\mathcal{U}) = m$. Find $\dim(U)$.

Chapter 9

General Linear Transformations

Section 9.1 Linear Transformations

Discussion 9.1.1 In Chapter 7, a linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ is defined to be a function $T : \mathbb{R}^n \to \mathbb{R}^m$ such that $T(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{R}^n$ where $\boldsymbol{A}$ is an $m \times n$ real matrix and vectors in $\mathbb{R}^n$ are written as column vectors. In this chapter, we shall generalize the concept of linear transformations to abstract vector spaces. As a consequence, we can regard this abstract version of linear transformations as a generalized form of matrices. See Proposition 9.4.3.

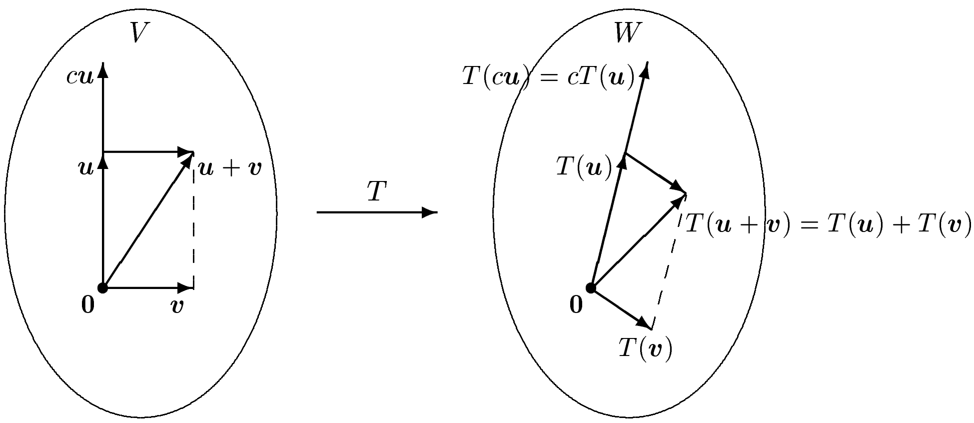
Definition 9.1.2 Let V and W be two vector spaces over a field $\mathbb{F}$. A *linear transformation* $T : V \to W$ is a mapping from V to W that satisfies the following two axioms:

(T1) For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$.

(T2) For all $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$.

If $W = V$, the linear transformation $T : V \to V$ is called a *linear operator* on V.

If $W = \mathbb{F}$, the linear transformation $T : V \to \mathbb{F}$ is called a *linear functional* on V.



Remark 9.1.3 (T1) and (T2) can be combined together: Let V and W be two vector spaces over a field $\mathbb{F}$. A mapping $T: V \to W$ is a linear transformation if and only if

$$T(a\boldsymbol{u} + b\boldsymbol{v}) = aT(\boldsymbol{u}) + bT(\boldsymbol{v}) \quad \text{for all } a, b \in \mathbb{F} \text{ and } \boldsymbol{u}, \boldsymbol{v} \in V.$$

Example 9.1.4

1. Let $\boldsymbol{A}$ be an $m \times n$ matrix over a field $\mathbb{F}$. Define a mapping $L_{\boldsymbol{A}}: \mathbb{F}^n \to \mathbb{F}^m$ by

   $$L_{\boldsymbol{A}}(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u} \quad \text{for } \boldsymbol{u} \in \mathbb{F}^n$$

   where vectors in $\mathbb{F}^n$ are written as column vectors.

   (T1) For any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^n$, $L_{\boldsymbol{A}}(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{v} = L_{\boldsymbol{A}}(\boldsymbol{u}) + L_{\boldsymbol{A}}(\boldsymbol{v})$.

   (T2) For any $c \in \mathbb{F}$ and $\boldsymbol{u} \in \mathbb{F}^n$, $L_{\boldsymbol{A}}(c\boldsymbol{u}) = \boldsymbol{A}(c\boldsymbol{u}) = c\boldsymbol{A}\boldsymbol{u} = cL_{\boldsymbol{A}}(\boldsymbol{u})$.

   So $L_{\boldsymbol{A}}$ is a linear transformation.

   (From this example, we see that Definition 9.1.2 can be regarded as a generalization of Definition 7.1.1.)

2. The *identity mapping* on a vector space V is defined to be the mapping $I_V: V \to V$ such that

   $$I_V(\boldsymbol{u}) = \boldsymbol{u} \quad \text{for } \boldsymbol{u} \in V.$$

   It is a linear operator on V and is also called the *identity operator* on V.

3. The *zero mapping* $O_{V,W}: V \to W$, where V and W are vector spaces over the same field, is defined by

   $$O_{V,W}(\boldsymbol{u}) = \boldsymbol{0} \quad \text{for } \boldsymbol{u} \in V.$$

   It is a linear transformation and is also called the *zero transformation* from V to W.

   If $W = V$, we use O_V to denote $O_{V,V}$ and call it the *zero operator* on V.

4. (a) Let $S: \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ be the mapping defined by

   $$S(p(x)) = p(x)^2 \quad \text{for } p(x) \in \mathcal{P}(\mathbb{R}).$$

   Is S a linear operator?

   (b) Let $T: \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ be the mapping defined by

   $$T(p(x)) = xp(x) \quad \text{for } p(x) \in \mathcal{P}(\mathbb{R}).$$

   Is T a linear operator?

Solution

(a) S is not a linear operator. For example

$$S(1+x) = (1+x)^2 = 1 + 2x + x^2 \neq 1 + x^2 = S(1) + S(x).$$

(b) T is a linear operator:

(T1) For any $p(x)$, $q(x) \in \mathcal{P}(\mathbb{R})$,

$$T(p(x) + q(x)) = x(p(x) + q(x)) = xp(x) + xq(x) = T(p(x)) + T(q(x)).$$

(T2) For any $c \in \mathbb{R}$ and $p(x) \in \mathcal{P}(\mathbb{R})$,

$$T(cp(x)) = x(cp(x)) = cxp(x) = cT(p(x)).$$

5. Let V be the set of all convergent sequences over $\mathbb{R}$. We know that V is a subspace of $\mathbb{R}^{\mathbb{N}}$ (see Question 8.13). Define a mapping $T : V \to \mathbb{R}$ by

$$T((a_n)_{n\in\mathbb{N}}) = \lim_{n\to\infty} a_n \quad \text{for } (a_n)_{n\in\mathbb{N}} \in V.$$

Since for any convergent sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ and any $c \in \mathbb{R}$,

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \quad \text{and} \quad \lim_{n\to\infty} ca_n = c \lim_{n\to\infty} a_n,$$

T is a linear functional.

6. Let $[a, b]$, with $a < b$, be a closed interval of the real line. We use the real vector space $C^\infty([a, b])$ defined in Example 8.3.6.5. Let $D : C^\infty([a, b]) \to C^\infty([a, b])$ be the *differential operator* such that for every $f \in C^\infty([a, b])$, $D(f)$ is a function on $C^\infty([a, b])$ defined by

$$D(f)(x) = \frac{df(x)}{dx} \quad \text{for } x \in [a, b]$$

and let $F : C^\infty([a, b]) \to C^\infty([a, b])$ be the *integral operator* such that for every $f \in C^\infty([a, b])$, $F(f)$ is a function on $C^\infty([a, b])$ defined by

$$F(f)(x) = \int_a^x f(t)dt \quad \text{for } x \in [a, b].$$

Both D and F are linear operators.

Proposition 9.1.5 Let V and W be vector space over the same field. If $T : V \to W$ is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

Proof Since $\mathbf{0} + \mathbf{0} = \mathbf{0}$,

$$\begin{aligned}
& & T(\mathbf{0}+\mathbf{0}) &= T(\mathbf{0}) \\
&\Rightarrow & T(\mathbf{0}) + T(\mathbf{0}) &= T(\mathbf{0}) \\
&\Rightarrow & T(\mathbf{0}) + T(\mathbf{0}) - T(\mathbf{0}) &= T(\mathbf{0}) - T(\mathbf{0}) \\
&\Rightarrow & T(\mathbf{0}) + \mathbf{0} &= \mathbf{0} \\
&\Rightarrow & T(\mathbf{0}) &= \mathbf{0}.
\end{aligned}$$

Remark 9.1.6 Let V and W be vector space over the same field. Suppose V has a basis B. Let $T : V \to W$ be a linear transformation. For every $\boldsymbol{u} \in V$,

$$\boldsymbol{u} = a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_m\boldsymbol{v_m}$$

for some scalar $a_1, a_2, \ldots, a_m$ and some $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m} \in B$. By using (T1) and (T2) repeatedly, we get

$$T(\boldsymbol{u}) = a_1T(\boldsymbol{v_1}) + a_2T(\boldsymbol{v_2}) + \cdots + a_mT(\boldsymbol{v_m}).$$

It follows that T is completely determined by the images of vectors from B.

On the other hand, to define a linear transformation S from V to W, we first set the image $S(\boldsymbol{v})$ for each $\boldsymbol{v} \in B$. For any $\boldsymbol{u} \in V$, since $\boldsymbol{u} = a_1\boldsymbol{v_1}+a_2\boldsymbol{v_2}+\cdots+a_m\boldsymbol{v_m}$ for some $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m} \in B$ and scalars $a_1, a_2, \ldots, a_m$, define $S(\boldsymbol{u}) = a_1S(\boldsymbol{v_1}) + a_2S(\boldsymbol{v_2}) + \cdots + a_mS(\boldsymbol{v_m})$. Then we have a linear transformation.

Example 9.1.7 Take the standard basis $\{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{R})$. Define a linear transformation $S : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ by

$$S(1) = (1, 2, 1), \quad S(x) = (0, 1, 1) \quad \text{and} \quad S(x^2) = (-1, 1, 0).$$

Then for any $p(x) = a + bx + cx^2 \in \mathcal{P}_2(\mathbb{R})$,

$$\begin{aligned} S(p(x)) &= aS(1) + bS(x) + cS(x^2) \\ &= a(1, 2, 1) + b(0, 1, 1) + c(-1, 1, 0) = (a - c,\ 2a + b + c,\ a + b). \end{aligned}$$

Section 9.2 Matrices for Linear Transformations

Theorem 9.2.1 Let $T : V \to W$ be a linear transformation where V and W are finite dimensional vector spaces over a field $\mathbb{F}$ such that $n = \dim(V) \geq 1$ and $m = \dim(W) \geq 1$. For any ordered bases B and C for V and W respectively, there exists an $m \times n$ matrix $\boldsymbol{A}$ such that

$$[T(\boldsymbol{u})]_C = \boldsymbol{A}\,[\boldsymbol{u}]_B \quad \text{for all } \boldsymbol{u} \in V.$$

(Together with Example 9.1.4.1, the linear transformation defined in Definition 9.1.2 is the same as that in Definition 7.1.1 when $V = \mathbb{F}^n$, $W = \mathbb{F}^m$ and the standard bases B and C, respectively, are used.)

Proof Let $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$. Then every $\boldsymbol{u} \in V$ can be expressed uniquely as

$$\boldsymbol{u} = a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_n\boldsymbol{v_n}$$

for some scalar $a_1, a_2, \ldots, a_n$. Using the notation of coordinate vectors (see Definition 8.5.6), we have

$$[\boldsymbol{u}]_B = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Then by Remark 9.1.6 and Lemma 8.5.7,

$$\begin{aligned} [T(\boldsymbol{u})]_C &= a_1[T(\boldsymbol{v_1})]_C + a_2[T(\boldsymbol{v_2})]_C + \cdots + a_n[T(\boldsymbol{v_n})]_C \\ &= \begin{pmatrix} [T(\boldsymbol{v_1})]_C & [T(\boldsymbol{v_2})]_C & \cdots & [T(\boldsymbol{v_n})]_C \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \\ &= \begin{pmatrix} [T(\boldsymbol{v_1})]_C & [T(\boldsymbol{v_2})]_C & \cdots & [T(\boldsymbol{v_n})]_C \end{pmatrix} [\boldsymbol{u}]_B. \end{aligned}$$

Let $\boldsymbol{A} = \begin{pmatrix} [T(\boldsymbol{v_1})]_C & [T(\boldsymbol{v_2})]_C & \cdots & [T(\boldsymbol{v_n})]_C \end{pmatrix}$. Then $\boldsymbol{A}$ is an $m \times n$ matrix such that $[T(\boldsymbol{u})]_C = \boldsymbol{A}\,[\boldsymbol{u}]_B$ for all $\boldsymbol{u} \in V$.

Definition 9.2.2 The matrix $\boldsymbol{A} = \begin{pmatrix} [T(\boldsymbol{v_1})]_C & [T(\boldsymbol{v_2})]_C & \cdots & [T(\boldsymbol{v_n})]_C \end{pmatrix}$ in the proof of Theorem 9.2.1 is called the *matrix for T relative to the ordered bases B and C*. This matrix $\boldsymbol{A}$ is usually denoted by $[T]_{C,B}$.

If $W = V$ and $C = B$, we simply denote $[T]_{B,B}$ by $[T]_B$ and the matrix is called the *matrix for T relative to the ordered basis B*.

Lemma 9.2.3 Let $T_1, T_2 : V \to W$ be linear transformations where V and W are finite dimensional vector spaces where $\dim(V) \geq 1$ and $\dim(W) \geq 1$. Take any ordered bases B and C for V and W respectively. Then $T_1 = T_2$ if and only if $[T_1]_{C,B} = [T_2]_{C,B}$.

Proof By Remark 9.1.6, every linear transformation $T : V \to W$ is completely determined by the images of vectors from B. Thus the matrix in Definition 9.2.2 uniquely define a linear transformation.

Example 9.2.4

1. In Example 9.1.7, let $B = \{1, x, x^2\}$ and $C = \{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)\}$. The matrix for S relative to B and C is

   $$[S]_{C,B} = \begin{pmatrix} [S(1)]_C & [S(x)]_C & [S(x^2)]_C \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

   We can also recover the formula of S from the matrix $[S]_{C,B}$:

For $p(x) = a + bx + cx^2 \in \mathcal{P}_2(\mathbb{R})$, $[p(x)]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$. Then

$$[S(p(x)]_C = [S]_{C,B}\,[p(x)]_B = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a-c \\ 2a+b+c \\ a+b \end{pmatrix}.$$

So

$$\begin{aligned} S(p(x)) &= (a-c)(1,0,0) + (2a+b+c)(0,1,0) + (a+b)(0,0,1) \\ &= (a-c,\, 2a+b+c,\, a+b). \end{aligned}$$

2. Let V be a vector space of dimension 3 over $\mathbb{C}$. Define a linear transformation $T : V \to \mathbb{C}^3$ such that $T(\boldsymbol{v_1}) = (1,-1,-1)$, $T(\boldsymbol{v_2}) = (\mathrm{i},2,\mathrm{i})$ and $T(\boldsymbol{v_3}) = (0,\mathrm{i},0)$ where $B = \{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \boldsymbol{v_3}\}$ is an ordered basis for V. Let $C = \{(1,1,1),\, (1,0,-1),\, (1,1,0)\}$. Find the matrix for T relative to B and C.

Solution Since $[T]_{C,B} = \Big([T(\boldsymbol{v_1})]_C \quad [T(\boldsymbol{v_2})]_C \quad [T(\boldsymbol{v_3})]_C \Big)$, we need to find the coordinate vectors $[(1,-1,-1)]_C$, $[(\mathrm{i},2,\mathrm{i})]_C$ and $[(0,\mathrm{i},0)]_C$, i.e. to find a_j, b_j, c_j, $j = 1,2,3$ such that

$$a_1(1,1,1) + b_1(1,0,-1) + c_1(1,1,0) = (1,-1,-1) \quad \Leftrightarrow \quad \begin{cases} a_1 + b_1 + c_1 = 1 \\ a_1 \qquad\; + c_1 = -1 \\ a_1 - b_1 \qquad\;\; = -1, \end{cases}$$

$$a_2(1,1,1) + b_2(1,0,-1) + c_2(1,1,0) = (\mathrm{i},2,\mathrm{i}) \quad \Leftrightarrow \quad \begin{cases} a_2 + b_2 + c_2 = \mathrm{i} \\ a_2 \qquad\; + c_2 = 2 \\ a_2 - b_2 \qquad\;\; = \mathrm{i}, \end{cases}$$

$$a_3(1,1,1) + b_3(1,0,-1) + c_3(1,1,0) = (0,\mathrm{i},0) \quad \Leftrightarrow \quad \begin{cases} a_3 + b_3 + c_3 = 0 \\ a_3 \qquad\; + c_3 = \mathrm{i} \\ a_3 - b_3 \qquad\;\; = 0. \end{cases}$$

We solve the three linear systems together (see Example 3.7.4.1):

$$\left(\begin{array}{ccc|c|c|c} 1 & 1 & 1 & 1 & \mathrm{i} & 0 \\ 1 & 0 & 1 & -1 & 2 & \mathrm{i} \\ 1 & -1 & 0 & -1 & \mathrm{i} & 0 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & 1 & -2+2\mathrm{i} & -\mathrm{i} \\ 0 & 1 & 0 & 2 & -2+\mathrm{i} & -\mathrm{i} \\ 0 & 0 & 1 & -2 & 4-2\mathrm{i} & 2\mathrm{i} \end{array}\right).$$

Then $[(1,-1,-1)]_C = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$, $[(\mathrm{i},2,\mathrm{i})]_C = \begin{pmatrix} -2+2\mathrm{i} \\ -2+\mathrm{i} \\ 4-2\mathrm{i} \end{pmatrix}$ and $[(0,\mathrm{i},0)]_C = \begin{pmatrix} -\mathrm{i} \\ -\mathrm{i} \\ 2\mathrm{i} \end{pmatrix}$. So

$$[T]_{C,B} = \Big([T(\boldsymbol{v_1})]_C \quad [T(\boldsymbol{v_2})]_C \quad [T(\boldsymbol{v_3})]_C \Big) = \begin{pmatrix} 1 & -2+2\mathrm{i} & -\mathrm{i} \\ 2 & -2+\mathrm{i} & -\mathrm{i} \\ -2 & 4-2\mathrm{i} & 2\mathrm{i} \end{pmatrix}.$$

3. Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_n} \end{pmatrix}$ be an $n \times n$ matrix over a field $\mathbb{F}$ where $\boldsymbol{c_i}$ is the ith column of $\boldsymbol{A}$. Let $L_{\boldsymbol{A}} : \mathbb{F}^n \to \mathbb{F}^n$ be the linear operator defined in Example 9.1.4.1, i.e. $L_{\boldsymbol{A}}(\boldsymbol{u}) = \boldsymbol{Au}$ for $\boldsymbol{u} \in \mathbb{F}^n$ where vectors in $\mathbb{F}^n$ are written as column vectors.

   Take the standard basis $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ for $\mathbb{F}^n$. For all $\boldsymbol{u} = (u_1, u_2, \dots, u_n)^{\mathrm{T}} \in \mathbb{F}^n$,

$$\boldsymbol{u} = u_1\boldsymbol{e_1} + u_2\boldsymbol{e_2} + \cdots + u_n\boldsymbol{e_n} \quad \Rightarrow \quad [\boldsymbol{u}]_E = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \boldsymbol{u}.$$

   Thus

$$\begin{aligned} [L_{\boldsymbol{A}}]_E &= [L_{\boldsymbol{A}}]_{E,E} \\ &= \begin{pmatrix} [L_{\boldsymbol{A}}(\boldsymbol{e_1})]_E & [L_{\boldsymbol{A}}(\boldsymbol{e_2})]_E & \cdots & [L_{\boldsymbol{A}}(\boldsymbol{e_n})]_E \end{pmatrix} \\ &= \begin{pmatrix} [\boldsymbol{Ae_1}]_E & [\boldsymbol{Ae_2}]_E & \cdots & [\boldsymbol{Ae_n}]_E \end{pmatrix} \\ &= \begin{pmatrix} [\boldsymbol{c_1}]_E & [\boldsymbol{c_2}]_E & \cdots & [\boldsymbol{c_n}]_E \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_n} \end{pmatrix} \\ &= \boldsymbol{A}. \end{aligned}$$

   (See also Discussion 7.1.8.)

Discussion 9.2.5 Use the identity operator $I_V : V \to V$ in Example 9.1.4.2 where V is a finite dimensional vector space with $\dim(V) \geq 1$. Suppose B and C are two ordered bases for V. Then from Theorem 9.2.1, we have

$$[\boldsymbol{u}]_C = [I_V(\boldsymbol{u})]_C = [I_V]_{C,B}\,[\boldsymbol{u}]_B \quad \text{for all } \boldsymbol{u} \in V.$$

So for any $\boldsymbol{u} \in V$, the matrix $[I_V]_{C,B}$ is doing the job of converting the coordinate vector of $\boldsymbol{u}$ relative to B to the coordinate vector of $\boldsymbol{u}$ relative to C. Thus $[I_V]_{C,B}$ is called the *transition matrix* from B to C.

Suppose $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$. Then

$$\begin{aligned} [I_V]_{C,B} &= \begin{pmatrix} [I_V(\boldsymbol{v_1})]_C & [I_V(\boldsymbol{v_2})]_C & \cdots & [I_V(\boldsymbol{v_n})]_C \end{pmatrix} \\ &= \begin{pmatrix} [\boldsymbol{v_1}]_C & [\boldsymbol{v_2}]_C & \cdots & [\boldsymbol{v_n}]_C \end{pmatrix} \end{aligned}$$

and it resembles the transition matrix defined in Definition 3.7.3.

Theorem 9.2.6 The matrix $[I_V]_{C,B}$ in Discussion 9.2.5 is invertible and its inverse is the transition matrix from C to B, i.e. $(\,[I_V]_{C,B})^{-1} = [I_V]_{B,C}$. (See also Theorem 3.7.5.)

Proof It is easier to prove the theorem using the concept of compositions of linear transformations. So we leave it as an exercise for next section. See Question 9.16.

Example 9.2.7 Let $E = \{(1,0,0),\ (0,1,0),\ (0,0,1)\}$ and $B = \{(1,\mathrm{i},1),\ (0,1,1),\ (\mathrm{i},-1,-1)\}$. They are bases for $\mathbb{C}^3$. Find the transition matrix from E to B.

Solution Since $[I_{\mathbb{C}^3}]_{E,B} = \Big([(1,\mathrm{i},1)]_E \quad [(0,1,1)]_E \quad [(\mathrm{i},-1,-1)]_E \Big) = \begin{pmatrix} 1 & 0 & \mathrm{i} \\ \mathrm{i} & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$, the transition matrix from E to B is

$$[I_{\mathbb{C}^3}]_{B,E} = \left([I_{\mathbb{C}^3}]_{E,B} \right)^{-1} = \begin{pmatrix} 1 & 0 & \mathrm{i} \\ \mathrm{i} & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & \frac{-1-\mathrm{i}}{2} & \frac{1+\mathrm{i}}{2} \\ -\mathrm{i} & 1 & 0 \\ -\mathrm{i} & \frac{1-\mathrm{i}}{2} & \frac{-1+\mathrm{i}}{2} \end{pmatrix}.$$

Section 9.3 Compositions of Linear Transformations

Theorem 9.3.1 Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then the *composition mapping* $T \circ S : U \to W$, defined by

$$(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) \quad \text{for } \boldsymbol{u} \in U,$$

is also a linear transformation.

Proof We use the result of Remark 9.1.3 to show that $T \circ S$ is a linear transformation:
Take any scalars a, b and vectors $\boldsymbol{u}, \boldsymbol{v} \in U$.

$$\begin{aligned} (T \circ S)(a\boldsymbol{u} + b\boldsymbol{v}) &= T(S(a\boldsymbol{u} + b\boldsymbol{v})) \\ &= T(aS(\boldsymbol{u}) + bS(\boldsymbol{v})) && \text{because } S \text{ is a linear transformation} \\ &= aT(S(\boldsymbol{u})) + bT(S(\boldsymbol{v})) && \text{because } T \text{ is a linear transformation} \\ &= a(T \circ S)(\boldsymbol{u}) + b(T \circ S)(\boldsymbol{v}). \end{aligned}$$

So $T \circ S$ is a linear transformation.

Example 9.3.2

1. Let $S : \mathbb{C}^3 \to \mathcal{M}_{2\times 2}(\mathbb{C})$ be the linear transformation defined by

$$S((a,b,c)) = \begin{pmatrix} a + \mathrm{i}c & 0 \\ 2b & a - \mathrm{i}c \end{pmatrix} \quad \text{for } (a,b,c) \in \mathbb{C}^3$$

and let $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$ be the linear transformation defined by

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) = a + (\mathrm{i}b + c)x - dx^2 \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{C}).$$

Then for $(a, b, c) \in \mathbb{C}^3$

$$\begin{aligned}(T \circ S)((a, b, c)) &= T(S((a, b, c))) \\ &= T\left(\begin{pmatrix} a + \mathrm{i}c & 0 \\ 2b & a - \mathrm{i}c \end{pmatrix}\right) \\ &= (a + \mathrm{i}c) + 2bx - (a - \mathrm{i}c)x^2\end{aligned}$$

is a linear transformation from $\mathbb{C}^3$ to $\mathcal{P}_2(\mathbb{C})$.

2. Let $[a, b]$, with $a < b$, be a closed interval on the real line and let $V = C^\infty([a, b])$. Consider the differential and integral operators D and F on V defined in Example 9.1.4.6. For every $f \in V$,

$$\begin{aligned}&(D \circ F)(f)(x) = D(F(f))(x) = \frac{d}{dx}\int_a^x f(t)dt = f(x) \quad \text{for } x \in [a, b] \\ \Rightarrow \quad &(D \circ F)(f) = f\end{aligned}$$

and

$$\begin{aligned}&(F \circ D)(f)(x) = F(D(f))(x) = \int_a^x \frac{d\,f(t)}{dt}\,dt = f(x) - f(a) \quad \text{for } x \in [a, b] \\ \Rightarrow \quad &(F \circ D)(f) \begin{cases} = f & \text{if } f(a) = 0 \\ \neq f & \text{if } f(a) \neq 0. \end{cases}\end{aligned}$$

Thus $D \circ F = I_V$ but $F \circ D \neq I_V$.

Theorem 9.3.3 Let $S : U \to V$ and $T : V \to W$ be linear transformations. Suppose U, V and W are finite dimensional where $\dim(U) \geq 1$, $\dim(V) \geq 1$ and $\dim(W) \geq 1$. Let A, B and C be ordered bases for U, V and W respectively. Then

$$[T \circ S]_{C,A} = [T]_{C,B}\,[S]_{B,A}.$$

(See also Theorem 7.1.11.)

Proof Let $A = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$, where $n = \dim(V)$, and let $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ be the standard basis for $\mathbb{F}^n$, where $\mathbb{F}$ is the field of scalars. In here, each $\boldsymbol{e_i}$ is written as a column vector. Note that for $i = 1, 2, \dots, n$,

$$\boldsymbol{v_i} = 0\boldsymbol{v_1} + \cdots + 0\boldsymbol{v_{i-1}} + 1\boldsymbol{v_i} + 0\boldsymbol{v_{i+1}} + \cdots + 0\boldsymbol{v_n} \quad \Rightarrow \quad [\boldsymbol{v_i}]_A = \boldsymbol{e_i}.$$

Hence

$$\begin{aligned}\text{the } i\text{th column of the matrix } [T \circ S]_{C,A} &= [(T \circ S)(\boldsymbol{v_i})]_C \\ &= [T(S(\boldsymbol{v_i}))]_C \\ &= [T]_{C,B}[S(\boldsymbol{v_i})]_B \\ &= [T]_{C,B}[S]_{B,A}[\boldsymbol{v_i}]_A \\ &= [T]_{C,B}[S]_{B,A}\boldsymbol{e_i} \\ &= \text{the } i\text{th column of the matrix } [T]_{C,B}[S]_{B,A}.\end{aligned}$$

So $[T \circ S]_{C,A} = [T]_{C,B}\,[S]_{B,A}$.

Example 9.3.4 Let $S : \mathbb{C}^3 \to \mathcal{M}_{2\times 2}(\mathbb{C})$ and $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$ be linear transformations defined in Example 9.3.2.1. Take the standard bases

$$
\begin{aligned}
A &= \{(1,0,0),\ (0,1,0),\ (0,0,1)\},\\
B &= \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}\\
\text{and}\ \ C &= \{1, x, x^2\}
\end{aligned}
$$

for $\mathbb{C}^4$, $\mathcal{M}_{2\times 2}(\mathbb{C})$ and $\mathcal{P}_2(\mathbb{C})$ respectively. Then

$$
\begin{aligned}
[S]_{B,A} &= \Big([S((1,0,0))]_B \quad [S((0,1,0))]_B \quad [S((0,0,1))]_B \Big)\\
&= \left(\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]_B \quad \left[\begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}\right]_B \quad \left[\begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}\right]_B \right)\\
&= \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & -\mathrm{i} \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
[T]_{C,B} &= \left(\left[T\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right)\right]_C \quad \left[T\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right)\right]_C \quad \left[T\left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right)\right]_C \quad \left[T\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right)\right]_C \right)\\
&= \Big([1]_C \quad [\mathrm{i}x]_C \quad [x]_C \quad [-x^2]_C \Big)\\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathrm{i} & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
[T\circ S]_{C,A} &= \Big([(T\circ S)((1,0,0))]_C \quad [(T\circ S)((0,1,0))]_C \quad [(T\circ S)((0,0,1))]_C \Big)\\
&= \Big([1-x^2]_C \quad [2x]_C \quad [\mathrm{i}+\mathrm{i}x^2]_C \Big)\\
&= \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 2 & 0 \\ -1 & 0 & \mathrm{i} \end{pmatrix}.
\end{aligned}
$$

Note that

$$
[T\circ S]_{C,A} = \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 2 & 0 \\ -1 & 0 & \mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathrm{i} & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & -\mathrm{i} \end{pmatrix} = [T]_{C,B}\,[S]_{B,A}.
$$

Definition 9.3.5 Let $T : V \to V$ be a linear operator. For any nonnegative integer m, we define T^m as follows:

$$T^m = \begin{cases} I_V & \text{if } m = 0 \\ \underbrace{T \circ T \circ \cdots \circ T}_{m \text{ times}} & \text{if } m \geq 1. \end{cases}$$

Corollary 9.3.6 Let $T : V \to V$ be a linear operator where V is finite dimensional where $\dim(V) \geq 1$. Let B be an ordered basis. Then $[T^m]_B = ([T]_B)^m$.

Proof Use Theorem 9.3.3 repeatedly.

Example 9.3.7

1. Let $\boldsymbol{A}$ be an $n \times n$ matrix over a field $\mathbb{F}$ and let $L_{\boldsymbol{A}} : \mathbb{F}^n \to \mathbb{F}^n$ be the linear operator defined in Example 9.1.4.1. Then $L_{\boldsymbol{A}}^m(\boldsymbol{u}) = \boldsymbol{A}^m \boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{F}^n$.

2. Let $[a, b]$, with $a < b$, be a closed interval on the real line and let $D : C^\infty([a, b]) \to C^\infty([a, b])$ be the differential operator defined in Example 9.1.4.6. Then for every $f \in C^\infty([a, b])$, $D^m(f)$ is a function on $C^\infty([a, b])$ such that

$$D^m(f)(x) = \frac{d^m f(x)}{dx^m} \quad \text{for } x \in [a, b].$$

Lemma 9.3.8 Let $T : V \to V$ be a linear operator where V is a finite dimensional vector space with $\dim(V) \geq 1$. Let B, C are two ordered bases for V and $\boldsymbol{P}$ the transition matrix from B to C, i.e. $\boldsymbol{P} = [I_V]_{C,B}$. Then

$$[T]_B = \boldsymbol{P}^{-1}[T]_C \boldsymbol{P}.$$

Proof Since $T = I_V \circ T \circ I_V$,

$$\begin{aligned} [T]_B &= [T]_{B,B} \\ &= [I_V \circ T \circ I_V]_{B,B} \\ &= [I_V \circ T]_{B,C} \, [I_V]_{C,B} \\ &= [I_V]_{B,C} \, [T]_{C,C} \, [I_V]_{C,B} \\ &= (\, [I_V]_{C,B})^{-1} \, [T]_C \, [I_V]_{C,B}. \end{aligned}$$

So we have $[T]_B = \boldsymbol{P}^{-1}[T]_C \boldsymbol{P}$.

Definition 9.3.9 Let $\mathbb{F}$ be a field and $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n \times n}(\mathbb{F})$. Then $\boldsymbol{B}$ is said to be *similar* to $\boldsymbol{A}$ if there exists an invertible matrix $\boldsymbol{P} \in \mathcal{M}_{n \times n}(\mathbb{F})$ such that $\boldsymbol{B} = \boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P}$.

Theorem 9.3.10 Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$, with $\dim(V) = n \geq 1$, and let C be an ordered basis for V. Then an $n \times n$ matrix $\boldsymbol{D}$ over $\mathbb{F}$ is similar to $[T]_C$ if and only if there exists an ordered basis B for V such that $\boldsymbol{D} = [T]_B$.

Proof

($\Leftarrow$) The result follows from Lemma 9.3.8.

($\Rightarrow$) Let $C = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_n}\}$. Suppose $\boldsymbol{D} = \boldsymbol{P}^{-1}[T]_C\boldsymbol{P}$ where $\boldsymbol{P} = (p_{ij})_{n\times n}$ is an invertible matrix. Define $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ where

$$\boldsymbol{v_j} = p_{1j}\boldsymbol{u_1} + p_{2j}\boldsymbol{u_2} + \cdots + p_{nj}\boldsymbol{u_n}$$

for $j = 1, 2, \dots, n$. Since $\boldsymbol{P}$ is invertible, by Question 8.18, $\text{span}(B) = \text{span}(C) = V$ and hence by Theorem 8.5.13, B is a basis for V. Using B as an ordered basis for V, $[I_V]_{C,B} = \begin{pmatrix} [\boldsymbol{v_1}]_C & [\boldsymbol{v_2}]_C & \cdots & [\boldsymbol{v_n}]_C \end{pmatrix} = \boldsymbol{P}$ and hence

$$[T]_B = [I_V]_{B,C}\,[T]_C\,[I_V]_{C,B} = \boldsymbol{P}^{-1}[T]_C\boldsymbol{P} = \boldsymbol{D}.$$

Example 9.3.11 Consider the real matrix $\boldsymbol{A} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}$. Let $B = \{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \boldsymbol{v_3}\}$ where

$$\boldsymbol{v_1} = \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}, \quad \boldsymbol{v_2} = \begin{pmatrix} -1 \\ \sqrt{2} \\ 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{v_3} = \begin{pmatrix} -1 \\ -\sqrt{2} \\ 1 \end{pmatrix}.$$

Then $L_{\boldsymbol{A}}(\boldsymbol{v_1}) = \boldsymbol{v_1}$, $L_{\boldsymbol{A}}(\boldsymbol{v_2}) = \sqrt{2}\,\boldsymbol{v_2}$ and $L_{\boldsymbol{A}}(\boldsymbol{v_3}) = -\sqrt{2}\,\boldsymbol{v_3}$. (See Example 6.1.12.3.) Hence

$$\begin{aligned} [L_{\boldsymbol{A}}]_B &= \begin{pmatrix} [L_{\boldsymbol{A}}(\boldsymbol{v_1})]_B & [L_{\boldsymbol{A}}(\boldsymbol{v_2})]_B & [L_{\boldsymbol{A}}(\boldsymbol{v_3})]_B \end{pmatrix} \\ &= \begin{pmatrix} [\boldsymbol{v_1}]_B & \left[\sqrt{2}\,\boldsymbol{v_2}\right]_B & \left[-\sqrt{2}\,\boldsymbol{v_3}\right]_B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & -\sqrt{2} \end{pmatrix}. \end{aligned}$$

Using the standard basis $E = \{(1,0,0),\ (0,1,0),\ (0,0,1)\}$ for $\mathbb{R}^3$, $[L_{\boldsymbol{A}}]_E = \boldsymbol{A}$ and

$$[I_V]_{E,B} = \begin{pmatrix} [\boldsymbol{v_1}]_E & [\boldsymbol{v_2}]_E & [\boldsymbol{v_3}]_E \end{pmatrix} = \begin{pmatrix} -2 & -1 & -1 \\ 2 & \sqrt{2} & -\sqrt{2} \\ 1 & 1 & 1 \end{pmatrix}.$$

With $\boldsymbol{P} = [I_V]_{E,B}$, we have $[L_{\boldsymbol{A}}]_B = [I_V]_{B,E}\,[L_{\boldsymbol{A}}]_E\,[I_V]_{E,B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$. (See also Example 6.2.6.2.)

Remark 9.3.12 Using linear operators, we have a new interpretation of the problem of diagonalization discussed in Section 6.2. Given a square matrix $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$, we use the linear operator $L_{\boldsymbol{A}}$ on $\mathbb{F}^n$ defined in Example 9.1.4.1. By Example 9.2.4.3, $[L_{\boldsymbol{A}}]_E = \boldsymbol{A}$ where E is the standard basis for $\mathbb{F}^n$. Then to diagonalize $\boldsymbol{A}$ is the same as to find an ordered basis B for $\mathbb{F}^n$ such that the matrix for $L_{\boldsymbol{A}}$ relative to B is a diagonal matrix. (See also Chapter 11.)

Section 9.4 The Vector Space $\mathcal{L}(V, W)$

Definition 9.4.1 Let V and W be vector spaces over the same field $\mathbb{F}$.

1. Let T_1, $T_2 : V \to W$ be linear transformations. We define a mapping $T_1 + T_2 : V \to W$ by
$$(T_1 + T_2)(\boldsymbol{u}) = T_1(\boldsymbol{u}) + T_2(\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \in V.$$
2. Let $T : V \to W$ be a linear transformation and $c \in \mathbb{F}$. We define a mapping $cT : V \to W$ by
$$(cT)(\boldsymbol{u}) = cT(\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \in V.$$

The mappings $T_1 + T_2$ and cT are linear transformations.

Example 9.4.2 Let T_1, $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ be given by
$$T_1((x, y)) = (x + 2y,\ 5x - 6y) \quad \text{and} \quad T_2((x, y)) = (3x + 4y,\ y - x) \quad \text{for } (x, y) \in \mathbb{R}^2.$$
Find the formulae for $T_1 + T_2$, $3T_1$, $-5T_2$ and $3T_1 - 5T_2$.

Solution For $(x, y) \in \mathbb{R}^2$,
$$\begin{aligned}
(T_1 + T_2)((x, y)) &= T_1((x, y)) + T_2((x, y)) \\
&= (x + 2y,\ 5x - 6y) + (3x + 4y,\ y - x) = (4x + 6y,\ 4x - 5y), \\
(3T_1)((x, y)) &= 3T_1((x, y)) = 3(x + 2y,\ 5x - 6y) = (3x + 6y,\ 15x - 18y), \\
(-5T_2)((x, y)) &= -5T_2((x, y)) = -5(3x + 4y,\ y - x) = (-15x - 20y,\ 5x - 5y), \\
(3T_1 - 5T_2)((x, y)) &= (3T_1)((x, y)) + (-5T_2)((x, y)) \\
&= (3x + 6y,\ 15x - 18y) + (-15x - 20y,\ 5x - 5y) \\
&= (-12x - 14y,\ 20x - 23y).
\end{aligned}$$

Proposition 9.4.3 Let V and W be finite dimensional vector space over the same field $\mathbb{F}$ where $\dim(V) \geq 1$ and $\dim(W) \geq 1$ and let B and C be ordered basis for V and W respectively.

1. If T_1, $T_2 : V \to W$ are linear transformations, then $[T_1 + T_2]_{C,B} = [T_1]_{C,B} + [T_2]_{C,B}$.
2. If $T : V \to W$ be a linear transformation and $c \in \mathbb{F}$, then $[cT]_{C,B} = c[T]_{C,B}$.

Proof The proof is left as exercise. See Question 9.20

Remark 9.4.4 Matrices and linear transformations have a lot of similarities. The observations above show their relations in addition and scalar multiplication. By Theorem 9.3.3, we have also seen that the composition of linear transformations is equivalent to matrix multiplication. In the later sections, we shall learnt that the matrix inverse has a corresponding analog in linear transformations (see Theorem 9.6.6). Thus we may sometimes regard linear transformations as generalized matrices.

Theorem 9.4.5 Let V and W be vector spaces over the same field $\mathbb{F}$, and let $\mathcal{L}(V, W)$ be the set of all linear transformations from V to W. Then $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$ with addition and scalar multiplication defined in Definition 9.4.1.

Furthermore, if V and W are finite dimensional, then

$$\dim(\mathcal{L}(V, W)) = \dim(V)\ \dim(W).$$

Proof It is straight forward to check that $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$. In particular, the zero vector in $\mathcal{L}(V, W)$ is the zero transformation $O_{V,W}$ and the negative of T is the linear transformation $-T = (-1)T$.

Now, let V and W be finite dimensional. If V or W is a zero space, then $\mathcal{L}(V, W) = \{O_{V,W}\}$ and $\dim(\mathcal{L}(V, W)) = 0 = \dim(V)\ \dim(W)$. Suppose $\dim(V) \geq 1$ and $\dim(W) \geq 1$. Let $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ and $C = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_m}\}$ be bases for V and W respectively. For each pair of i, j, where $1 \leq i \leq m$ and $1 \leq j \leq n$, define a linear transformation $T_{ij} : V \to W$ such that for $k = 1, 2, \ldots, n$,

$$T_{ij}(\boldsymbol{v_k}) = \begin{cases} \boldsymbol{w_i} & \text{if } k = j \\ \boldsymbol{0} & \text{if } k \neq j. \end{cases}$$

(See Remark 9.1.6.) It can be shown that $\{\, T_{ij} \mid 1 \leq i \leq m \text{ and } 1 \leq j \leq n\}$ is a basis for $\mathcal{L}(V, W)$. Since this basis has mn elements, the dimension formula follows.

(Note that $[T_{ij}]_{C,B} = \boldsymbol{E_{ij}}$ where $\boldsymbol{E_{ij}}$ is the matrix defined in Example 8.4.6.7.)

Definition 9.4.6 If V is a vector space over $\mathbb{F}$, then the vector space $\mathcal{L}(V, \mathbb{F})$ of all linear functionals on V is called the *dual space* of V and is denoted by V^*. By Theorem 9.4.5, if V is finite dimensional, then $\dim(V) = \dim(V^*)$.

Example 9.4.7 Let $V = \mathbb{F}^n$ where vectors in V are written as column vectors. For each $f \in V^*$, there exists scalars $a_1, a_2, \ldots, a_n$ such that for all $(x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in V$,

$$f\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

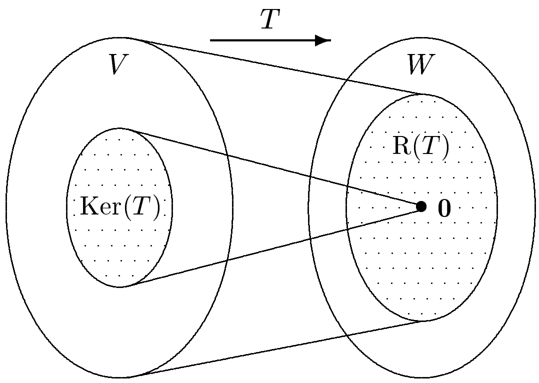
So each linear functional in V^* can be represented by a row vector over $\mathbb{F}$. (See also Example 9.6.14.4.)

Section 9.5 Kernels and Ranges

Definition 9.5.1 Let $T : V \to W$ be a linear transformation.

1. The subset $\mathrm{Ker}(T) = \{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) = \mathbf{0}\}$ of V is called the *kernel* of T.
   (In some textbooks, $\mathrm{Ker}(T)$ is called the *nullspace* of T and is denoted by $\mathrm{N}(T)$.)

2. The subset $\mathrm{R}(T) = \{T(\boldsymbol{u}) \mid \boldsymbol{u} \in V\}$ of W is called the *range* of T.



Theorem 9.5.2 Let $T : V \to W$ be a linear transformation. Then

1. $\mathrm{Ker}(T)$ is a subspace of V; and

2. $\mathrm{R}(T)$ is a subspace of W.

Proof The proof is left as an exercise. See Question 9.28.

Example 9.5.3

1. Let $T : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathcal{M}_{2\times 3}(\mathbb{R})$ be the linear transformation defined by

$$T\left(\begin{pmatrix} w & x \\ y & z \end{pmatrix}\right) = \begin{pmatrix} w - x + z & -x + 2y & x - z \\ x - 2y & w + x - z & 0 \end{pmatrix} \quad \text{for } \begin{pmatrix} w & x \\ y & z \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{R}).$$

$$T\left(\begin{pmatrix} w & x \\ y & z \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \Leftrightarrow \quad \left\{\begin{array}{rrrl} w - x & & + z & = 0 \\ -x + 2y & & & = 0 \\ x & & - z & = 0 \\ x - 2y & & & = 0 \\ w + x & & - z & = 0 \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} w = 0 \\ x = t \\ y = \frac{1}{2}t \\ z = t \end{array}\right. \quad \text{for } t \in \mathbb{R}.$$

So $\operatorname{Ker}(T) = \left\{ \begin{pmatrix} 0 & t \\ \frac{1}{2}t & t \end{pmatrix} \middle| \, t \in \mathbb{R} \right\}$ and $\left\{ \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & 1 \end{pmatrix} \right\}$ is a basis for $\operatorname{Ker}(T)$.

Since

$$\begin{aligned} T\left(\begin{pmatrix} w & x \\ y & z \end{pmatrix}\right) &= \begin{pmatrix} w-x+z & -x+2y & x-z \\ x-2y & w+x-z & 0 \end{pmatrix} \\ &= w\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + x\begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix} + y\begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \end{pmatrix} + z\begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \end{aligned}$$

$$\mathrm{R}(T) = \operatorname{span}\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} \right\}.$$

Using the standard basis $\{\boldsymbol{E_{11}}, \boldsymbol{E_{12}}, \boldsymbol{E_{13}}, \boldsymbol{E_{21}}, \boldsymbol{E_{22}}, \boldsymbol{E_{23}}\}$, the four matrices are converted to coordinate vectors $(1,0,0,0,1,0)$, $(-1,-1,1,1,1,0)$, $(0,2,0,-2,0,0)$, $(1,0,-1,0,-1,0)$.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & -2 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \end{pmatrix} \xrightarrow[\text{Elimination}]{\text{Gaussian}} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus $\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 1 \\ 1 & 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 2 \\ 0 & 4 & 0 \end{pmatrix} \right\}$ is a basis for $\mathrm{R}(T)$.

2. Let $I_V : V \to V$ be the identity operator defined in Example 9.1.4.2. Then $\operatorname{Ker}(I_V) = \{\mathbf{0}\}$ and $\mathrm{R}(T) = V$.

3. Let $O_{V,W} : V \to W$ be the zero transformation defined in Example 9.1.4.3. Then $\operatorname{Ker}(O_{V,W}) = V$ and $\mathrm{R}(O_{V,W}) = \{\mathbf{0}\}$.

4. Let $[a,b]$, with $a < b$, be a closed interval on the real line and let $D : C^\infty([a,b]) \to C^\infty([a,b])$ be the differential operator defined in Example 9.1.4.6.

   For $f \in C^\infty([a,b])$, $D(f) = \mathit{0}$, where $\mathit{0}$ is the zero function, if and only if f is a constant function, i.e. $f = c\mathit{1}$ for some $c \in \mathbb{R}$ where $\mathit{1}$ is the function on $C^\infty([a,b])$ defined by $\mathit{1}(x) = 1$ for $x \in [a,b]$. So $\operatorname{Ker}(D) = \operatorname{span}\{\mathit{1}\}$.

   Take $f \in C^\infty([a,b])$. Let $g \in C^\infty([a,b])$ be the function defined by $g(x) = \displaystyle\int_a^x f(t)dt$ for $x \in [a,b]$. Then $D(g) = f$. So $\mathrm{R}(D) = C^\infty([a,b])$.

5. Let $[a,b]$, with $a < b$, be a closed interval on the real line and let $F : C^\infty([a,b]) \to C^\infty([a,b])$ be the integral operator defined in Example 9.1.4.6. Then $\operatorname{Ker}(F) = \{\mathit{0}\}$ and $\mathrm{R}(F) = \{h \in C^\infty([a,b]) \mid h(a) = 0\}$.

Definition 9.5.4 Let $T : V \to W$ be a linear transformation.

1. If $\mathrm{Ker}(T)$ is finite dimensional, then $\dim(\mathrm{Ker}(T))$ is called the *nullity* of T and is denoted by $\mathrm{nullity}(T)$.

2. If $\mathrm{R}(T)$ is finite dimensional, then $\dim(\mathrm{R}(T))$ is called the *rank* of T and is denoted by $\mathrm{rank}(T)$.

Example 9.5.5 Consider linear transformations defined in Example 9.5.3.

1. $\mathrm{nullity}(T) = 1$ and $\mathrm{rank}(T) = 3$.

2. $\mathrm{nullity}(I_V) = 0$ and if V is finite dimensional, then $\mathrm{rank}(I_V) = \dim(V)$.

3. $\mathrm{rank}(O_{V,W}) = 0$ and if V is finite dimensional, then $\mathrm{nullity}(O_{V,W}) = \dim(V)$.

4. $\mathrm{nullity}(D) = 1$.

5. $\mathrm{nullity}(F) = 0$.

Lemma 9.5.6 Let $T : V \to W$ be a linear transformation where V and W are finite dimensional with $\dim(V) \geq 1$ and $\dim(W) \geq 1$. For any ordered bases B and C for V and W respectively,

1. $\{ [\boldsymbol{u}]_B \mid \boldsymbol{u} \in \mathrm{Ker}(T)\}$ is the nullspace of $[T]_{C,B}$ and $\mathrm{nullity}(T) = \mathrm{nullity}([T]_{C,B})$; and

2. $\{ [\boldsymbol{v}]_C \mid \boldsymbol{v} \in \mathrm{R}(T)\}$ is the column space of $[T]_{C,B}$ and $\mathrm{rank}(T) = \mathrm{rank}([T]_{C,B})$.

Proof The proof is left as an exercise. See Question 9.30.

Theorem 9.5.7 (Dimension Theorem for Linear Transformations) Let $T : V \to W$ be a linear transformation where V and W are finite dimensional. Then

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

Proof If $V = \{\mathbf{0}\}$ is a zero space, then $\mathrm{Ker}(T) = V = \{\mathbf{0}\}$ and $\mathrm{R}(T) = \{\mathbf{0}_W\}$, where $\mathbf{0}_W$ is the zero vector in W, and hence

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{R}(T)) = 0 + 0 = 0 = \dim(V).$$

If $W = \{\mathbf{0}\}$ is a zero space, then $\mathrm{Ker}(V) = V$ and $\mathrm{R}(T) = W = \{\mathbf{0}\}$ and hence

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{R}(T)) = \dim(V) + 0 = \dim(V).$$

Suppose $\dim(V) \geq 1$ and $\dim(W) \geq 1$. Let B and C be ordered bases for V and W respectively. By Lemma 9.5.6 and the Dimension Theorem for Matrices (Theorem 4.3.4),

$$\begin{aligned}\operatorname{rank}(T) + \operatorname{nullity}(T) &= \dim(\operatorname{Ker}(T)) + \dim(\operatorname{R}(T)) \\ &= \operatorname{rank}([T]_{C,B}) + \operatorname{nullity}([T]_{C,B}) \\ &= \text{ the number of columns in } [T]_{C,B} \\ &= \dim(V).\end{aligned}$$

Example 9.5.8 In Example 9.5.3.1, $\operatorname{rank}(T) + \operatorname{nullity}(T) = 3 + 1 = 4 = \dim(\mathcal{M}_{2\times 2}(\mathbb{R}))$.

Theorem 9.5.9 Let $T : V \to W$ be a linear transformation. Suppose B and C are subsets of V such that B is a basis for $\operatorname{Ker}(T)$, $\{T(\boldsymbol{v}) \mid \boldsymbol{v} \in C\}$ is a basis for $\operatorname{R}(T)$ and for any $\boldsymbol{v}, \boldsymbol{v'} \in C$, if $\boldsymbol{v} \neq \boldsymbol{v'}$, then $T(\boldsymbol{v}) \neq T(\boldsymbol{v'})$. Then $B \cup C$ is a basis for V.

Proof The proof is left as exercise. (See Question 9.31.)

Remark 9.5.10 When V is finite dimensional, Theorem 9.5.9 gives us a direct proof of the Dimension Theorem for Linear Transformations without using matrices.

Definition 9.5.11 Let $f : A \to B$ be a mapping.

1. f is called *injective* or *one-to-one* if for every $z \in B$, there exists at most one $x \in A$ such that $f(x) = z$.

2. f is called *surjective* or *onto* if for every $z \in B$, there exists at least one $x \in A$ such that $f(x) = z$.

3. f is called *bijective* if f is both injective and surjective, i.e. for every $z \in B$, there exists one and only one $x \in A$ such that $f(x) = z$.

Proposition 9.5.12 Let $T : V \to W$ be a linear transformation.

1. T is injective (one-to-one) if and only if $\operatorname{Ker}(T) = \{\mathbf{0}\}$ if and only if $\operatorname{nullity}(T) = 0$.

2. T is surjective (onto) if and only if $\operatorname{R}(T) = W$.

   If W is finite dimensional, then T is surjective if and only if $\operatorname{rank}(T) = \dim(W)$.

Proof We only prove that T is injective (one-to-one) if and only if $\operatorname{Ker}(T) = \{\mathbf{0}\}$. The other parts are obvious.

($\Rightarrow$) By Proposition 9.1.5, $T(\mathbf{0}) = \mathbf{0}$. On the other hand, since T is injective, there is only one vector maps to the zero vector in W. Hence $\operatorname{Ker}(T) = \{\mathbf{0}\}$.

($\Leftarrow$) Suppose $\text{Ker}(T) = \{\mathbf{0}\}$.

Assume for some $\boldsymbol{w} \in W$, there exist $\boldsymbol{u}, \boldsymbol{v} \in V$ such that $T(\boldsymbol{u}) = \boldsymbol{w}$ and $T(\boldsymbol{v}) = \boldsymbol{w}$. Then

$$T(\boldsymbol{u} - \boldsymbol{v}) = T(\boldsymbol{u}) - T(\boldsymbol{v}) = \boldsymbol{w} - \boldsymbol{w} = \mathbf{0}.$$

It means $\boldsymbol{u} - \boldsymbol{v} \in \text{Ker}(T)$. As $\text{Ker}(T) = \{\mathbf{0}\}$, we have $\boldsymbol{u} - \boldsymbol{v} = \mathbf{0}$ and hence $\boldsymbol{u} = \boldsymbol{v}$.

So T is injective.

Example 9.5.13 Consider linear transformations defined in Example 9.5.3. T and $O_{V,W}$ (when both V and W are not zero spaces) are neither injective nor surjective; I_V is both injective and surjective, i.e. I_V is bijective; D is surjective but not injective; and F is injective but not surjective.

Section 9.6 Isomorphisms

Definition 9.6.1 Let $T : V \to W$ be a linear transformation. Then T is called an *isomorphism from V onto W* if T is bijective.

Example 9.6.2

1. For any vector space V, the identity operator on V is an isomorphism.

2. Let $\mathbb{F}$ be a field and let $T : \mathbb{F}^3 \to \mathcal{P}_2(\mathbb{F})$ be the linear transformation defined by

$$T((a, b, c)) = a + (a + b)x + (a + b + c)x^2 \quad \text{for } (a, b, c) \in \mathbb{F}^3.$$

Since

$$T((a,b,c)) = 0 \quad \Leftrightarrow \quad \begin{cases} a & = 0 \\ a+b & = 0 \\ a+b+c & = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} a = 0 \\ b = 0 \\ c = 0, \end{cases}$$

$\text{Ker}(T) = \{(0, 0, 0)\}$ and hence T is injective.

For any $d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F})$,

$$T((a,b,c)) = d + ex + fx^2 \quad \Leftrightarrow \quad \begin{cases} a & = d \\ a+b & = e \\ a+b+c & = f \end{cases} \quad \Leftrightarrow \quad \begin{cases} a = d \\ b = e - d \\ c = f - e. \end{cases}$$

Thus $T((d, e - d, f - e)) = d + ex + fx^2$ for all $d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F})$. It means that T is surjective.

As T is a bijective linear transformation, T is an isomorphism.

3. Define $P : \mathbb{R}^2 \to \mathbb{R}^3$ to be a linear transformation such that $P((x, y)) = (x, y, 0)$ for $(x, y) \in \mathbb{R}^2$. It is obvious that P is injective but not surjective. So P is not an isomorphism.

   Let W be the xy-plane in $\mathbb{R}^3$. Note that $\mathrm{R}(P) = W$. Define $P' : \mathbb{R}^2 \to W$ such that $P((x, y)) = (x, y, 0)$ for $(x, y) \in \mathbb{R}^2$. Then P' is a bijective linear transformation and hence P' is an isomorphism.

Definition 9.6.3 A mapping $T : V \to W$ is bijective if and only if there exists a mapping $S : W \to V$ such that $S \circ T = I_V$ and $T \circ S = I_W$ where I_V and I_W are identity operators on V and W respectively. The mapping S is known as the *inverse* of T and is denoted by T^{-1}. Thus a bijective mapping is also called an *invertible mapping*.

Theorem 9.6.4 If T is an isomorphism, then T^{-1} is a linear transformation and hence is also an isomorphism.

Proof Suppose T maps V to W where V and W are vector spaces over a field $\mathbb{F}$. Take any $\boldsymbol{w_1}, \boldsymbol{w_2} \in W$. Let $\boldsymbol{v_1} = T^{-1}(\boldsymbol{w_1})$ and $\boldsymbol{v_2} = T^{-1}(\boldsymbol{w_2})$. Note that

$$T(\boldsymbol{v_1}) = T(T^{-1}(\boldsymbol{w_1})) = (T \circ T^{-1})(\boldsymbol{w_1}) = I_W(\boldsymbol{w_1}) = \boldsymbol{w_1}$$

and similarly, $T(\boldsymbol{v_2}) = \boldsymbol{w_2}$. Then for any $a, b \in \mathbb{F}$,

$$\begin{aligned} T^{-1}(a\boldsymbol{w_1} + b\boldsymbol{w_2}) &= T^{-1}(aT(\boldsymbol{v_1}) + bT(\boldsymbol{v_2})) \\ &= T^{-1}(T(a\boldsymbol{v_1} + b\boldsymbol{v_2})) \qquad \text{(because } T \text{ is a linear transformation)} \\ &= (T^{-1} \circ T)(a\boldsymbol{v_1} + b\boldsymbol{v_2}) \\ &= I_V(a\boldsymbol{v_1} + b\boldsymbol{v_2}) \\ &= a\boldsymbol{v_1} + b\boldsymbol{v_2} = aT^{-1}(\boldsymbol{w_1}) + bT^{-1}(\boldsymbol{w_2}). \end{aligned}$$

So T^{-1} is a linear transformation. As the inverse of a bijective mapping is also bijective, T^{-1} is also an isomorphism.

Example 9.6.5 Let $\mathbb{F}$ be a field and let $T : \mathbb{F}^3 \to \mathcal{P}_2(\mathbb{F})$ be the isomorphism defined in Example 9.6.2.2. Define a mapping $S : \mathcal{P}_2(\mathbb{F}) \to \mathbb{F}^3$ such that

$$S(d + ex + fx^2) = (d,\ e - d,\ f - e) \ \text{ for } d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F}).$$

Then for all $(a, b, c) \in \mathbb{F}^3$,

$$\begin{aligned} (S \circ T)((a, b, c)) &= S(T((a, b, c))) \\ &= S(a + (a + b)x + (a + b + c)x^2) \\ &= (a,\ (a + b) - a,\ (a + b + c) - (a + b)) = (a, b, c). \end{aligned}$$

and for all $d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F})$,

$$\begin{aligned}(T \circ S)(d + ex + fx^2) &= T(S(d + ex + fx^2)) \\ &= T((d,\, e - d,\, f - e)) \\ &= d + [d + (e - d)]x + [d + (e - d) + (f - e)]x^2 = d + ex + fx^2.\end{aligned}$$

It means $S \circ T = I_{\mathbb{F}^3}$ and $T \circ S = I_{\mathcal{P}_2(\mathbb{F})}$. So $S = T^{-1}$. Note that S is also a linear transformation and hence is an isomorphism.

Theorem 9.6.6 Let $T : V \to W$ be a linear transformation where V and W are finite dimensional with $\dim(V) = \dim(W) \geq 1$. Let B and C be ordered bases for V and W respectively.

1. T is an isomorphism if and only if $[T]_{C,B}$ is an invertible matrix.
2. If T is an isomorphism, $[T^{-1}]_{B,C} = (\,[T]_{C,B})^{-1}$.

Proof The proof is left as exercise. See Question 9.42.

Example 9.6.7 Let $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{M}_{2\times 2}(\mathbb{R})$ be the mapping defined by

$$T(p(x)) = \begin{pmatrix} p(0) & p(1) \\ p(-1) & p(2) \end{pmatrix} \quad \text{for } p(x) \in \mathcal{P}_3(\mathbb{R}).$$

T is a linear transformation. (Check it.) Let $B = \{1, x, x^2, x^3\}$ and $C = \{\boldsymbol{A_1},\, \boldsymbol{A_2},\, \boldsymbol{A_3},\, \boldsymbol{A_4}\}$ where $\boldsymbol{A_1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{A_2} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{A_3} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{A_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Note that $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\boldsymbol{A_1} + (-a + b)\boldsymbol{A_2} + (-a + c)\boldsymbol{A_3} + (a - b - c + d)\boldsymbol{A_4}$. Then

$$\begin{aligned}[T]_{C,B} &= \begin{pmatrix} [T(1)]_C & [T(x)]_C & [T(x^2)]_C & [T(x^3)]_C \end{pmatrix} \\ &= \left(\left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right]_C \; \left[\begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}\right]_C \; \left[\begin{pmatrix} 0 & 1 \\ 1 & 4 \end{pmatrix}\right]_C \; \left[\begin{pmatrix} 0 & 1 \\ -1 & 8 \end{pmatrix}\right]_C \right) \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 2 & 2 & 8 \end{pmatrix}.\end{aligned}$$

Since $\det([T]_{C,B}) = 12 \neq 0$, $[T]_{C,B}$ is invertible and hence T is an isomorphism.

Furthermore,

$$[T^{-1}]_{B,C} = (\,[T]_{C,B})^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 2 & 2 & 8 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{5}{6} & -\frac{1}{2} & -\frac{1}{6} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & 0 & \frac{1}{6} \end{pmatrix}.$$

For $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{R})$,

$$\begin{aligned}
\left[T^{-1}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)\right]_B &= [T^{-1}]_{B,C}\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right]_C \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{5}{6} & -\frac{1}{2} & -\frac{1}{6} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{1}{3} & 0 & \frac{1}{6} \end{pmatrix}\begin{pmatrix} a \\ -a+b \\ -a+c \\ a-b-c+d \end{pmatrix} = \begin{pmatrix} a \\ -\frac{1}{2}a+b-\frac{1}{3}c-\frac{1}{6}d \\ -a+\frac{1}{2}b+\frac{1}{2}c \\ \frac{1}{2}a-\frac{1}{2}b-\frac{1}{6}c+\frac{1}{6}d \end{pmatrix}.
\end{aligned}$$

Thus

$$T^{-1}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = a + (-\tfrac{1}{2}a + b - \tfrac{1}{3}c - \tfrac{1}{6}d)x + (-a + \tfrac{1}{2}b + \tfrac{1}{2}c)x^2 + (\tfrac{1}{2}a - \tfrac{1}{2}b - \tfrac{1}{6}c + \tfrac{1}{6}d)x^3.$$

Theorem 9.6.8 Let $S : W \to V$ and $T : V \to W$ be linear transformations such that $T \circ S = I_W$.

1. S is injective and T is surjective.

2. If V and W are finite dimensional and $\dim(V) = \dim(W)$, then S and T are isomorphisms, $S^{-1} = T$ and $T^{-1} = S$. (See also Theorem 2.4.12.)

Proof The proof is left as an exercise. See Question 9.46.

Remark 9.6.9 If V and W are infinite dimensional, Theorem 9.6.8.2 is in general not true. For example, in Example 9.3.2.2, we have $D \circ F = I_V$ but neither D nor F is an isomorphism.

Definition 9.6.10 Let V and W be vector spaces over a field $\mathbb{F}$. If there exists an isomorphism from V onto W, then V is said to be *isomorphic* to W and we write $V \cong_{\mathbb{F}} W$ or simply $V \cong W$.

Example 9.6.11 Let $\mathbb{F}$ be a field.

1. Let $T_1 : \mathcal{P}_n(\mathbb{F}) \to \mathbb{F}^{n+1}$ be the mapping defined by

$$T_1(a_0 + a_1x + \cdots + a_nx^n) = (a_0, a_1, \ldots, a_n) \quad \text{for } a_0 + a_1x + \cdots + a_nx^n \in \mathcal{P}_n(\mathbb{F}).$$

   T_1 is an isomorphism and hence $\mathcal{P}_n(\mathbb{F}) \cong_{\mathbb{F}} \mathbb{F}^{n+1}$.

2. Let $T_2 : \mathcal{F}(\mathbb{N}, \mathbb{F}) \to \mathbb{F}^{\mathbb{N}}$ be the mapping defined by

$$T_2(f) = (f(n))_{n\in\mathbb{N}} = (f(1),\ f(2),\ f(3),\ \ldots) \quad \text{for } f : \mathbb{N} \to \mathbb{F} \text{ in } \mathcal{F}(\mathbb{N}, \mathbb{F}).$$

   T_2 is an isomorphism and hence $\mathcal{F}(\mathbb{N}, \mathbb{F}) \cong_{\mathbb{F}} \mathbb{F}^{\mathbb{N}}$.

Remark 9.6.12 The term "isomorphic" is used in abstract algebra to indicate that two algebraic objects have the same structure. For example, by Example 9.6.11.1, $\mathcal{P}_n(\mathbb{F})$ is isomorphic to $\mathbb{F}^{n+1}$ and hence $\mathcal{P}_n(\mathbb{F})$ and $\mathbb{F}^{n+1}$ are the "same" as vector spaces. However, we do not want to say that $\mathcal{P}_n(\mathbb{F})$ is "equal" to $\mathbb{F}^{n+1}$ because they are still different in other aspects. In particular, we can multiply two polynomials in $\mathcal{P}_n(\mathbb{F})$ but cannot multiply two vectors in $\mathbb{F}^{n+1}$.

Theorem 9.6.13 Let V and W be finite dimensional vector spaces over the same field. Then V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

Proof

($\Rightarrow$) Suppose $V \cong W$, i.e. there exits an isomorphism $T : V \to W$. Let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ be a basis for V. We claim that $\{T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \ldots, T(\boldsymbol{v_n})\}$ is a basis for W:

(a) Take any $\boldsymbol{w} \in W$. Since T is surjective, there exists $\boldsymbol{u} \in V$ such that $T(\boldsymbol{u}) = \boldsymbol{w}$. As $\boldsymbol{u} \in \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$, $\boldsymbol{w} = T(\boldsymbol{u}) \in \operatorname{span}\{T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \ldots, T(\boldsymbol{v_n})\}$.

So we have shown that $W = \operatorname{span}\{T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \ldots, T(\boldsymbol{v_n})\}$.

(b) Consider the vector equation

$$c_1 T(\boldsymbol{v_1}) + c_2 T(\boldsymbol{v_2}) + \cdots + c_n T(\boldsymbol{v_n}) = \mathbf{0}.$$

But then $T(c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n}) = c_1 T(\boldsymbol{v_1}) + c_2 T(\boldsymbol{v_2}) + \cdots + c_n T(\boldsymbol{v_n}) = \mathbf{0}$ and hence $c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n} \in \operatorname{Ker}(T)$. Since T is injective, $\operatorname{Ker}(T) = \{\mathbf{0}\}$. Thus $c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n} = \mathbf{0}$. As $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}$ are linearly independent, $c_1 = c_2 = \cdots = c_n = 0$.

So we have shown that $T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \ldots, T(\boldsymbol{v_n})$ are linearly independent.

By (a) and (b), $\{T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \ldots, T(\boldsymbol{v_n})\}$ is a basis for W.

Hence $\dim(W) = n = \dim(V)$.

($\Leftarrow$) Suppose $\dim(V) = \dim(W) = n$. Let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ and $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n}\}$ be bases for V and W respectively.

Define a linear transformation $T : V \to W$ such that $T(\boldsymbol{v_i}) = \boldsymbol{w_i}$ for $i = 1, 2, \ldots, n$, i.e. for any $\boldsymbol{u} = a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_n\boldsymbol{v_n} \in V$,

$$T(\boldsymbol{u}) = a_1\boldsymbol{w_1} + a_2\boldsymbol{w_2} + \cdots + a_n\boldsymbol{w_n}.$$

It can be shown that T is an isomorphism (check it). So $V \cong W$.

Example 9.6.14 Let $\mathbb{F}$ be a field.

1. $\mathcal{M}_{m\times n}(\mathbb{F}) \cong_{\mathbb{F}} \mathbb{F}^{mn}$.

2. $\mathcal{P}_n(\mathbb{F}) \cong_{\mathbb{F}} \mathbb{F}^{n+1}$.

3. $\mathbb{C}^n \cong_{\mathbb{R}} \mathbb{R}^{2n}$. (In here, $\mathbb{C}^n$ is regarded as a vector space over $\mathbb{R}$.)

4. Suppose V and W are finite dimensional vector spaces over $\mathbb{F}$ such that $\dim(V) = n$ and $\dim(W) = m$. Then $\mathcal{L}(V, W) \cong_{\mathbb{F}} \mathbb{F}^{mn} \cong_{\mathbb{F}} \mathcal{M}_{m\times n}(\mathbb{F})$.

   In particular, V is isomorphic to its dual space V^* defined in Definition 9.4.6. (This result is not true if V is infinite dimensional.)

Theorem 9.6.15 (The First Isomorphism Theorem) Let $T : V \to W$ be a linear transformation. Then $V/\mathrm{Ker}(T) \cong \mathrm{R}(T)$.

Proof Define a mapping $S : V/\mathrm{Ker}(T) \to \mathrm{R}(T)$ such that

$$S(\mathrm{Ker}(T) + \boldsymbol{u}) = T(\boldsymbol{u}) \quad \text{for } \mathrm{Ker}(T) + \boldsymbol{u} \in V/\mathrm{Ker}(T).$$

Note that every coset A of $\mathrm{Ker}(T)$ can be represented as $\mathrm{Ker}(T) + \boldsymbol{u}$ by many different choices of $\boldsymbol{u}$. So we need to make sure that our definition of $S(A)$ will always give us the same answer in despite of the choices of $\boldsymbol{u}$: Suppose $\boldsymbol{u}, \boldsymbol{v} \in V$ such that $\mathrm{Ker}(T) + \boldsymbol{u} = \mathrm{Ker}(T) + \boldsymbol{v}$. By Theorem 8.7.3.1, $\boldsymbol{u} - \boldsymbol{v} \in \mathrm{Ker}(T)$ and hence $T(\boldsymbol{u}) - T(\boldsymbol{v}) = T(\boldsymbol{u} - \boldsymbol{v}) = \boldsymbol{0}$, i.e. $T(\boldsymbol{u}) = T(\boldsymbol{v})$. Thus

$$S(\mathrm{Ker}(T) + \boldsymbol{u}) = T(\boldsymbol{u}) = T(\boldsymbol{v}) = S(\mathrm{Ker}(T) + \boldsymbol{v}).$$

We have shown that the mapping S is well-defined.

It is easy to check that S is an isomorphism. (See Question 9.48.) Thus $V/\mathrm{Ker}(T) \cong \mathrm{R}(T)$.

Remark 9.6.16 Let $\boldsymbol{A}$ be an $m \times n$ matrix over a field $\mathbb{F}$.

Consider the linear transformation $L_{\boldsymbol{A}} : \mathbb{F}^n \to \mathbb{F}^m$ defined by $L_{\boldsymbol{A}}(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{F}^n$. Then $\mathrm{Ker}(L_{\boldsymbol{A}})$ is the nullspace of $\boldsymbol{A}$ and $\mathrm{R}(L_{\boldsymbol{A}})$ is the column space of $\boldsymbol{A}$. (See Section 7.2 or Lemma 9.5.6.) So Theorem 9.6.15 gives us an algebraic relation between the nullspace of $\boldsymbol{A}$ and the column space of $\boldsymbol{A}$.

By Example 8.7.2.4, for each element $\mathrm{Ker}(L_{\boldsymbol{A}}) + \boldsymbol{v} \in \mathbb{F}^n/\mathrm{Ker}(L_{\boldsymbol{A}})$,

$$\mathrm{Ker}(L_{\boldsymbol{A}}) + \boldsymbol{v} = \{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \in \mathrm{Ker}(L_{\boldsymbol{A}})\}$$

is the solution set of the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{v}$ is an element of the column space of $\boldsymbol{A}$. (See Theorem 4.3.6.)

Exercise 9

Question 9.1 to Question 9.7 are exercises for Section 9.1.

1. For each of the following functions T, determine whether it is a linear transformation.

(a) $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T((x, y)) = (x,\, xy)$ for $(x, y) \in \mathbb{R}^2$.

(b) $T : \mathbb{R}^m \to \mathbb{R}^n$ such that $T(\boldsymbol{u}) = \boldsymbol{u}\boldsymbol{A}$ for $\boldsymbol{u} \in \mathbb{R}^m$, where $\boldsymbol{A}$ is an $m \times n$ real matrix and vectors $\boldsymbol{u} \in \mathbb{R}^m$ are written as row vectors.

(c) $T : \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathbb{C}$ such that $T(\boldsymbol{A}) = \det(\boldsymbol{A})$ for $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$.

(d) $T : \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathbb{C}$ such that $T(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A})$ for $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$.

(e) $T : \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathcal{M}_{n\times n}(\mathbb{C})$ such that $T(\boldsymbol{A}) = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ for $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$, where $\boldsymbol{P}$ is an $n \times n$ invertible real matrix.

(f) $T : \mathcal{P}_1(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ such that $T(a + bx) = (a + 1) + (a + b)x$ for $a + bx \in \mathcal{P}_1(\mathbb{R})$.

(g) $T : [0, 1] \to \mathbb{R}$ such that $T(x) = 2x$ for $x \in [0, 1]$.

(h) $T : C^\infty([0, 1]) \to C^\infty([0, 1])$ such that for $f \in C^\infty([0, 1])$, $T(f)$ is the function defined by $T(f)(x) = f(x) + x$ for $x \in [0, 1]$.

(i) $T : C^\infty([a, b]) \to C^\infty([a, b])$, where $a < b$, such that for $f \in C^\infty([a, b])$, $T(f)$ is the function defined by

$$T(f)(x) = \frac{d^2 f(x)}{dx^2} - 3\frac{d\,f(x)}{dx} + 2f(x) \quad \text{for } x \in [a, b].$$

(j) $T : V \to V/W$ such that $T(\boldsymbol{u}) = W + \boldsymbol{u}$ for $\boldsymbol{u} \in V$, where V is a vector space and W is a subspace of V.

2. For each of the following linear transformation T, (i) determine whether the given conditions is sufficient for us to find a formula for T; (ii) if possible, write down a formula for T; and if not, give two different examples of T that satisfies the given conditions.

(a) $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ such that

$$T(1 + x + x^2) = 2 + 3x, \quad T(2 + x + 3x^2) = -1, \quad T(-1 + x + 2x^2) = x.$$

(b) $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathbb{C}$ such that

$$T\left(\begin{pmatrix} \mathrm{i} & \mathrm{i} \\ 0 & 0 \end{pmatrix}\right) = \mathrm{i}, \quad T\left(\begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}\right) = 0, \quad T\left(\begin{pmatrix} 0 & 0 \\ \mathrm{i} & \mathrm{i} \end{pmatrix}\right) = \mathrm{i}.$$

3. Let $T : V \to W$ be a linear transformation. For $X \subseteq V$, define $T[X] = \{T(\boldsymbol{u}) \mid \boldsymbol{u} \in X\}$ which is called the *image* of X under T.

(a) Show that if X is a subspace of V, then $T[X]$ is a subspace of W.

(b) If $T[X]$ is a subspace of W, is it true that X must be a subspace of V?

(c) Let $V = W = \mathbb{R}^3$, $T((x, y, z)) = (x - y,\, y - z,\, z - x)$ for $(x, y, z) \in \mathbb{R}^3$ and $X = \{(x, x, z) \mid x, z \in \mathbb{R}\}$. Write down $T[X]$ explicitly and find its dimension.

4. (a) Give an example of a mapping that satisfies (T1) but not (T2).

   (b) Give an example of a mapping that satisfies (T2) but not (T1).

5. Let V and W be two vector spaces over a field $\mathbb{F}$. Suppose $T : V \to W$ is a mapping that satisfies (T1).

   (a) Prove that $T(\mathbf{0}) = \mathbf{0}$ and $T(-\boldsymbol{u}) = -T(\boldsymbol{u})$ for all $\boldsymbol{u} \in V$.

   (b) Suppose $\mathbb{F} = \mathbb{Q}$. Prove that T is a linear transformation.

6. Let W_1 and W_2 be subspaces of a vector space V such that $V = W_1 \oplus W_2$. Let $P : V \to V$ be a mapping such that $P(\boldsymbol{u}) \in W_1$ and $\boldsymbol{u} - P(\boldsymbol{u}) \in W_2$ for all $\boldsymbol{u} \in V$.

   (a) For $\boldsymbol{v} \in W_1$ and $\boldsymbol{w} \in W_2$, show that $P(\boldsymbol{v}) = \boldsymbol{v}$ and $P(\boldsymbol{w}) = \mathbf{0}$. (Hint: Use the property that every vector in V can only be expressed uniquely as a sum of vectors from W_1 and W_2.)

   (b) Prove that the mapping P is a linear operator.

   (Parts (a) and (b) imply that for all $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w} \in V$ with $\boldsymbol{v} \in W_1$ and $\boldsymbol{w} \in W_2$, $P(\boldsymbol{u}) = \boldsymbol{v}$. The linear operator P is sometimes called the *projection* on W_1 along W_2. You can compare it with the orthogonal projections in Definition 5.2.13 and Definition 12.4.8.)

   (c) Let $V = \mathcal{M}_{2\times 2}(\mathbb{R})$, $W_1 = \left\{ \begin{pmatrix} a & a \\ b & b \end{pmatrix} \middle|\, a, b \in \mathbb{R} \right\}$ and $W_2 = \left\{ \begin{pmatrix} a & b \\ -a & b \end{pmatrix} \middle|\, a, b \in \mathbb{R} \right\}$.

      (i) Prove that $V = W_1 \oplus W_2$.

      (ii) Write down a formula for P.

7. Let $T : V \to W$ be a linear transformation and let $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}$ be vectors in V.

   (a) Suppose $T(\boldsymbol{v_1})$, $T(\boldsymbol{v_2})$, $\dots$, $T(\boldsymbol{v_n})$ are linearly independent. Prove that $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}$ are linearly independent.

   (b) Suppose $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}$ are linearly independent. Are $T(\boldsymbol{v_1})$, $T(\boldsymbol{v_2})$, $\dots$, $T(\boldsymbol{v_n})$ linearly independent?

Question 9.8 to Question 9.13 are exercises for Section 9.2.

8. For each of the following linear operator $T : V \to V$ and bases B, C for V, write down the matrix $[T]_{C,B}$.

   (a) $V = \mathbb{R}^3$, $T((x, y, z)) = (x,\ x + y + z,\ x + 2y + 3z)$ for $(x, y, z) \in \mathbb{R}^3$,
   $B = \{(1, 0, 0),\ (1, 1, 0),\ (1, 1, 1)\}$ and $C = \{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)\}$.

(b) Use the same V and T as in Part (a) but $B = \{(1,0,0),\ (0,1,0),\ (0,0,1)\}$ and $C = \{(1,0,0),\ (1,1,0),\ (1,1,1)\}$.

(c) Let $V = \mathcal{P}_n(\mathbb{R})$, $T(p(x)) = \dfrac{dp(x)}{dx}$ for $p(x) \in \mathcal{P}_n(\mathbb{R})$ and $B = C = \{1, x, x^2, \dots, x^n\}$.

(d) Let $V = \{(a_n)_{n\in\mathbb{N}} \mid a_{n+2} = a_n + a_{n+1} \text{ for } n = 1,2,3,\dots\}$, $T((a_n)_{n\in\mathbb{N}}) = (a_{n+1})_{n\in\mathbb{N}}$ for $(a_n)_{n\in\mathbb{N}} \in V$ and $B = C = \{(b_n)_{n\in\mathbb{N}},\ T((b_n)_{n\in\mathbb{N}})\}$ where $(b_n)_{n\in\mathbb{N}}$ is a sequence in V such that $(b_n)_{n\in\mathbb{N}}$ and $T((b_n)_{n\in\mathbb{N}})$ are linearly independent.

9. Let T be a linear operation on $\mathcal{P}_2(\mathbb{R})$ such that

$$[T]_{C,B} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

where $B = \{1,\ 1+x,\ 1+x+x^2\}$ and $C = \{1-x,\ x-x^2,\ x^2\}$. Find a formula of T.

10. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix}$ be a real matrix. Suppose $\boldsymbol{C_1}$, $\boldsymbol{C_2}$ and $\boldsymbol{C_3}$ are matrices obtained from $\boldsymbol{A}$ by the following elementary row operations:

$$\boldsymbol{A} \xrightarrow{2R_3} \boldsymbol{C_1}, \qquad \boldsymbol{A} \xrightarrow{R_1 \leftrightarrow R_2} \boldsymbol{C_2}, \qquad \boldsymbol{A} \xrightarrow{R_3 + 2R_2} \boldsymbol{C_3}.$$

(Follow Notation 1.4.8.)

(a) Write down $\boldsymbol{C_1}$, $\boldsymbol{C_2}$ and $\boldsymbol{C_3}$ and find elementary matrices $\boldsymbol{E_1}$, $\boldsymbol{E_2}$ and $\boldsymbol{E_3}$ such that $\boldsymbol{E_i}\boldsymbol{A} = \boldsymbol{C_i}$ for $i = 1,2,3$.

(b) Suppose V, W are real vector spaces and $T : V \to W$ is a linear transformation such that $[T]_{C,B} = \boldsymbol{A}$ where B is an ordered basis for V and $C = \{\boldsymbol{w_1},\ \boldsymbol{w_2},\ \boldsymbol{w_3}\}$ is an ordered basis for W. For each $i = 1,2,3$, find an ordered basis D_i for W such that $[T]_{D_i,B} = \boldsymbol{C_i}$.

(This question shows that elementary row operations done to a matrix is equivalent to changing bases for the corresponding linear transformation.)

11. Let $T : \mathcal{P}_3(\mathbb{R}) \to \mathbb{R}^3$ be the linear transformation defined by

$$T(a + bx + cx^2 + dx^3) = (a - b - c,\ \ 2a - b + d,\ \ -b - 2c)$$

for all $a + bx + cx^2 + dx^3 \in \mathcal{P}_3(\mathbb{R})$.

(a) Let $B = \{1,\ x,\ x^2,\ x^3\}$ and $C = \{(1,0,0),\ (0,1,0),\ (0,0,1)\}$. Write down the matrix $[T]_{C,B}$.

(b) Use Gauss-Jordan Elimination to reduce $[T]_{C,B}$ to its reduce row-echelon form $\boldsymbol{R}$.

(c) Write down an ordered basis D for $\mathbb{R}^3$ such that $[T]_{D,B} = \boldsymbol{R}$.

12. Let V be a real vector space with an ordered basis $B = \{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \boldsymbol{v_3}\}$. Suppose T is an operator on V such that

$$[T]_B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Prove that $C = \{\boldsymbol{v_1} - \boldsymbol{v_3},\ \boldsymbol{v_1},\ \boldsymbol{v_1} + \boldsymbol{v_2} + \boldsymbol{v_3}\}$ is a basis for V.

(b) Using C as an ordered basis, compute $[T]_C$.

13. Let $B = \{1 + x + x^2,\ x + x^2,\ x - x^2\}$.

(a) Prove that B is a basis for $\mathcal{P}_2(\mathbb{R})$.

(b) Find the transition matrix from E to B where $E = \{1,\ x,\ x^2\}$ is the standard basis for $\mathcal{P}_2(\mathbb{R})$.

Question 9.14 to Question 9.19 are exercises for Section 9.3.

14. Let $S : \mathcal{M}_{2\times2}(\mathbb{C}) \to \mathbb{C}^3$ and $T : \mathbb{C}^3 \to \mathcal{M}_{2\times2}(\mathbb{C})$ be linear transformations such that

$$S\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = (a + \mathrm{i}b,\ c + \mathrm{i}a,\ c + \mathrm{i}d) \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times2}(\mathbb{C})$$

and

$$T((x, y, z)) = \begin{pmatrix} x - \mathrm{i}y & y \\ -\mathrm{i}x & x - \mathrm{i}z \end{pmatrix} \quad \text{for } (x, y, z) \in \mathbb{C}^3.$$

(a) Let $B = \{\boldsymbol{E_{11}}, \boldsymbol{E_{12}}, \boldsymbol{E_{21}},\ \boldsymbol{E_{22}}\}$ and $C = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ be standard bases for $\mathcal{M}_{2\times2}(\mathbb{C})$ and $\mathbb{C}^3$ respectively. Compute $[S]_{C,B}$, $[T]_{B,C}$, $[S \circ T]_B$ and $[T \circ S]_C$.

(b) Write down a formula for each of $S \circ T$ and $T \circ S$.

15. Let $B = \{1, x, x^2\}$ and $C = \{(1, 0),\ (0, 1)\}$. Suppose $T_1 : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^2$ and $T_2 : \mathbb{R}^2 \to \mathcal{P}_2(\mathbb{R})$ are linear transformations such that

$$T_1(x) = (-1, 1), \quad T_2((1, 0)) = 1 - x^2 \quad \text{and} \quad [T_2 \circ T_1]_B = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & -2 \end{pmatrix}.$$

(a) Find the matrices $[T_1]_{C,B}$ and $[T_2]_{B,C}$.

(b) Write down the formulae for T_1 and T_2.

16. Prove Theorem 9.2.6.

    Let V be a finite dimensional vector space and B, C two bases for V. Prove that the matrix $[I_V]_{C,B}$ is invertible and $\left([I_V]_{C,B}\right)^{-1} = [I_V]_{B,C}$.

17. Let T be a linear operator on $\mathcal{P}_2(\mathbb{C})$ such that

$$[T]_B = \begin{pmatrix} 1 & \mathrm{i} & 0 \\ 0 & 1 & 1 \\ \mathrm{i} & 0 & 1 \end{pmatrix}$$

    where $B = \{1, x, x^2\}$.

    (a) Let $C = \{1, 1 + \mathrm{i}x, 1 + \mathrm{i}x^2\}$. Compute $[T]_C$.

    (b) Find a matrix $\boldsymbol{P}$ so that $\boldsymbol{P}^{-1}[T]_B\boldsymbol{P} = [T]_C$.

18. (a) Let P be a plane in $\mathbb{R}^3$ that contains the origin. Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ to be the reflection about P. (See Section 7.3.) Let $W_1 = \{\boldsymbol{u} \in \mathbb{R}^3 \mid F(\boldsymbol{u}) = \boldsymbol{u}\}$ and $W_2 = \{\boldsymbol{u} \in \mathbb{R}^3 \mid F(\boldsymbol{u}) = -\boldsymbol{u}\}$.

    (i) What are W_1 and W_2 geometrically?

    (ii) Show that $\mathbb{R}^3 = W_1 \oplus W_2$.

    (b) Let $\mathbb{F}$ be a field with $1 + 1 \neq 0$.

    (i) Suppose T is an linear operator on a vector space V over $\mathbb{F}$ such that $T^2 = I_V$. Let $W_1 = \{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) = \boldsymbol{u}\}$ and $W_2 = \{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) = -\boldsymbol{u}\}$. Prove that $V = W_1 \oplus W_2$.

    (ii) If $\boldsymbol{A}$ is a square matrix over $\mathbb{F}$ such that $\boldsymbol{A}^2 = \boldsymbol{I}$, show that there exists an invertible matrix $\boldsymbol{P}$ over $\mathbb{F}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

19. Let V be a finite dimensional vector space such that $\dim(V) = n \geq 1$ and let $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v}_n\}$ be an ordered basis for V. Define a linear operator S on V such that

$$S(\boldsymbol{v_k}) = \begin{cases} \boldsymbol{0} & \text{if } k = 1 \\ \boldsymbol{v_{k-1}} & \text{if } 2 \leq k \leq n. \end{cases}$$

    (a) Write down $[S]_B$.

    (b) Prove that $S^{n-1} \neq O_V$ and $S^n = O_V$.

Question 9.21 to Question 9.25 are exercises for Section 9.4.

20. Prove Proposition 9.4.3:

    Let V and W be finite dimensional vector space over the same field $\mathbb{F}$, where $\dim(V) \geq 1$ and $\dim(W) \geq 1$, and let B and C be ordered basis for V and W respectively.

(a) If T_1, $T_2 : V \to W$ are linear transformations, prove that $[T_1 + T_2]_{C,B} = [T_1]_{C,B} + [T_2]_{C,B}$.

(b) If $T : V \to W$ be a linear transformation and $c \in \mathbb{F}$, prove that $[cT]_{C,B} = c[T]_{C,B}$.

21. Let S and T be linear operators on $\mathcal{M}_{2\times 2}(\mathbb{R})$ such that

$$S(\boldsymbol{X}) = \boldsymbol{AX} \quad \text{and} \quad T(\boldsymbol{X}) = \boldsymbol{XA} \quad \text{for } \boldsymbol{X} \in \mathcal{M}_{2\times 2}(\mathbb{R})$$

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

(a) Let $E = \{\boldsymbol{E_{11}}, \boldsymbol{E_{12}}, \boldsymbol{E_{21}}, \boldsymbol{E_{22}}\}$ be the standard bases for $\mathcal{M}_{2\times 2}(\mathbb{R})$. Compute $[S]_E$, $[T]_E$, $[I_V]_E$, $[S + T]_E$, $[T - 2I_V]_E$ and $[(T - 2I_V)^2]_E$.

(b) Write down a formula for each of $S + T$, $T - 2I_V$ and $(T - 2I_V)^2$.

22. Let V be a finite dimensional vector space such that $\dim(V) = n \geq 1$ and let T be an linear operator on V. Define $Q = T - \lambda I_V$ where λ is a scalar. Suppose there exists a positive integer n such that $Q^n = O_V$ and $Q^{n-1}(\boldsymbol{v}) \neq \boldsymbol{0}$ for some $\boldsymbol{v} \in V$.

(a) Show that $C = \{Q^{n-1}(\boldsymbol{v}), \ldots, Q(\boldsymbol{v}), \boldsymbol{v}\}$ is a basis for V.

(b) Using C in Part (a) as an ordered basis, compute $[Q]_C$ and $[T]_C$. (The matrix $[T]_C$ is called a *Jordan block*. See Section 11.6.)

23. Let $V = \mathbb{R}^2$ and $W = \mathcal{P}_2(\mathbb{R})$. Suppose U is the subspace of $\mathcal{L}(V, W)$ spanned by T_1, T_2, T_3 and T_4 where

$$\begin{aligned} T_1((a,b)) &= (a+b) + ax^2, & T_2((a,b)) &= (a+b) + (a+b)x - bx^2, \\ T_3((a,b)) &= (a+b)(x - x^2), & T_4((a,b)) &= 2(a+b) + (a+b)x + (a-b)x^2 \end{aligned}$$

for $(a, b) \in V$. Find a basis for U and determine the dimension of U.

24. Let $T : V \to W$ be a linear transformation. Define $T^* : W^* \to V^*$ such that for $f \in W^*$, $T^*(f)$ is a functional on V defined by

$$T^*(f)(\boldsymbol{u}) = f(T(\boldsymbol{u})) \quad \text{for } \boldsymbol{u} \in V.$$

(See Definition 9.4.6 for the definition of the dual spaces V^* and W^*.)

(a) Show that T^* is a linear transformation.

(b) Suppose V and W are finite dimensional with bases $B = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ and $C = \{\boldsymbol{w_1}, \dots, \boldsymbol{w_m}\}$ respectively. For $i = 1, \dots, n$, define $g_i \in V^*$ such that

$$g_i(\boldsymbol{v_j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise;} \end{cases}$$

and for $i = 1, \dots, m$, define $h_i \in W^*$ such that

$$h_i(\boldsymbol{w_j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $B^* = \{g_1, \dots, g_n\}$ is a basis for V^* and $C^* = \{h_1, \dots, h_m\}$ is a basis for W^*.

(ii) Prove that $[T^*]_{B^*,C^*} = ([T]_{C,B})^{\mathrm{T}}$.

25. Let V and W be vector spaces over the same field. For any subset A of V, define

$$A^0 = \{T \in \mathcal{L}(V, W) \mid T(\boldsymbol{u}) = \boldsymbol{0} \text{ for all } \boldsymbol{u} \in A\}.$$

(a) If A is a subset of V, prove that A^0 is a subspace of $\mathcal{L}(V, W)$.

(b) If A and B are subsets of V such that $A \subseteq B$, prove that $B^0 \subseteq A^0$.

(c) If U_1 and U_2 are subspaces of V, prove that $(U_1 + U_2)^0 = U_1^0 \cap U_2^0$.

(d) If U_1 and U_2 are finite dimensional subspaces of V, prove that $U_1^0 + U_2^0 = (U_1 \cap U_2)^0$.

Question 9.26 to Question 9.40 are exercises for Section 9.5.

26. For each of the following linear transformation T, (i) find $\mathrm{Ker}(T)$ and $\mathrm{R}(T)$; and (ii) if possible, find $\mathrm{nullity}(T)$ and $\mathrm{rank}(T)$.

(a) $T : \mathbb{F}_2^3 \to \mathbb{F}_2^3$ such that $T((x, y, z)) = (x + y,\ y + z,\ x + z)$ for $(x, y, z) \in \mathbb{F}_2^3$.

(b) $T : \mathcal{P}(\mathbb{R}) \to \mathbb{R}^2$ such that $T(p(x)) = (p(0),\ p(1))$ for $p(x) \in \mathcal{P}(\mathbb{R})$.

(c) $T : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathcal{M}_{n\times n}(\mathbb{R})$ such that $T(\boldsymbol{A}) = \boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}$ for $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R})$.

(d) $T : \mathbb{F}^{\mathbb{N}} \to \mathbb{F}^{\mathbb{N}}$ such that $T((a_n)_{n\in\mathbb{N}}) = (a_{n+1} - a_n)_{n\in\mathbb{N}}$ for $(a_n)_{n\in\mathbb{N}} \in \mathbb{F}^{\mathbb{N}}$.

(e) $T : V \to V/W$ such that $T(\boldsymbol{u}) = W + \boldsymbol{u}$ for $\boldsymbol{u} \in V$, where V is a vector space and W is a subspace of V.

27. Let $T : \mathbb{R}^3 \to \mathcal{P}_1(\mathbb{R})$ be a linear transformation such that

$$[T]_{C,B} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

where $B = \{(1, 0, 0),\ (0, 1, 1),\ (0, 1, -1)\}$ and $C = \{1,\ 1 + x\}$.

(a) Write down a formula for T.

(b) Find $\text{Ker}(T)$ and $\text{R}(T)$.

28. Prove Theorem 9.5.2:

Let $T : V \to W$ be a linear transformation. Show that

(a) $\text{Ker}(T)$ is a subspace of V; and

(b) $\text{R}(T)$ is a subspace of W.

29. Let T be a linear operator on a vector space V such that $T^2 = T$.

(a) Prove that $V = \text{R}(T) \oplus \text{Ker}(T)$.

(b) Prove that T is the projection on $\text{R}(T)$ along $\text{Ker}(T)$. (See Question 9.6 for the definition of "projection".)

30. Prove Lemma 9.5.6:

Let $T : V \to W$ be a linear transformation where V and W are finite dimensional with $\dim(V) \geq 1$ and $\dim(W) \geq 1$. Prove that for any ordered bases B and C for V and W respectively,

(a) $\{ [\boldsymbol{u}]_B \mid \boldsymbol{u} \in \text{Ker}(T)\}$ is the nullspace of $[T]_{C,B}$ and $\text{nullity}(T) = \text{nullity}([T]_{C,B})$; and

(b) $\{ [\boldsymbol{v}]_C \mid \boldsymbol{v} \in \text{R}(T)\}$ is the column space of $[T]_{C,B}$ and $\text{rank}(T) = \text{rank}([T]_{C,B})$.

31. Prove Theorem 9.5.9:

Let $T : V \to W$ be a linear transformation. Suppose B and C are subsets of V such that B is a basis for $\text{Ker}(T)$, $\{T(\boldsymbol{v}) \mid \boldsymbol{v} \in C\}$ is a basis for $\text{R}(T)$ and for any $\boldsymbol{v}, \boldsymbol{v'} \in C$ with $\boldsymbol{v} \neq \boldsymbol{v'}$, $T(\boldsymbol{v}) \neq T(\boldsymbol{v'})$. Prove that $B \cup C$ is a basis for V.

32. Let $T : V \to W$ be a linear transformation. For $\boldsymbol{w} \in W$, we define the *pre-image* of $\boldsymbol{w}$ under T to be the set

$$T^{-1}[\boldsymbol{w}] = \{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) = \boldsymbol{w}\}.$$

(a) Show that $T^{-1}[\boldsymbol{w}] = \text{Ker}(T) + \boldsymbol{v}$ for some $\boldsymbol{v} \in V$.

Let U be a subspace of W. Define the *pre-image* of U under T to be the set

$$T^{-1}[U] = \{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) \in U\}.$$

(b) Show that $T^{-1}[U]$ is a subspace of V.

(c) If both $\text{Ker}(T)$ and U are finite dimensional, say $\dim(\text{Ker}(T)) = k$ and $\dim(U) = m$, find $\dim(T^{-1}[U])$.

33. (a) Let $T : V \to W$ be a linear transformation where V and W are finite dimensional vector spaces such that $\dim(V) = n \geq 1$ and $\dim(W) = m \geq 1$. Show that there exist ordered bases B and C for V and W, respectively, such that

$$[T]_{C,B} = \begin{pmatrix} \boldsymbol{I_k} & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{0}_{(m-k)\times k} & \boldsymbol{0}_{(m-k)\times(n-k)} \end{pmatrix}$$

where $k = \operatorname{rank}(T)$. (Hint: The result of Theorem 9.5.9 can be useful.)

(b) Let $\boldsymbol{A} \in \mathcal{M}_{m\times n}(\mathbb{F})$. Show that there exists invertible matrices $\boldsymbol{P} \in \mathcal{M}_{m\times m}(\mathbb{F})$ and $\boldsymbol{Q} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I_k} & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{0}_{(m-k)\times k} & \boldsymbol{0}_{(m-k)\times(n-k)} \end{pmatrix}$$

where $k = \operatorname{rank}(\boldsymbol{A})$.

(c) Let $\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix}$ be a real matrix. Find two 3×3 invertible real matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that $\boldsymbol{PAQ} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

34. (a) Let $S : U \to V$ and $T : V \to W$ be linear transformations. Prove that $\operatorname{Ker}(S) \subseteq \operatorname{Ker}(T \circ S)$ and $\mathrm{R}(T) \supseteq \mathrm{R}(T \circ S)$.

(b) Let $S : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ and $T : \mathbb{R}^3 \to \mathcal{M}_{2\times 2}(\mathbb{R})$ be linear transformations defined by

$$S(a + bx + cx^2) = (a - b,\ b - c,\ c - a) \quad \text{for } a + bx + cx^2 \in \mathcal{P}_2(\mathbb{R})$$
$$T((x, y, z)) = \begin{pmatrix} x & y + z \\ y + z & x \end{pmatrix} \quad \text{for } (x, y, z) \in \mathbb{R}^3.$$

Find $\operatorname{Ker}(S)$, $\operatorname{Ker}(T \circ S)$, $\mathrm{R}(T)$, $\mathrm{R}(T \circ S)$ and hence verify the results in Part (a).

35. (a) Let $S, T : V \to W$ be linear transformations. Prove that $\mathrm{R}(S + T) \subseteq \mathrm{R}(S) + \mathrm{R}(T)$ and $\operatorname{Ker}(S + T) \supseteq \operatorname{Ker}(S) \cap \operatorname{Ker}(T)$.

(b) Let $S, T : \mathbb{R}^3 \to \mathbb{R}^4$ be linear transformations defined by

$$S((x, y, z)) = (x,\ x,\ -y,\ -y) \quad \text{for } (x, y, z) \in \mathbb{R}^3$$
$$T((x, y, z)) = (x,\ x + z,\ y,\ y + z) \quad \text{for } (x, y, z) \in \mathbb{R}^3.$$

Find $\mathrm{R}(S + T)$, $\mathrm{R}(S) + \mathrm{R}(T)$, $\operatorname{Ker}(S + T)$ and $\operatorname{Ker}(S) \cap \operatorname{Ker}(T)$ and hence verify the results in Part (a).

36. Let T_1, T_2, ..., T_k be linear operators on a vector space V. Suppose

   (i) $I_V = T_1 + T_2 + \cdots + T_k$;

   (ii) $T_i \circ T_j = O_V$ whenever $i \neq j$; and

   (iii) ${T_i}^2 = T_i$ for $i = 1, 2, \ldots, k$.

   Prove that $V = \mathrm{R}(T_1) \oplus \mathrm{R}(T_2) \oplus \cdots \oplus \mathrm{R}(T_k)$.

37. Let V be a finite dimensional vector space.

   (a) Prove that if T is a linear operator on V such that $\mathrm{Ker}(T) \cap \mathrm{R}(T) = \{\mathbf{0}\}$, then $V = \mathrm{Ker}(T) \oplus \mathrm{R}(T)$.

   (b) Give an example of a linear operator T on $\mathbb{R}^2$ such that $V \neq \mathrm{Ker}(T) + \mathrm{R}(T)$.

   (c) Prove that for any linear operator T on V, $\mathrm{Ker}(T^i) \subseteq \mathrm{Ker}(T^{i+1})$ for $i = 1, 2, 3, \ldots$.

   (d) Show that for any linear operator T on V, there exists a positive integer m such that $V = \mathrm{Ker}(T^m) \oplus \mathrm{R}(T^m)$.

38. Let $S : W \to V$ and $T : V \to W$ be linear transformations such that $T \circ S = I_W$. Prove that $V = \mathrm{Ker}(T) \oplus \mathrm{R}(S)$.

39. Let V and W be finite dimensional vector spaces such that $\dim(V) = \dim(W)$ and let $T : V \to W$ be a linear transformation. Prove that the following statements are equivalent:

   (i) T is injective.

   (ii) T is surjective.

   (iii) T is bijective.

40. Let V and W be finite dimensional vector spaces and let $T : V \to W$ be a linear transformation.

   (a) Prove that if $\dim(V) < \dim(W)$, then T is not surjective.

   (b) Prove that if $\dim(V) > \dim(W)$, then T is not injective.

Question 9.41 to Question 9.49 are exercises for Section 9.6.

41. For each of the following, determine whether the linear transformation T is an isomorphism. If so, find the inverse of T.

   (a) $T : \mathbb{F}_2^3 \to \mathbb{F}_2^3$ such that $T((x, y, z)) = (x + y,\ y + z,\ z + x)$ for $(x, y, z) \in \mathbb{F}_2^3$.

   (b) $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that $T((x, y, z)) = (x + y,\ y + z,\ z + x)$ for $(x, y, z) \in \mathbb{R}^3$.

(c) $T : \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathcal{M}_{n\times n}(\mathbb{C})$ such that $T(\boldsymbol{A}) = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ for $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$ where $\boldsymbol{P}$ is an $n \times n$ invertible complex matrix.

(d) $T : \mathcal{P}_n(\mathbb{R}) \to \mathbb{R}^n$ such that $T(p(x)) = (p(1), p(2), \ldots, p(n))$ for $p(x) \in \mathcal{P}_n(\mathbb{R})$.

(e) $T : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ such that $T(p(x)) = \dfrac{d(xp(x))}{dx}$ for $p(x) \in \mathcal{P}(\mathbb{R})$.

42. Prove Theorem 9.6.6:

Let $T : V \to W$ be a linear transformation where V and W are finite dimensional with $\dim(V) = \dim(W) \geq 1$. Let B and C be ordered bases for V and W respectively.

(a) Prove that T is an isomorphism if and only if $[T]_{C,B}$ is an invertible matrix.

(b) If T is an isomorphism, show that $[T^{-1}]_{B,C} = (\,[T]_{C,B})^{-1}$.

43. Let $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathcal{M}_{2\times 2}(\mathbb{C})$ be a linear transformation such that T^{-1} exists and

$$T^{-1}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} a & a + \mathrm{i}b \\ a + \mathrm{i}c & b + c + \mathrm{i}d \end{pmatrix} \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{C}).$$

(a) Write down a formula for T.

(b) Find $\mathrm{Ker}(T)$ and $\mathrm{R}(T)$.

44. Let $\mathbb{F}$ be a field and $c_0, c_1, \ldots, c_n \in \mathbb{F}$ such that $c_i \neq c_j$ for $i \neq j$. For $i = 0, 1, \ldots, n$, define

$$q_i(x) = d_i^{-1}(x - c_0)\cdots(x - c_{i-1})(x - c_{i+1})\cdots(x - c_n) \in \mathcal{P}_n(\mathbb{F})$$

where $d_i = (c_i - c_0)\cdots(c_i - c_{i-1})(c_i - c_{i+1})\cdots(c_i - c_n) \in \mathbb{F}$. (The polynomials $q_i(x)$'s are called the *Lagrange polynomials*.)

(a) Show that $\{q_0(x), q_1(x), \ldots, q_n(x)\}$ is a basis for $\mathcal{P}_n(\mathbb{F})$.

(b) Hence, or otherwise, prove that the linear transformation $T : \mathcal{P}_n(\mathbb{F}) \to \mathbb{F}^{n+1}$ defined by $T(p(x)) = (p(c_0), p(c_1), \ldots, p(c_n))$, for $p(x) \in \mathcal{P}_n(\mathbb{F})$, is an isomorphism.

45. (a) Let $U = \{a_0 + a_1x^2 + a_2x^4 + \cdots + a_mx^{2m} \mid m \in \mathbb{N} \text{ and } a_0, a_1, a_2, \ldots, a_m \in \mathbb{R}\}$.

(i) Is U a proper subspace of $\mathcal{P}(\mathbb{R})$?

(ii) Show that $U \cong \mathcal{P}(\mathbb{R})$.

(b) Given a finite dimensional vector space V, can there exist a proper subspace W of V such that $W \cong V$?

46. Prove Theorem 9.6.8:

Let $S : W \to V$ and $T : V \to W$ be linear transformations such that $T \circ S = I_W$, the identity operator on W.

(a) Prove that S is injective and T is surjective.

(b) If V and W are finite dimensional and $\dim(V) = \dim(W)$, prove that S and T are isomorphisms and they are inverses of each other. (See also Theorem 2.4.12.)

47. Let U, V and W be vector spaces over a field $\mathbb{F}$. Suppose $R : U \to V$ and $S, T : V \to W$ are isomorphisms.

    (a) Is $T \circ R$ an isomorphism? If $S \neq -T$, is $S + T$ an isomorphism? If $c \in \mathbb{F}$ and $c \neq 0$, is cT an isomorphism?

    (b) Prove that if $S + T$ is an isomorphism, then $S^{-1} + T^{-1}$ is also an isomorphism.

48. Complete the proof of Theorem 9.6.15:

    Let $T : V \to W$ be a linear transformation and let $S : V/\mathrm{Ker}(T) \to \mathrm{R}(T)$ be the mapping defined in the proof of Theorem 9.6.15, i.e.

$$S(\mathrm{Ker}(T) + \boldsymbol{u}) = T(\boldsymbol{u}) \quad \text{for } \mathrm{Ker}(T) + \boldsymbol{u} \in V/\mathrm{Ker}(T).$$

    Prove that S is an isomorphism.

49. (a) Let V be a subspace of a vector space U and let W be a subspace of V. Show that V/W is a subspace of U/W.

    (b) **(The Second Isomorphism Theorem)** Let V and W be subspaces of a vector space U. Prove that $(V + W)/W \cong V/(V \cap W)$.

    (c) **(The Third Isomorphism Theorem)** Let V be a subspace of a vector space U and let W be a subspace of V. Prove that $(U/W)/(V/W) \cong U/V$.

Chapter 10

Multilinear Forms and Determinants

Section 10.1 Permutations

Discussion 10.1.1 In Section 2.5, we define the determinants of a square matrices by using the cofactor expansion. Such a definition is usually referred as a "working" definition which is good for computation but not quite suitable for us to investigate the properties of determinants. In this chapter, we shall study two different ways of defining determinants: one by multilinear forms and another by permutations.

Definition 10.1.2 A *permutation* σ of $\{1, 2, \ldots, n\}$ is a bijective mapping from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$. We usually represent σ by $\begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$.
The set of all permutations of $\{1, 2, \ldots, n\}$ is denoted by $\mathcal{S}_n$. Note that $|\mathcal{S}_n| = n!$.

Example 10.1.3

$\mathcal{S}_1 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$, $\mathcal{S}_2 = \left\{ \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\}$ and

$\mathcal{S}_3 = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \right\}$.

For example, $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$ is the mapping from $\{1, 2, 3\}$ to $\{1, 2, 3\}$ such that $\sigma(1) = 3$, $\sigma(2) = 1$ and $\sigma(3) = 2$.

Notation 10.1.4

1. For $\sigma, \tau \in \mathcal{S}_n$, $\sigma \circ \tau$ is also a permutation and we usually denote $\sigma \circ \tau$ by $\sigma\tau$.

2. In the following, for $\alpha, \beta \in \{1, 2, \ldots, n\}$, we use $\phi_{\alpha,\beta}$ to denote the permutation of $\{1, 2, \ldots, n\}$ such that
$$\phi_{\alpha,\beta}(k) = \begin{cases} k & \text{if } k \neq \alpha, \beta \\ \beta & \text{if } k = \alpha \\ \alpha & \text{if } k = \beta. \end{cases}$$
This permutation is called the *transposition* of α and β. Note that $\phi_{\alpha,\beta} = \phi_{\beta,\alpha}$ and $\phi_{\alpha,\beta}^{-1} = \phi_{\alpha,\beta}$.

Example 10.1.5

1. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}$. Then
$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}.$$

2. In $\mathcal{S}_4$, $\phi_{1,2} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}$, $\phi_{2,3} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$, $\phi_{3,4} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}$,
$$\phi_{1,2}\,\phi_{2,3}\,\phi_{1,2} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = \phi_{1,3} \quad \text{and} \quad \phi_{1,2}\,\phi_{2,3}\,\phi_{3,4}\,\phi_{2,3}\,\phi_{1,2} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix} = \phi_{1,4}.$$

Lemma 10.1.6

1. $\{\sigma^{-1} \mid \sigma \in \mathcal{S}_n\} = \mathcal{S}_n$.
2. For any $\tau \in \mathcal{S}_n$, $\{\tau\sigma \mid \sigma \in \mathcal{S}_n\} = \{\sigma\tau \mid \sigma \in \mathcal{S}_n\} = \mathcal{S}_n$.

Proof The proof is left as an exercise. See Question 10.4(a) and (b).

Lemma 10.1.7 For every $\sigma \in \mathcal{S}_n$, there exists $\alpha_1, \alpha_2, \ldots, \alpha_k \in \{1, 2, \ldots, n-1\}$ such that $\sigma = \phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1} \cdots \phi_{\alpha_k,\alpha_k+1}$.

Proof Since every transposition $\phi_{\alpha,\beta} \in \mathcal{S}_n$, $\alpha < \beta$, can be written as
$$\phi_{\alpha,\beta} = \phi_{\alpha,\alpha+1}\,\phi_{\alpha+1,\alpha+2} \cdots \phi_{\beta-2,\beta-1}\,\phi_{\beta-1,\beta}\,\phi_{\beta-2,\beta-1} \cdots \phi_{\alpha+1,\alpha+2}\,\phi_{\alpha,\alpha+1},$$
we only need to show that σ can be written as a product of transpositions.

Let $\sigma_0 = \sigma$. Suppose we have already defined σ_{m-1}, where $1 \leq m < n-1$, such that
$$\sigma_{m-1}(k) = k \quad \text{for } k = 1, 2, \ldots, m-1.$$

Define $\sigma_m = \phi_{m,\gamma_m}\,\sigma_{m-1}$ where $\gamma_m = \sigma_{m-1}(m) \geq m$. Then
$$\sigma_m(k) = \phi_{m,\gamma_m}(\sigma_{m-1}(k)) = \begin{cases} \phi_{m,\gamma_m}(k) = k & \text{if } k = 1, 2, \ldots, m-1 \\ \phi_{m,\gamma_m}(\gamma_m) = m & \text{if } k = m. \end{cases}$$

Inductively, we obtain transpositions ϕ_{1,γ_1}, ϕ_{2,γ_2}, ..., $\phi_{n-1,\gamma_{n-1}}$ such that

$$\phi_{n-1,\gamma_{n-1}} \cdots \phi_{2,\gamma_2}\, \phi_{1,\gamma_1}\, \sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$$

which is the identity mapping. Thus

$$\sigma = (\phi_{n-1,\gamma_{n-1}} \cdots \phi_{2,\gamma_2}\, \phi_{1,\gamma_1})^{-1} = \phi_{1,\gamma_1}\, \phi_{2,\gamma_2} \cdots \phi_{n-1,\gamma_{n-1}}$$

is a product of transpositions.

Example 10.1.8 Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$. Following the procedure of the proof of Lemma 10.1.7, we have

$$\phi_{34}\, \phi_{23}\, \phi_{13}\, \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

and hence $\sigma = \phi_{13}\, \phi_{23}\, \phi_{34}$. Since $\phi_{13} = \phi_{12}\, \phi_{23}\, \phi_{12}$, we have $\sigma = \phi_{12}\, \phi_{23}\, \phi_{12}\, \phi_{23}\, \phi_{34}$.

The decomposition of a permutation discussed in Lemma 10.1.7 is not unique, for this example, we can also write $\sigma = \phi_{23}\, \phi_{34}\, \phi_{12}$.

Definition 10.1.9 Let $\sigma \in \mathcal{S}_n$. An *inversion* is said to occur in σ if $\sigma(i) > \sigma(j)$ for $i < j$. If the total number of inversions in σ is even, σ is called an *even permutation*; otherwise, σ is called an *odd permutation*.

The *sign* (or *parity*) of σ is defined to be

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Example 10.1.10

1. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}$.

   In σ, inversions occur when $(i, j) = (1, 2)$, $(1, 4)$ and $(3, 4)$. So σ is an odd permutation and $\operatorname{sgn}(\sigma) = -1$. In τ, inversions occur when $(i, j) = (1, 2)$, $(1, 3)$, $(1, 4)$ and $(3, 4)$. So τ is an even permutation and $\operatorname{sgn}(\tau) = 1$.

2. For $\phi_{\alpha,\beta} \in \mathcal{S}_n$, where $1 \le \alpha < \beta \le n$, inversions occur when $(i, j) = (\alpha,\, \alpha+1)$, $(\alpha,\, \alpha+2)$, ..., $(\alpha,\, \beta)$, $(\alpha+1,\, \beta)$, ..., $(\beta-1,\, \beta)$. There are $(\beta-\alpha)+(\beta-\alpha-1) = 2(\beta-\alpha)-1$ inversions. Hence $\phi_{\alpha,\beta}$ is an odd permutation and $\operatorname{sgn}(\phi_{\alpha,\beta}) = -1$.

   In particular, $\operatorname{sgn}(\phi_{\alpha,\alpha+1}) = -1$ for $1 \le \alpha \le n-1$.

Theorem 10.1.11 For any $\sigma, \tau \in \mathcal{S}_n$, $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$.

Proof First we show that for all $\theta \in \mathcal{S}_n$ and $1 \le \alpha \le n-1$, $\operatorname{sgn}(\theta\,\phi_{\alpha,\alpha+1}) = -\operatorname{sgn}(\theta)$. As

$$\theta = \begin{pmatrix} 1 & \cdots & \alpha-1 & \alpha & \alpha+1 & \alpha+2 & \cdots & n \\ \theta(1) & \cdots & \theta(\alpha-1) & \theta(\alpha) & \theta(\alpha+1) & \theta(\alpha+2) & \cdots & \theta(n) \end{pmatrix}$$

and

$$\theta\,\phi_{\alpha,\alpha+1} = \begin{pmatrix} 1 & \cdots & \alpha-1 & \alpha & \alpha+1 & \alpha+2 & \cdots & n \\ \theta(1) & \cdots & \theta(\alpha-1) & \theta(\alpha+1) & \theta(\alpha) & \theta(\alpha+2) & \cdots & \theta(n) \end{pmatrix},$$

$$\begin{aligned} &\text{the number of inversions in } \theta\,\phi_{\alpha,\alpha+1} \\ &= \begin{cases} (\text{the number of inversions in } \theta) + 1 & \text{if } \theta(\alpha) < \theta(\alpha+1) \\ (\text{the number of inversions in } \theta) - 1 & \text{if } \theta(\alpha) > \theta(\alpha+1). \end{cases} \end{aligned}$$

So

$$\operatorname{sgn}(\theta\,\phi_{\alpha,\alpha+1}) = -\operatorname{sgn}(\theta). \tag{10.1}$$

By Lemma 10.1.7, we can write $\tau = \phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_k,\alpha_k+1}$ for some $\alpha_1, \alpha_2, \ldots, \alpha_k \in \{1, 2, \ldots, n-1\}$. By applying (10.1) repeatedly,

$$\begin{aligned} \operatorname{sgn}(\tau) &= \operatorname{sgn}(\phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_{k-1},\alpha_{k-1}+1}\,\phi_{\alpha_k,\alpha_k+1}) \\ &= -\operatorname{sgn}(\phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_{k-1},\alpha_{k-1}+1}) \\ &\;\;\vdots \\ &= (-1)^{k-1}\operatorname{sgn}(\phi_{\alpha_1,\alpha_1+1}) \end{aligned}$$

and by Example 10.1.10.2, $\operatorname{sgn}(\tau) = (-1)^k$. On the other hands, by applying (10.1) repeatedly,

$$\begin{aligned} \operatorname{sgn}(\sigma\tau) &= \operatorname{sgn}(\sigma\,\phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_{k-1},\alpha_{k-1}+1}\,\phi_{\alpha_k,\alpha_k+1}) \\ &= -\operatorname{sgn}(\sigma\,\phi_{\alpha_1,\alpha_1+1}\,\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_{k-1},\alpha_{k-1}+1}) \\ &\;\;\vdots \\ &= (-1)^k\operatorname{sgn}(\sigma) \\ &= \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau). \end{aligned}$$

Corollary 10.1.12

1. If $\sigma \in \mathcal{S}_n$ is a product of k transpositions, then $\operatorname{sgn}(\sigma) = (-1)^k$.
2. A permutation is even (respectively, odd) if it is a product of even (respectively, odd) number of transpositions.
3. For any $\sigma \in \mathcal{S}_n$, $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$.

Proof Part 1 is a consequence of Example 10.1.10.2 and Theorem 10.1.11 while Part 2 is a consequence of Part 1. The proof of Part 3 is left as exercise. See Question 10.4(c).

Section 10.2 Multilinear Forms

Definition 10.2.1 Let V be a vector space over a field $\mathbb{F}$ and let $V^n = \underbrace{V \times \cdots \times V}_{n \text{ times}}$. A mapping $T : V^n \to \mathbb{F}$ is called a *multilinear form* on V if for each i, $1 \le i \le n$,

$$\begin{aligned} &T(\boldsymbol{u_1}, \ldots, \boldsymbol{u_{i-1}}, a\boldsymbol{v} + b\boldsymbol{w}, \boldsymbol{u_{i+1}}, \ldots, \boldsymbol{u_n}) \\ &= aT(\boldsymbol{u_1}, \ldots, \boldsymbol{u_{i-1}}, \boldsymbol{v}, \boldsymbol{u_{i+1}}, \ldots, \boldsymbol{u_n}) + bT(\boldsymbol{u_1}, \ldots, \boldsymbol{u_{i-1}}, \boldsymbol{w}, \boldsymbol{u_{i+1}}, \ldots, \boldsymbol{u_n}) \end{aligned} \tag{10.2}$$

for all $a, b \in \mathbb{F}$ and $\boldsymbol{u_1}, \ldots, \boldsymbol{u_{i-1}}, \boldsymbol{u_{i+1}}, \ldots, \boldsymbol{u_n}, \boldsymbol{v}, \boldsymbol{w} \in V$. If $n = 2$, T is also called a *bilinear form* on V.

A multilinear form T on V is called *alternative* if $T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) = 0$ whenever $\boldsymbol{u_\alpha} = \boldsymbol{u_\beta}$ for some $\alpha \neq \beta$.

(If $1 + 1 \neq 0$ in $\mathbb{F}$, a multilinear form T on V is alternative if and only if for all transposition $\tau = \phi_{\alpha,\beta} \in \mathcal{S}_n$ and $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n} \in V$, $T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) = -T(\boldsymbol{u_{\tau(1)}}, \boldsymbol{u_{\tau(2)}}, \ldots, \boldsymbol{u_{\tau(n)}})$. See Theorem 10.2.3 and Question 10.9.)

Example 10.2.2 In the following examples, vectors in $\mathbb{F}^n$ are written as column vectors.

1. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Define $T : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ by

   $$T(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^n.$$

   Then T is a bilinear form on $\mathbb{F}^n$. (See Question 10.8.)

   In particular, if $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{F})$, then

   $$T\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) = ad - bc \quad \text{for } \begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \in \mathbb{F}^2$$

   which is an alternative bilinear form.

2. Define $P : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ by

   $$P(\boldsymbol{A}) = \sum_{\sigma \in \mathcal{S}_n} a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(n),n} \quad \text{for } \boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F}).$$

   The value $P(\boldsymbol{A})$ is known as the *permanent* of $\boldsymbol{A}$.

   Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n} \end{pmatrix}$ where $\boldsymbol{a_i}$ is the ith column of $\boldsymbol{A}$. Since each $\boldsymbol{a_i}$ is a vector in $\mathbb{F}^n$, P can be regarded as a mapping from $\mathbb{F}^n \times \cdots \times \mathbb{F}^n$ to $\mathbb{F}$. Take any $b, c \in \mathbb{F}$ and $\boldsymbol{a_1}, \ldots, \boldsymbol{a_{i-1}}, \boldsymbol{a_{i+1}}, \ldots, \boldsymbol{a_n}, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^n$ where $1 \le i \le n$. Let $\boldsymbol{a_j} = (a_{1j}, a_{2j}, \ldots, a_{nj})^{\mathrm{T}}$,

$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$. Then

$$\begin{aligned}
&P\Big(\begin{pmatrix}\boldsymbol{a_1} & \cdots & \boldsymbol{a_{i-1}} & b\boldsymbol{x}+c\boldsymbol{y} & \boldsymbol{a_{i+1}} & \cdots & \boldsymbol{a_n}\end{pmatrix}\Big)\\
&= \sum_{\sigma\in\mathcal{S}_n} a_{\sigma(1),1}\cdots a_{\sigma(i-1),i-1}\left(bx_{\sigma(i)}+cy_{\sigma(i)}\right) a_{\sigma(i+1),i+1}\cdots a_{\sigma(n),n}\\
&= b\sum_{\sigma\in\mathcal{S}_n} a_{\sigma(1),1}\cdots a_{\sigma(i-1),i-1}\, x_{\sigma(i)}\, a_{\sigma(i+1),i+1}\cdots a_{\sigma(n),n}\\
&\qquad\qquad + c\sum_{\sigma\in\mathcal{S}_n} a_{\sigma(1),1}\cdots a_{\sigma(i-1),i-1}\, y_{\sigma(i)}\, a_{\sigma(i+1),i+1}\cdots a_{\sigma(n),n}\\
&= bP\Big(\begin{pmatrix}\boldsymbol{a_1} & \cdots & \boldsymbol{a_{i-1}} & \boldsymbol{x} & \boldsymbol{a_{i+1}} & \cdots & \boldsymbol{a_n}\end{pmatrix}\Big)\\
&\qquad\qquad + cP\Big(\begin{pmatrix}\boldsymbol{a_1} & \cdots & \boldsymbol{a_{i-1}} & \boldsymbol{y} & \boldsymbol{a_{i+1}} & \cdots & \boldsymbol{a_n}\end{pmatrix}\Big).
\end{aligned}$$

So P is a multilinear form on $\mathbb{F}^n$. Note that P is not alternative if $1+1\neq 0$ in $\mathbb{F}$.

3. Define $Q : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ by

$$Q(\boldsymbol{A}) = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2}\cdots a_{\sigma(n),n} \quad \text{for } \boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F}).$$

By Theorem 10.3.2, we shall learn that $Q(\boldsymbol{A})$ is actually the determinant of $\boldsymbol{A}$.

Following the same arguments as in Part 2, Q can be regarded as a multilinear form on $\mathbb{F}^n$.

Let $\boldsymbol{A} = \begin{pmatrix}\boldsymbol{a_1} & \boldsymbol{a_2} & \cdots & \boldsymbol{a_n}\end{pmatrix}$ where $\boldsymbol{a_i} = (a_{1i}, a_{2i}, \ldots, a_{ni})^{\mathrm{T}}$. Suppose $\boldsymbol{a_\alpha} = \boldsymbol{a_\beta}$ for some $\alpha\neq\beta$. Let $\tau = \phi_{\alpha,\beta}$. Then $a_{i,\tau(j)} = a_{ij}$ for all i, j. For any $\sigma\in\mathcal{S}_n$,

$$\begin{aligned}
&a_{\sigma(1),1}\, a_{\sigma(2),2}\cdots a_{\sigma(n),n}\\
&= a_{\sigma(1),\tau(1)}\, a_{\sigma(2),\tau(2)}\cdots a_{\sigma(n),\tau(n)}\\
&= a_{\sigma(\tau^{-1}(1)),1}\, a_{\sigma(\tau^{-1}(2)),2}\cdots a_{\sigma(\tau^{-1}(n)),n} && \text{(by rearranging the order of the terms)}\\
&= a_{\sigma\tau(1),1}\, a_{\sigma\tau(2),2}\cdots a_{\sigma\tau(n),n} && \text{(because } \tau^{-1}=\tau).
\end{aligned}$$

By Lemma 10.1.6.2, we have $\{\sigma\tau \mid \sigma\in\mathcal{S}_n\} = \mathcal{S}_n$; and by (10.1) in the proof of Theorem 10.1.11, we have $\operatorname{sgn}(\sigma\tau) = -\operatorname{sgn}(\sigma)$ for all $\sigma\in\mathcal{S}_n$. So

$$\begin{aligned}
Q(\boldsymbol{A}) &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2}\cdots a_{\sigma(n),n}\\
&= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma\tau)\, a_{\sigma\tau(1),1}\, a_{\sigma\tau(2),2}\cdots a_{\sigma\tau(n),n}\\
&= -\sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2}\cdots a_{\sigma(n),n}\\
&= -Q(\boldsymbol{A}).
\end{aligned}$$

Suppose $1+1 \neq 0$ in $\mathbb{F}$. Then $Q(\boldsymbol{A}) = 0$. (Actually, it is also true for the case when $1+1=0$ in $\mathbb{F}$, see Question 10.10.)

Hence Q is an alternative multilinear form.

Theorem 10.2.3 Let $T : V^n \to \mathbb{F}$ be an alternative multilinear form on a vector space V. Then for all $\sigma \in \mathcal{S}_n$ and $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n} \in V$,

$$T(\boldsymbol{u_1},\, \boldsymbol{u_2},\, \ldots,\, \boldsymbol{u_n}) = \operatorname{sgn}(\sigma)\, T(\boldsymbol{u_{\sigma(1)}},\, \boldsymbol{u_{\sigma(2)}},\, \ldots,\, \boldsymbol{u_{\sigma(n)}}).$$

Proof By Corollary 10.1.12, we only need to show that for all $\tau = \phi_{\alpha,\beta}$, where $1 \le \alpha < \beta \le n$, $T(\boldsymbol{u_1},\, \boldsymbol{u_2},\, \ldots,\, \boldsymbol{u_n}) = -T(\boldsymbol{u_{\tau(1)}},\, \boldsymbol{u_{\tau(2)}},\, \ldots,\, \boldsymbol{u_{\tau(n)}})$.

In the following computation, we use $T(\ldots,\, \boldsymbol{x},\, \ldots,\, \boldsymbol{y},\, \ldots)$ to denote the function value of T when the αth term is $\boldsymbol{x}$, the βth term is $\boldsymbol{y}$ and the ith term, for $i \neq \alpha, \beta$, is $\boldsymbol{u_i}$.

Since T is alternative,

$$\begin{aligned}
0 &= T(\ldots,\, \boldsymbol{u_\alpha} + \boldsymbol{u_\beta},\, \ldots,\, \boldsymbol{u_\alpha} + \boldsymbol{u_\beta},\, \ldots) \\
&= T(\ldots,\, \boldsymbol{u_\alpha},\, \ldots,\, \boldsymbol{u_\alpha} + \boldsymbol{u_\beta},\, \ldots) + T(\ldots,\, \boldsymbol{u_\beta},\, \ldots,\, \boldsymbol{u_\alpha} + \boldsymbol{u_\beta},\, \ldots) \\
&= T(\ldots,\, \boldsymbol{u_\alpha},\, \ldots,\, \boldsymbol{u_\alpha},\, \ldots) + T(\ldots,\, \boldsymbol{u_\alpha},\, \ldots,\, \boldsymbol{u_\beta},\, \ldots) \\
&\qquad + T(\ldots,\, \boldsymbol{u_\beta},\, \ldots,\, \boldsymbol{u_\alpha},\, \ldots) + T(\ldots,\, \boldsymbol{u_\beta},\, \ldots,\, \boldsymbol{u_\beta},\, \ldots) \\
&= T(\ldots,\, \boldsymbol{u_\alpha},\, \ldots,\, \boldsymbol{u_\beta},\, \ldots) + T(\ldots,\, \boldsymbol{u_\beta},\, \ldots,\, \boldsymbol{u_\alpha},\, \ldots).
\end{aligned}$$

Thus $T(\boldsymbol{u_1},\, \boldsymbol{u_2},\, \ldots,\, \boldsymbol{u_n}) = -T(\boldsymbol{u_{\tau(1)}},\, \boldsymbol{u_{\tau(2)}},\, \ldots,\, \boldsymbol{u_{\tau(n)}})$.

Remark 10.2.4 Let $T : V^n \to \mathbb{F}$ be a multilinear form on a finite dimensional vector space V over a field $\mathbb{F}$. Fix a basis $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \ldots, \boldsymbol{v_m}\}$ for V. Take any $\boldsymbol{u_1},\, \boldsymbol{u_2},\, \ldots,\, \boldsymbol{u_n} \in V$, let

$$\begin{aligned}
\boldsymbol{u_1} &= a_{11}\boldsymbol{v_1} + a_{21}\boldsymbol{v_2} + \cdots + a_{m1}\boldsymbol{v_m}, \\
\boldsymbol{u_2} &= a_{12}\boldsymbol{v_1} + a_{22}\boldsymbol{v_2} + \cdots + a_{m2}\boldsymbol{v_m}, \\
&\ \ \vdots \\
\boldsymbol{u_n} &= a_{1n}\boldsymbol{v_1} + a_{2n}\boldsymbol{v_2} + \cdots + a_{mn}\boldsymbol{v_m}
\end{aligned}$$

where $a_{11}, a_{12}, \ldots, a_{mn} \in \mathbb{F}$.

1. Let $\mathcal{F}$ be the set of all mapping from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, m\}$. Then apply (10.2) repeatedly (see Example 10.2.5), we have

$$T(\boldsymbol{u_1},\, \boldsymbol{u_2},\, \ldots,\, \boldsymbol{u_n}) = \sum_{f \in \mathcal{F}} a_{f(1),1}\, a_{f(2),2} \cdots a_{f(n),n}\, T(\boldsymbol{v_{f(1)}},\, \boldsymbol{v_{f(2)}},\, \ldots,\, \boldsymbol{v_{f(n)}}). \tag{10.3}$$

2. Suppose T is an alternative form. Then in (10.3), $T(\boldsymbol{v_{f(1)}},\, \boldsymbol{v_{f(2)}},\, \ldots,\, \boldsymbol{v_{f(n)}}) = 0$ when f is not injective, i.e. there exists $\alpha, \beta \in \{1, 2, \ldots, n\}$ such that $\alpha \neq \beta$ and $f(\alpha) = f(\beta)$.

(a) If $m < n$, then $T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) = 0$ for all $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n} \in V$, i.e. T is a zero mapping.

(b) If $m \geq n$, then (10.3) is still holds if we change the set $\mathcal{F}$ to the set of all injective mapping from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, m\}$.

In particular, When $m = n$, we can replace $\mathcal{F}$ by $\mathcal{S}_n$. By Theorem 10.2.3, we get

$$\begin{aligned} T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) &= \sum_{\sigma \in \mathcal{S}_n} a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(n),n}\, T(\boldsymbol{v_{\sigma(1)}}, \boldsymbol{v_{\sigma(2)}}, \ldots, \boldsymbol{v_{\sigma(n)}}) \\ &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(n),n}\, T(\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}). \end{aligned} \tag{10.4}$$

Example 10.2.5 Let T be a bilinear form on $\mathbb{F}^2$. Take the standard basis $\{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ for $\mathbb{F}^2$. Then for all $(a_{11}, a_{21})^{\mathrm{T}}$, $(a_{12}, a_{22})^{\mathrm{T}} \in \mathbb{F}^2$,

$$\begin{aligned} & T\left(\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}\right) \\ &= T(a_{11}\boldsymbol{e_1} + a_{21}\boldsymbol{e_2},\, a_{12}\boldsymbol{e_1} + a_{22}\boldsymbol{e_2}) \\ &= a_{11}\, T(\boldsymbol{e_1},\, a_{12}\boldsymbol{e_1} + a_{22}\boldsymbol{e_2}) + a_{21}\, T(\boldsymbol{e_2},\, a_{12}\boldsymbol{e_1} + a_{22}\boldsymbol{e_2}) \\ &= a_{11}a_{12}\, T(\boldsymbol{e_1}, \boldsymbol{e_1}) + a_{11}a_{22}\, T(\boldsymbol{e_1}, \boldsymbol{e_2}) + a_{21}a_{12}\, T(\boldsymbol{e_2}, \boldsymbol{e_1}) + a_{21}a_{22}\, T(\boldsymbol{e_2}, \boldsymbol{e_2}). \end{aligned}$$

Furthermore, suppose T is an alternative bilinear form. Then $T(\boldsymbol{e_1}, \boldsymbol{e_1}) = 0$, $T(\boldsymbol{e_2}, \boldsymbol{e_2}) = 0$ and $T(\boldsymbol{e_1}, \boldsymbol{e_2}) = -T(\boldsymbol{e_2}, \boldsymbol{e_1})$. Hence

$$T\left(\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}\right) = (a_{11}a_{22} - a_{21}a_{12})\, T(\boldsymbol{e_1}, \boldsymbol{e_2}).$$

Section 10.3 Determinants

Definition 10.3.1 A mapping $D : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is called a *determinant function* on $\mathcal{M}_{n\times n}(\mathbb{F})$ if it satisfies the following axioms:

(D1) By regarding the columns of matrices in $\mathcal{M}_{n\times n}(\mathbb{F})$ as vectors in $\mathbb{F}^n$ (see Example 10.2.2.2), D is a multilinear form on $\mathbb{F}^n$.

(D2) $D(\boldsymbol{A}) = 0$ if $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ has two identical columns, i.e. as a multilinear form on $\mathbb{F}^n$, D is alternative.

(D3) $D(\boldsymbol{I_n}) = 1$.

Theorem 10.3.2 There exists one and only one determinant function on $\mathcal{M}_{n\times n}(\mathbb{F})$ and it is the function $\det : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ defined by

$$\det(\boldsymbol{A}) = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(n),n} \quad \text{for } \boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F}). \tag{10.5}$$

This formula is known as the *classical definition of determinants.*

Proof Note that the function det is the function Q in Example 10.2.2.3. By Example 10.2.2.3 and Question 10.10, det is an alternative multilinear form on $\mathbb{F}^n$ and hence it satisfies (D1) and (D2). Since $\boldsymbol{I_n} = (\delta_{ij})$ where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$, for any $\sigma \in \mathcal{S}_n$, $\delta_{\sigma(1),1}\, \delta_{\sigma(2),2} \cdots \delta_{\sigma(n),n} = 0$ whenever σ is not the identity mapping. Then

$$\det(\boldsymbol{I_n}) = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, \delta_{\sigma(1),1}\, \delta_{\sigma(2),2} \cdots \delta_{\sigma(n),n} = \delta_{11}\, \delta_{22} \cdots \delta_{nn} = 1$$

and hence it satisfies (D3). So det is a determinant function on $\mathcal{M}_{n\times n}(\mathbb{F})$.

On the other hand, suppose D is a determinant function on $\mathcal{M}_{n\times n}(\mathbb{F})$. Take any $\boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. By applying (10.4) to the standard basis $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ for $\mathbb{F}^n$,

$$\begin{aligned} D(\boldsymbol{A}) &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(n),n}\, D\Big(\begin{pmatrix} \boldsymbol{e_1} & \boldsymbol{e_2} & \cdots & \boldsymbol{e_n} \end{pmatrix}\Big) \\ &= \det(\boldsymbol{A})\, D(\boldsymbol{I_n}) \\ &= \det(\boldsymbol{A}). \end{aligned}$$

Hence $D = \det$.

Example 10.3.3

1. In $\mathcal{S}_1$, there is only one permutation $\sigma = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

   Let $\boldsymbol{A} = (a_{11}) \in \mathcal{M}_{1\times 1}(\mathbb{F})$. Then

   $$\det(\boldsymbol{A}) = \operatorname{sgn}(\sigma)\, a_{\sigma(1),1} = a_{11}.$$

   It coincides with Definition 2.5.2.

2. In $\mathcal{S}_2$, there are two permutations: $\sigma_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ and $\sigma_2 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.

   Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{F})$. Then

   $$\det(\boldsymbol{A}) = \operatorname{sgn}(\sigma_1)\, a_{\sigma_1(1),1}\, a_{\sigma_1(2),2} + \operatorname{sgn}(\sigma_2)\, a_{\sigma_2(1),1}\, a_{\sigma_2(2),2} = a_{11}\, a_{22} - a_{21}\, a_{12}.$$

   This gives us the same formula as in Example 2.5.4.1.

3. In $\mathcal{S}_3$, there are six permutations: $\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, $\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, $\sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$, $\sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ and $\sigma_6 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$.

   Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in \mathcal{M}_{3\times 3}(\mathbb{F})$. Then

$$\begin{aligned}\det(\boldsymbol{A}) &= \sum_{j=1}^{6} \operatorname{sgn}(\sigma_j)\, a_{\sigma_j(1),1}\, a_{\sigma_j(2),2}\, a_{\sigma_j(3),3} \\ &= a_{11}\, a_{22}\, a_{33} + a_{31}\, a_{12}\, a_{23} + a_{21}\, a_{32}\, a_{13} - a_{31}\, a_{22}\, a_{13} - a_{11}\, a_{32}\, a_{23} - a_{21}\, a_{12}\, a_{33}.\end{aligned}$$

   This gives us the same formula as in Remark 2.5.5.

Lemma 10.3.4 Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Then $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.

Proof (Note that the theorem is the same as Theorem 2.5.10. However, since we need this theorem in the proof of the cofactor expansions, Theorem 10.3.5, while the proof of Theorem 2.5.10 uses the property of cofactor expansions, we need to reprove the theorem using only results learnt in this chapter so far.)

Let $\boldsymbol{A} = (a_{ij})$. For any $\sigma \in \mathcal{S}_n$, by rearranging the order of the terms,

$$a_{\sigma(1),1}\, a_{\sigma(2),2}\, \cdots\, a_{\sigma(n),n} = a_{1,\sigma^{-1}(1)}\, a_{2,\sigma^{-1}(2)}\, \cdots\, a_{n,\sigma^{-1}(n)}.$$

By Lemma 10.1.6.1 and Corollary 10.1.12.3, we have $\{\sigma^{-1} \mid \sigma \in \mathcal{S}_n\} = \mathcal{S}_n$ and $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$. Thus

$$\begin{aligned}\det(\boldsymbol{A}) &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2}\, \cdots\, a_{\sigma(n),n} \\ &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{1,\sigma^{-1}(1)}\, a_{2,\sigma^{-1}(2)}\, \cdots\, a_{n,\sigma^{-1}(n)} \\ &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma^{-1})\, a_{1,\sigma^{-1}(1)}\, a_{2,\sigma^{-1}(2)}\, \cdots\, a_{n,\sigma^{-1}(n)} \\ &= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{1,\sigma(1)}\, a_{2,\sigma(2)}\, \cdots\, a_{n,\sigma(n)} \\ &= \det(\boldsymbol{A}^{\mathrm{T}}).\end{aligned}$$

Theorem 10.3.5 (Cofactor Expansions) Let $\boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. Define $\widetilde{\boldsymbol{A}}_{ij}$ to be the $(n-1)\times(n-1)$ matrix obtained from $\boldsymbol{A}$ by deleting the ith row and the jth column. Then for any $\alpha = 1, 2, \dots n$ and $\beta = 1, 2, \dots n$,

$$\begin{aligned}\det(\boldsymbol{A}) &= a_{\alpha 1}A_{\alpha 1} + a_{\alpha 2}A_{\alpha 2} + \cdots + a_{\alpha n}A_{\alpha n} \\ &= a_{1\beta}A_{1\beta} + a_{2\beta}A_{2\beta} + \cdots + a_{n\beta}A_{n\beta}\end{aligned}$$

where $A_{ij} = (-1)^{i+j}\det(\widetilde{\boldsymbol{A}}_{ij})$. (See Theorem 2.5.6.)

Proof For $m = 1, 2, \dots, n$, define mappings $E_m : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ such that

$$E_m(\boldsymbol{A}) = \sum_{k=1}^{n}(-1)^{m+k}a_{mk}\det(\widetilde{\boldsymbol{A}}_{mk}) \quad \text{for } \boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F}).$$

It can be checked that each E_m is a determinant function (see Question 10.17). Thus by Theorem 10.3.2,

$$\det(\boldsymbol{A}) = E_\alpha(\boldsymbol{A}) = \sum_{k=1}^{n}(-1)^{\alpha+k}a_{\alpha k}\det(\widetilde{\boldsymbol{A}}_{\alpha k}) = a_{\alpha 1}A_{\alpha 1} + a_{\alpha 2}A_{\alpha 2} + \cdots + a_{\alpha n}A_{\alpha n}.$$

Let $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}} = (b_{ij})$. Note that $b_{ij} = a_{ji}$ and $\widetilde{\boldsymbol{B}}_{ij} = (\widetilde{\boldsymbol{A}}_{ji})^{\mathrm{T}}$ for all i, j. Applying the result above to $\boldsymbol{B}$, by Lemma 10.3.4,

$$\begin{aligned}\det(\boldsymbol{A}) = \det(\boldsymbol{B}) = E_\beta(\boldsymbol{B}) &= \sum_{k=1}^{n}(-1)^{\beta+k}b_{\beta k}\det(\widetilde{\boldsymbol{B}}_{\beta k}) \\ &= \sum_{k=1}^{n}(-1)^{k+\beta}a_{k\beta}\det(\widetilde{\boldsymbol{A}}_{k\beta}) \\ &= a_{1\beta}A_{1\beta} + a_{2\beta}A_{2\beta} + \cdots + a_{n\beta}A_{n\beta}.\end{aligned}$$

Remark 10.3.6 By Example 10.3.3.1 and Theorem 10.3.5, the determinant function defined in Theorem 10.3.2 is the same as the determinant defined inductively in Definition 2.5.2.

Exercise 10

Question 10.1 to Question 10.5 are exercises for Section 10.1.

1. How many different permutations are there in $\mathcal{S}_4$? List all the permutations in $\mathcal{S}_4$ and determine their signs.

2. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix}$.

   (a) Write down σ^{-1}, τ^{-1}, $\sigma\tau$ and $\tau\sigma$.

(b) Find $x, y \in \mathcal{S}_5$ such that $\sigma x = \tau$ and $y\sigma = \tau$.

(c) Compute $\operatorname{sgn}(\sigma)$, $\operatorname{sgn}(\sigma^{-1})$, $\operatorname{sgn}(\tau)$, $\operatorname{sgn}(\tau^{-1})$, $\operatorname{sgn}(\sigma\tau)$ and $\operatorname{sgn}(\tau\sigma)$.

(d) Decompose σ into a product of transpositions of the form $\phi_{\alpha,\alpha+1}$ as stated in Lemma 10.1.7.

3. (a) Let $\sigma \in \mathcal{S}_n$. Prove that $\operatorname{sgn}(\sigma) = \displaystyle\prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i}$.

   (b) Use (a) to show that for any $\theta \in \mathcal{S}_n$ and $1 \le \alpha \le n-1$, $\operatorname{sgn}(\theta\, \phi_{\alpha,\,\alpha+1}) = -\operatorname{sgn}(\theta)$.

   (This is an alternative proof of the first part of the proof of Theorem 10.1.11.

   Hint: Compute $\dfrac{\operatorname{sgn}(\theta\, \phi_{\alpha,\,\alpha+1})}{\operatorname{sgn}(\theta)}$ using the formula in (a).)

4. Prove Lemma 10.1.6 and Corollary 10.1.12.3:

   (a) Show that $\{\sigma^{-1} \mid \sigma \in \mathcal{S}_n\} = \mathcal{S}_n$.

   (b) For any $\tau \in \mathcal{S}_n$, show that $\{\tau\sigma \mid \sigma \in \mathcal{S}_n\} = \{\sigma\tau \mid \sigma \in \mathcal{S}_n\} = \mathcal{S}_n$.

   (c) For any $\sigma \in \mathcal{S}_n$, prove that $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$.

5. Let $\mathcal{O}_1 = \{\sigma \in \mathcal{S}_n \mid \operatorname{sgn}(\sigma) = 1\}$ and $\mathcal{O}_2 = \{\sigma \in \mathcal{S}_n \mid \operatorname{sgn}(\sigma) = -1\}$ where $n \ge 2$.

   (a) Prove that for any $\tau \in \mathcal{O}_2$, $\mathcal{O}_2 = \{\sigma\tau \mid \sigma \in \mathcal{O}_1\}$.

   (b) Find $|\mathcal{O}_1|$ and $|\mathcal{O}_2|$.

Question 10.6 to Question 10.14 are exercises for Section 10.2.

6. Determine which of the following are bilinear forms on $\mathbb{R}^2$.

   (a) $T(\boldsymbol{u}, \boldsymbol{v}) = 0$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$.

   (b) $T(\boldsymbol{u}, \boldsymbol{v}) = 2$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$.

   (c) $T(\boldsymbol{u}, \boldsymbol{v}) = u_1 + u_2 + v_1 + v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

   (d) $T(\boldsymbol{u}, \boldsymbol{v}) = u_1 v_2 + u_2 v_1$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

   (e) $T(\boldsymbol{u}, \boldsymbol{v}) = u_1 u_2 + v_1 v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

7. Let V be a vector space over a field $\mathbb{F}$. A bilinear form T on V is called *symmetric* if $T(\boldsymbol{u}, \boldsymbol{v}) = T(\boldsymbol{v}, \boldsymbol{u})$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$. Determine which of the following bilinear forms on $\mathbb{R}^2$ are alternative and/or symmetric.

   (a) $T(\boldsymbol{u}, \boldsymbol{v}) = 0$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

   (b) $T(\boldsymbol{u}, \boldsymbol{v}) = -u_1 v_2 - u_2 v_1$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(c) $T(\boldsymbol{u}, \boldsymbol{v}) = -u_1v_2 + u_2v_1$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(d) $T(\boldsymbol{u}, \boldsymbol{v}) = 2u_1v_1 + u_1v_2 + u_2v_1 + 3u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(e) $T(\boldsymbol{u}, \boldsymbol{v}) = 2u_1v_1 - u_1v_2 + u_2v_1 + 3u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

8. (In this question, vectors in $\mathbb{F}^n$ are written as column vectors.)

   Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Define $T : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ by

   $$T(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v} \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^n.$$

   Show that T is a bilinear form on $\mathbb{F}^n$.

9. If $1 + 1 \neq 0$ in $\mathbb{F}$, prove that a multilinear form T on V is alternative if and only if for all transposition $\tau = \phi_{\alpha,\beta} \in \mathcal{S}_n$, $1 \leq \alpha < \beta \leq n$, and $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n} \in V$, $T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) = -T(\boldsymbol{u_{\tau(1)}}, \boldsymbol{u_{\tau(2)}}, \ldots, \boldsymbol{u_{\tau(n)}})$.

   (Hint: The ($\Rightarrow$) part has already been shown in Theorem 10.2.3. You only need to prove the ($\Leftarrow$) part.)

10. Prove that the multilinear form Q in Example 10.2.2.3 is alternative when $1 + 1 = 0$ in $\mathbb{F}$. (Hint: See Question 10.5.)

11. Let V be a vector space over a field $\mathbb{F}$ and $T : V^n \to \mathbb{F}$ an alternative multilinear form. If $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ are linearly dependent vectors in V, prove that $T(\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}) = 0$.

12. Suppose V is finite dimensional. Let B be an ordered basis for V and T a bilinear form on V.

    (a) Show that there exists a square matrix $\boldsymbol{A}$ over $\mathbb{F}$ such that $T(\boldsymbol{u}, \boldsymbol{v}) = ([\boldsymbol{u}]_B)^{\mathrm{T}}\boldsymbol{A}[\boldsymbol{v}]_B$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$.

    (b) Prove that T is symmetric if and only if the matrix $\boldsymbol{A}$ in (a) is symmetric. (See Question 10.7 for the definition of symmetric bilinear forms.)

    (c) If $1 + 1 \neq 0$ in $\mathbb{F}$, prove that T is alternative if and only if the matrix $\boldsymbol{A}$ in (a) is skew symmetric.

13. Let V be a vector space over a field $\mathbb{F}$. A function $Q : V \to \mathbb{F}$ is called a *quadratic form* on V if it satisfies the following two axioms.

    (Q1) For all $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $Q(c\boldsymbol{u}) = c^2Q(\boldsymbol{u})$.

    (Q2) The mapping $H : V^2 \to \mathbb{F}$ defined by

    $$H(\boldsymbol{u}, \boldsymbol{v}) = Q(\boldsymbol{u} + \boldsymbol{v}) - Q(\boldsymbol{u}) - Q(\boldsymbol{v}) \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in V$$

    is a bilinear form.

(a) Suppose $1+1 \neq 0$ in $\mathbb{F}$. Prove that a mapping $Q : V \to \mathbb{F}$ is a quadratic form on V if and only if there exists a symmetric bilinear form T on V such that $Q(\boldsymbol{u}) = T(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$.

(b) Give an example of a quadratic form Q on $V = \mathbb{F}_2^2$ such that we cannot find a symmetric bilinear form T on V such that $Q(\boldsymbol{u}) = T(\boldsymbol{u}, \boldsymbol{u})$ for all $\boldsymbol{u} \in V$.

14. Suppose V is finite dimensional vector space over $\mathbb{R}$ with $n = \dim(V)$. Let Q be a quadratic form on V. Prove that there exist $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ and a basis $B = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n}\}$ for V such that

$$Q(\boldsymbol{u}) = \lambda_1 a_1^2 + \lambda_2 a_2^2 + \cdots + \lambda_n a_n^2 \quad \text{for } \boldsymbol{u} = a_1\boldsymbol{w_1} + a_2\boldsymbol{w_2} + \cdots + a_n\boldsymbol{w_n} \in V.$$

(Hint: See Section 6.4.)

Question 10.15 to Question 10.23 are exercises for Section 10.3.

15. Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 4 & 1 \end{pmatrix}$, $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix}$ and

$$(a_{ij})_{5\times 5} = \begin{pmatrix} 3 & -1 & -9 & 3 & 1 \\ 1 & 2 & -1 & 7 & -1 \\ 0 & 6 & -5 & -1 & -1 \\ 4 & 3 & -2 & 1 & -1 \\ -2 & -3 & 2 & 9 & 2 \end{pmatrix}.$$

Compute

(a) $\operatorname{sgn}(\sigma)\, a_{\sigma(1),1}a_{\sigma(2),2}a_{\sigma(3),3}a_{\sigma(4),4}a_{\sigma(5),5}$ and

(b) $\operatorname{sgn}(\tau)\, a_{\tau(1),1}a_{\tau(2),2}a_{\tau(3),3}a_{\tau(4),4}a_{\tau(5),5}$.

16. Use the formula in Theorem 10.3.2 to write down a formula for $\det(\boldsymbol{A})$ where $\boldsymbol{A} = (a_{ij})_{4\times 4}$.

17. (This question is part of the proof of cofactor expansions, Theorem 10.3.5. You should not use the properties of determinants learnt from Section 2.5.)

Let $E_m : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$, $1 \leq m \leq n$, be the mapping defined in Theorem 10.3.5, i.e.

$$E_m(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{m+k} a_{mk} \det(\widetilde{\boldsymbol{A}}_{mk}) \quad \text{for } \boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F}).$$

Prove that E_m is a determinant function.

18. Reprove Theorem 10.3.5 using the classical definition of determinants:

    Let $\boldsymbol{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. Define $\widetilde{\boldsymbol{A}}_{ij}$ to be the $(n-1)\times(n-1)$ matrix obtained from $\boldsymbol{A}$ by deleting the ith row and the jth column. Using the formula (10.5), prove that for any $\alpha = 1, 2, \dots n$ and $\beta = 1, 2, \dots n$,

$$\begin{aligned}\det(\boldsymbol{A}) &= a_{\alpha 1}A_{\alpha 1} + a_{\alpha 2}A_{\alpha 2} + \cdots + a_{\alpha n}A_{\alpha n}\\ &= a_{1\beta}A_{1\beta} + a_{2\beta}A_{2\beta} + \cdots + a_{n\beta}A_{n\beta}\end{aligned}$$

    where $A_{\alpha\beta} = (-1)^{\alpha+\beta}\det(\widetilde{\boldsymbol{A}}_{\alpha\beta})$.

    (Hint: First, use Theorem 10.2.3 and Lemma 10.3.4 to find out what happen to the determinant if we interchanging two columns or two rows of $\boldsymbol{A}$.)

19. Let $\boldsymbol{D}$ be an $(m+n)\times(m+n)$ matrix such that

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0_{n\times m}} & \boldsymbol{C} \end{pmatrix}$$

    where $\boldsymbol{A}$ is an $m\times m$ matrix, $\boldsymbol{B}$ is an $m\times n$ matrix and $\boldsymbol{C}$ is an $n\times n$ matrix. Prove that $\det(\boldsymbol{D}) = \det(\boldsymbol{A})\det(\boldsymbol{C})$.

20. Let $\boldsymbol{W}$ be an $(m+n)\times(m+n)$ matrix such that

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Y} \\ \boldsymbol{Z} & \boldsymbol{0_{n\times m}} \end{pmatrix}$$

    where $\boldsymbol{X}$ is an $m\times n$ matrix, $\boldsymbol{Y}$ is an $m\times m$ matrix and $\boldsymbol{Z}$ is an $n\times n$ matrix. Prove that $\det(\boldsymbol{W}) = (-1)^{mn}\det(\boldsymbol{Y})\det(\boldsymbol{Z})$.

21. Let $\boldsymbol{A}$ be an $n\times n$ matrix. Explain why $p(x) = \det(x\boldsymbol{I_n} - \boldsymbol{A})$ is a polynomial of degree n. Also, show that the coefficient of x^n in $p(x)$ is equal to 1, the coefficient of x^{n-1} is equal to $-\mathrm{tr}(\boldsymbol{A})$ and the constant term is equal to $(-1)^n\det(\boldsymbol{A})$.

    (Note that $p(x)$ is the characteristic polynomial of $\boldsymbol{A}$, see Definition 6.1.6.)

22. Let $a_1, a_2, \dots, a_n$ be elements of a field. Prove that the value of the Vandermonde determinant

$$\det\Big((a_j^{i-1})_{n\times n}\Big) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & & \vdots \\ a_1^{n-1} & a_2^{n-1} & \dots & a_n^{n-1} \end{vmatrix}$$

    is equal to $\displaystyle\prod_{1\le i<j\le n}(a_j - a_i)$.

23. Let $a_0, a_1, \ldots, a_{n-1}$ and b be elements of a field. Prove that

$$\begin{vmatrix} b & & & & & -a_0 \\ -1 & b & & 0 & & -a_1 \\ & -1 & b & & & -a_2 \\ & & \ddots & \ddots & & \vdots \\ & 0 & & \ddots & b & -a_{k-2} \\ & & & & -1 & b - a_{k-1} \end{vmatrix} = -a_0 - a_1 b - \cdots - a_{k-1} b^{k-1} + b^k.$$

Chapter 11

Diagonalization and Jordan Canonical Forms

Section 11.1 Eigenvalues and Diagonalization

Discussion 11.1.1 In Section 6.2, we have studied the problem of diagonalization of real square matrices. The procedure to diagonalize square matrices over other fields is exactly the same. By Remark 9.3.12, we learnt that to diagonalize a square matrix $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ is the same as to find an ordered basis B for $\mathbb{F}^n$ such that the matrix for $L_{\boldsymbol{A}}$ relative to B is a diagonal matrix. In this section, we shall restate a few important results in Chapter 6 in terms of linear operators.

Definition 11.1.2 Let V be a vector space and $T : V \to V$ a linear operator. A nonzero vector $\boldsymbol{u} \in V$ is called an *eigenvector* of T if $T(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}\}$, i.e. $T(\boldsymbol{u}) = \lambda\boldsymbol{u}$ for some scalar λ. The scalar λ is called an *eigenvalue* of T and $\boldsymbol{u}$ is said to be an eigenvector of T *associated* with the eigenvalue λ.

Example 11.1.3

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator defined by $T((x, y)) = (x + 3y, x - y)$ for $(x, y) \in \mathbb{R}^2$. Since

$$T((3,1)) = (6,2) = 2(3,1) \quad \text{and} \quad T((1,-1)) = (-2,2) = -2(1,-1),$$

   2 and -2 are eigenvalues of T, $(3,1)$ is an eigenvector of T associated with 2 and $(1,-1)$ is an eigenvector associated with -2.

2. We define the *shift operator* S on the vector space of infinite sequences over a field $\mathbb{F}$ such that

$$S((a_n)_{n\in\mathbb{N}}) = (a_{n+1})_{n\in\mathbb{N}} = (a_2, a_3, a_4, \dots) \quad \text{for } (a_n)_{n\in\mathbb{N}} = (a_1, a_2, a_3, \dots) \in \mathbb{F}^{\mathbb{N}}.$$

For any $\lambda \in \mathbb{F}$, let $\boldsymbol{a_\lambda} = (\lambda^{n-1})_{n\in\mathbb{N}} = (1, \lambda, \lambda^2, \dots) \in \mathbb{F}^{\mathbb{N}}$. Then

$$S(\boldsymbol{a_\lambda}) = (\lambda^n)_{n\in\mathbb{N}} = (\lambda, \lambda^2, \lambda^3, \dots) = \lambda(1, \lambda, \lambda^2, \dots) = \lambda(\lambda^{n-1})_{n\in\mathbb{N}} = \lambda\boldsymbol{a_\lambda}.$$

Thus every scalar λ is an eigenvalue of S and $\boldsymbol{a_\lambda}$ is an eigenvector of S associated with λ.

3. Let $[a, b]$, with $a < b$, be a closed interval on the real line and let $D : C^\infty([a, b]) \to C^\infty([a, b])$ be the differential operator defined in Example 9.1.4.6. For any $\lambda \in \mathbb{R}$, let $f_\lambda \in C^\infty([a, b])$ be the function defined by $f_\lambda(x) = e^{\lambda x}$ for $x \in [a, b]$. Then

$$\begin{aligned} & D(f_\lambda)(x) = \frac{de^{\lambda x}}{dx} = \lambda e^{\lambda x} = \lambda f_\lambda(x) \quad \text{for all } x \in [a, b] \\ \Rightarrow \quad & D(f_\lambda) = \lambda f_\lambda. \end{aligned}$$

Thus every real number λ is an eigenvalue of D and f_λ is an eigenvector of D associated with λ.

Definition 11.1.4 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. The *determinant* of T, denoted by $\det(T)$, is defined to be the determinant of the matrix $[T]_B$ where B is any ordered basis for V.

Remark 11.1.5 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. Suppose B and C are two ordered bases for V. By Theorem 9.3.10, $[T]_B$ and $[T]_C$ are similar, i.e. $[T]_B = \boldsymbol{P}^{-1}[T]_C\boldsymbol{P}$ for some invertible matrix $\boldsymbol{P}$. Then

$$\begin{aligned} \det([T]_B) = \det(\boldsymbol{P}^{-1}[T]_C\boldsymbol{P}) &= \det(\boldsymbol{P}^{-1})\det([T]_C)\det(\boldsymbol{P}) \\ &= \det(\boldsymbol{P})^{-1}\det([T]_C)\det(\boldsymbol{P}) = \det([T]_C). \end{aligned}$$

So the definition of $\det(T)$ is independent of the choice of the basis B.

Theorem 11.1.6 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. For a scalar λ, let $\lambda I_V - T$ be the linear operator defined by $(\lambda I_V - T)(\boldsymbol{u}) = \lambda\boldsymbol{u} - T(\boldsymbol{u})$ for $\boldsymbol{u} \in V$.

1. λ is an eigenvalue of T if and only if $\det(\lambda I_V - T) = 0$.

   (The equation $\det(xI_V - T) = 0$ is called the *characteristic equation* of T and the polynomial $\det(xI_V - T)$ is called the *characteristic polynomial* of T.)

2. $\boldsymbol{u} \in V$ is an eigenvector of T associated with λ if and only if $\boldsymbol{u}$ is a nonzero vector in $\mathrm{Ker}(T - \lambda I_V)$ $(= \mathrm{Ker}(\lambda I_V - T))$.

   (The subspace $\mathrm{Ker}(T - \lambda I_V)$ of V is called the *eigenspace of T associated with* λ.)

Proof The proof of Part 1 follows the same argument as the proof for Remark 6.1.5. For Part 2, let $\boldsymbol{u} \in V$,

$$T(\boldsymbol{u}) = \lambda \boldsymbol{u} \quad \Leftrightarrow \quad (T - \lambda I_V)(\boldsymbol{u}) = T(\boldsymbol{u}) - \lambda \boldsymbol{u} = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{u} \in \operatorname{Ker}(T - \lambda I_V).$$

So $\boldsymbol{u}$ is an eigenvector of T associated with λ if and only if $\boldsymbol{u}$ is a nonzero vector in $\operatorname{Ker}(T - \lambda I_V)$.

Notation 11.1.7 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. We use $c_T(x)$ to denote the characteristic polynomial of T, i.e. $c_T(x) = \det(xI_V - T)$. For an eigenvalue λ of T, we use $E_\lambda(T)$ to denote the eigenspace of T associated with λ, i.e. $E_\lambda(T) = \operatorname{Ker}(T - \lambda I_V)$.

Also, for an $n \times n$ matrix $\boldsymbol{A}$, we use $c_{\boldsymbol{A}}(x)$ to denote the characteristic polynomial of $\boldsymbol{A}$. For an eigenvalue λ of $\boldsymbol{A}$, we use $E_\lambda(\boldsymbol{A})$ to denote the eigenspace of $\boldsymbol{A}$ associated with λ. (See Definition 6.1.6 and Definition 6.1.11.)

Remark 11.1.8 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) = n \geq 1$. Take any basis B for V. Then

$$c_T(x) = \det(xI_V - T) = \det([xI_V - T]_B) = \det(x[I_V]_B - [T]_B) = \det(x\boldsymbol{I_n} - [T]_B) = c_{[T]_B}(x).$$

Thus the characteristic polynomial of T is the same as the characteristic polynomial of the matrix $[T]_B$. By the result of Question 10.21, $c_T(x)$ is a monic polynomial of degree n.

Example 11.1.9 Let V be a finite dimensional real vector space with a basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ Let $T : V \to V$ be linear operator such that

$$T(\boldsymbol{v_1}) = \boldsymbol{v_1}, \quad T(\boldsymbol{v_2}) = -2\boldsymbol{v_1} - \boldsymbol{v_2} - 2\boldsymbol{v_3} \quad \text{and} \quad T(\boldsymbol{v_3}) = 2\boldsymbol{v_1} + 2\boldsymbol{v_2} + 3\boldsymbol{v_3}$$

Then

$$[T]_B = \Big([T(\boldsymbol{v_1})]_B \quad [T(\boldsymbol{v_2})]_B \quad [T(\boldsymbol{v_3})]_B \Big) = \begin{pmatrix} 1 & -2 & 2 \\ 0 & -1 & 2 \\ 0 & -2 & 3 \end{pmatrix}.$$

Thus

$$c_T(x) = c_{[T]_B}(x) = \det(x\boldsymbol{I} - [T]_B) = \begin{vmatrix} x-1 & 2 & -2 \\ 0 & x+1 & -2 \\ 0 & 2 & x-3 \end{vmatrix} = (x-1)^3$$

and hence T has only one eigenvalue 1.

For $a, b, c \in \mathbb{R}$,

$$\begin{aligned}
a\boldsymbol{v_1} + b\boldsymbol{v_2} + c\boldsymbol{v_3} \in E_1(T) \quad &\Leftrightarrow \quad (T - I_V)(a\boldsymbol{v_1} + b\boldsymbol{v_2} + c\boldsymbol{v_3}) = \boldsymbol{0} \\
&\Leftrightarrow \quad [T - I_V]_B \, [a\boldsymbol{v_1} + b\boldsymbol{v_2} + c\boldsymbol{v_3}]_B = [\boldsymbol{0}]_B \\
&\Leftrightarrow \quad \left[\begin{pmatrix} 1 & -2 & 2 \\ 0 & -1 & 2 \\ 0 & -2 & 3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} 0 & -2 & 2 \\ 0 & -2 & 2 \\ 0 & -2 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix} = s \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \quad \text{for } s, t \in \mathbb{R}.
\end{aligned}$$

So $E_1(T) = \text{span}\{\boldsymbol{v_1},\ \boldsymbol{v_2} + \boldsymbol{v_3}\}$.

Definition 11.1.10 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. Then T is called *diagonalizable* if there exists an ordered basis B for V such that $[T]_B$ is a diagonal matrix.

Theorem 11.1.11 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. Then T is diagonalizable if and only if V has a basis B such that every vector in B is an eigenvector of T.

Proof Let $n = \dim(V)$.

($\Rightarrow$) Suppose T is diagonalizable. There exists a basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ for V such that

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \text{ i.e. for } i = 1, 2, \dots, n,\ [T(\boldsymbol{v_i})]_B = \lambda_i \boldsymbol{e_i} \text{ and } T(\boldsymbol{v_i}) = \lambda_i \boldsymbol{v_i}.$$

So we have a basis B for V such that every vector in B is an eigenvector of T.

($\Leftarrow$) Suppose there a basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ for V such that every vector in B is an eigenvector of T, say, $T(\boldsymbol{v_i}) = \lambda_i \boldsymbol{v_i}$, where λ_i is a scalar, for $i = 1, 2, \dots, n$. Then

$$\begin{aligned}
[T]_B &= \Big([T(\boldsymbol{v_1})]_B \quad [T(\boldsymbol{v_2})]_B \quad \cdots \quad [T(\boldsymbol{v_n})]_B\Big) \\
&= \Big([\lambda_1 \boldsymbol{v_1}]_B \quad [\lambda_2 \boldsymbol{v_2}]_B \quad \cdots \quad [\lambda_n \boldsymbol{v_n}]_B\Big) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.
\end{aligned}$$

So T is diagonalizable.

Algorithm 11.1.12 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) = n \geq 1$. We want to determine whether T is diagonalizable and if it is diagonalizable, find an ordered basis so that the matrix of T relative to this basis is a diagonal matrix. (See Algorithm 6.2.4 and Remark 6.2.5.)

Step 1: Find a basis C for V and compute the matrix $\boldsymbol{A} = [T]_C$.

Step 2: Factorize the characteristic polynomial $c_T(x) = c_{\boldsymbol{A}}(x)$ into linear factors (if possible), i.e. to express it in the form

$$c_{\boldsymbol{A}}(x) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of $\boldsymbol{A}$ and $r_1 + r_2 + \cdots + r_k = n$. If we cannot factorize $c_{\boldsymbol{A}}(x)$ into linear factors, T is not diagonalizable.

Step 3: For each eigenvalue λ_i, find a basis B_{λ_i} for the eigenspace $E_{\lambda_i}(T) = \mathrm{Ker}(T - \lambda_i I_V)$. If $|B_{\lambda_i}| < r_i$ for some i, T is not diagonalizable. (See Theorem 11.5.10.)

Step 4: If T can pass the tests in Step 2 and Step 3, it is diagonalizable.
Let $B = B_{\lambda_1} \cup B_{\lambda_2} \cup \cdots \cup B_{\lambda_k}$. Then B is a basis for V and $\boldsymbol{D} = [T]_B$ is a diagonal matrix.
Note that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ where $\boldsymbol{P} = [I_V]_{C,B}$ is the transition matrix from B to C.

Example 11.1.13

1. Let $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator defined by

$$T_1((x, y)) = (y, -x) \quad \text{for } (x, y) \in \mathbb{R}^2.$$

Step 1: Take the standard basis $C = \{(1, 0), (0, 1)\}$ for $\mathbb{R}^2$. Then

$$\boldsymbol{A} = [T_1]_C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Step 2: The characteristic polynomial is

$$c_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} x & -1 \\ 1 & x \end{vmatrix} = 1 + x^2$$

which cannot be factorized into linear factors over $\mathbb{R}$. So T_1 is not diagonalizable.

2. Let $T_2 : \mathcal{P}_2(\mathbb{C}) \to \mathcal{P}_2(\mathbb{C})$ be linear operator defined by

$$T_2(a + bx + cx^2) = (4a - b + c) + (a + 2b - c)x - \mathrm{i}cx^2 \quad \text{for } a + bx + cx^2 \in \mathcal{P}_2(\mathbb{C}).$$

Step 1: Take the standard basis $C = \{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{C})$. Then

$$\boldsymbol{A} = [T_2]_C = \begin{pmatrix} 4 & -1 & 1 \\ 1 & 2 & -1 \\ 0 & 0 & -\mathrm{i} \end{pmatrix}.$$

Step 2: The characteristic polynomial is

$$c_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} x-4 & 1 & -1 \\ -1 & x-2 & 1 \\ 0 & 0 & x+\mathrm{i} \end{vmatrix} = (x-3)^2(x+\mathrm{i}).$$

Thus 3 and $-\mathrm{i}$ are the eigenvalues.

Step 3: To find a basis for $E_3(T_2)$:

$$\begin{aligned} a+bx+cx^2 \in E_3(T_2) \quad &\Leftrightarrow \quad (T_2 - 3I_{\mathcal{P}_2(\mathbb{C})})(a+bx+cx^2) = 0 \\ &\Leftrightarrow \quad [T_2 - 3I_{\mathcal{P}_2(\mathbb{C})}]_C\,[a+bx+cx^2]_C = [0]_C \\ &\Leftrightarrow \quad \left[\begin{pmatrix} 4 & -1 & 1 \\ 1 & 2 & -1 \\ 0 & 0 & -\mathrm{i} \end{pmatrix} - 3\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right]\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow \quad \begin{pmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \\ 0 & 0 & -3-\mathrm{i} \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix} = t\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{for } t \in \mathbb{C}. \end{aligned}$$

Thus $B_3 = \{1+x\}$ is a basis for $E_3(T_2)$.

As $(x-3)^2$ is a factor of $c_{\boldsymbol{A}}(x)$ and $|B_3| = 1 < 2$, T_2 is not diagonalizable.

3. Let $T_3 : \mathcal{M}_{2\times 2}(\mathbb{R}) \to \mathcal{M}_{2\times 2}(\mathbb{R})$ such that

$$T_3\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} -2a+b-2c-d & -a+b-c \\ 2a-b+2c+d & a+c+d \end{pmatrix} \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{R}).$$

Step 1: Take the standard basis $C = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ for $\mathcal{M}_{2\times 2}(\mathbb{R})$.

Then

$$\boldsymbol{A} = [T_3]_C = \begin{pmatrix} -2 & 1 & -2 & -1 \\ -1 & 1 & -1 & 0 \\ 2 & -1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Step 2: The characteristic polynomial is

$$c_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} x+2 & -1 & 2 & 1 \\ 1 & x-1 & 1 & 0 \\ -2 & 1 & x-2 & -1 \\ -1 & 0 & -1 & x-1 \end{vmatrix} = x^2(x-1)^2.$$

Thus 0 and 1 are the eigenvalues.

Step 3: To find a basis for $E_0(T_3)$:

$$
\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in E_0(T_3) \quad &\Leftrightarrow \quad (T_3 - 0I_{\mathcal{M}_{2\times 2}(\mathbb{R})})\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\
&\Leftrightarrow \quad [T_3]_C \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right]_C = \left[\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\right]_C \\
&\Leftrightarrow \quad \begin{pmatrix} -2 & 1 & -2 & -1 \\ -1 & 1 & -1 & 0 \\ 2 & -1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = s \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ -1 \\ 0 \\ 1 \end{pmatrix} \quad \text{for } s, t \in \mathbb{R}.
\end{aligned}
$$

Thus $B_0 = \left\{ \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix} \right\}$ is a basis for $E_0(T_3)$.

Similarly, $B_1 = \left\{ \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$ is a basis for $E_1(T_3)$.

Step 4: T_3 is diagonalizable. Let $B = \left\{ \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$.

We have

$$
\boldsymbol{D} = [T_3]_B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Let $\boldsymbol{P} = [I_{\mathcal{M}_{2\times 2}(\mathbb{R})}]_{C,B} = \begin{pmatrix} -1 & -1 & -1 & 0 \\ 0 & -1 & -1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

Remark 11.1.14 In Algorithm 11.1.12, we need to factorize the characteristic polynomial into linear factors. Sometimes, over a certain field $\mathbb{F}$, say $\mathbb{F} = \mathbb{R}$, we may not be able to factorize some polynomials in this manner. Luckily, we can always find a bigger field that contains $\mathbb{F}$ and at the same time, all polynomials over this bigger field can be factorized into linear factors. In particular, the field $\mathbb{C}$ contains $\mathbb{R}$ and all polynomials over $\mathbb{C}$ can be factorized into linear factors.

Example 11.1.15 In Example 11.1.13.1, the characteristic polynomial cannot be factorized into linear factors over $\mathbb{R}$. We can extend the linear operator T_1 on $\mathbb{R}^2$ to a linear operator on

$\mathbb{C}^2$, i.e.

$$T_1((x, y)) = (y, -x) \quad \text{for } (x, y) \in \mathbb{C}^2.$$

Using the standard basis $C = \{(1, 0), (0, 1)\}$ for $\mathbb{C}^2$, we get the same matrix $\boldsymbol{A} = [T_1]_C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. But over $\mathbb{C}$,

$$c_{\boldsymbol{A}}(x) = 1 + x^2 = (x - \mathrm{i})(x + \mathrm{i})$$

and hence i and $-\mathrm{i}$ are the eigenvalues of $\boldsymbol{A}$. Following Algorithm 11.1.12, we find a new basis $B = \{(1, \mathrm{i}), (1, -\mathrm{i})\}$ for $\mathbb{C}^2$ such that $\boldsymbol{D} = [T_1]_B = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}$.

Let $\boldsymbol{P} = [I_{\mathbb{C}^2}]_{C,B} = \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}$.

Section 11.2 Triangular Canonical Forms

Discussion 11.2.1 Although not all square matrices are diagonalizable, we can still reduce them into simpler form provided that the field used is big enough (see Remark 11.1.14). In this section, we shall see that if the characteristic polynomial of a square matrix can be factorized into linear factors, then the matrix is similar to an upper triangular matrix. In particular, every complex square matrix is similar to an upper triangular matrix.

Lemma 11.2.2 Suppose $\boldsymbol{A}$ is an $r \times m$ matrix, $\boldsymbol{B}$ is an $r \times n$ matrix, $\boldsymbol{C}$ is an $s \times m$ matrix, $\boldsymbol{D}$ is an $s \times n$ matrix, $\boldsymbol{E}$ is an $m \times t$ matrix, $\boldsymbol{F}$ is an $m \times u$ matrix, $\boldsymbol{G}$ is an $n \times t$ matrix, $\boldsymbol{H}$ is an $n \times u$ matrix. Then

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \boldsymbol{AE} + \boldsymbol{BG} & \boldsymbol{AF} + \boldsymbol{BH} \\ \boldsymbol{CE} + \boldsymbol{DG} & \boldsymbol{CF} + \boldsymbol{DH} \end{pmatrix}.$$

(To do such a matrix multiplication, you need to make sure that the sub-matrices can be multiplied with each other.)

Proof The lemma can be shown easily by applying the results in Question 2.23.

Theorem 11.2.3 (Triangular Canonical Forms) Let $\mathbb{F}$ be a field.

1. Let $\boldsymbol{A} \in \mathcal{M}_{n \times n}(\mathbb{F})$. If the characteristic polynomial of $\boldsymbol{A}$ can be factorized into linear factors over $\mathbb{F}$, then there exists an invertible matrix $\boldsymbol{P} \in \mathcal{M}_{n \times n}(\mathbb{F})$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is an upper triangular matrix.

2. Let T be a linear operator on a finite dimensional vector space V over $\mathbb{F}$ where $\dim(V) \geq 1$. If the characteristic polynomial of T can be factorized into linear factors over $\mathbb{F}$, then there exists an ordered basis B for V such that $[T]_B$ is an upper triangular matrix.

Proof We prove Part 1 by induction on n. (In the following, vectors in $\mathbb{F}^n$ are written as column vectors.)

Since 1×1 matrices are upper triangular matrices, the result is true for $n = 1$.

Assume that if the characteristic polynomial of a matrix $\boldsymbol{B} \in \mathcal{M}_{k\times k}(\mathbb{F})$ can be factorized into linear factors over $\mathbb{F}$, then there exists an invertible matrix $\boldsymbol{Q} \in \mathcal{M}_{k\times k}(\mathbb{F})$ such that $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}$ is an upper triangular matrix.

Now, let $\boldsymbol{A} \in \mathcal{M}_{(k+1)\times(k+1)}(\mathbb{F})$ such that $c_{\boldsymbol{A}}(x)$ can be factorized into linear factors over $\mathbb{F}$. Let E be the standard basis for $\mathbb{F}^{k+1}$. As $c_{\boldsymbol{A}}(x)$ has at least one linear factor $x - \lambda$ for some $\lambda \in \mathbb{F}$, $\boldsymbol{A}$ has at least one eigenvalue λ.

Let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{A}$ associated with λ. Extend $\{\boldsymbol{v}\}$ to a basis $D = \{\boldsymbol{v}, \boldsymbol{u_1}, \dots, \boldsymbol{u_k}\}$ for $\mathbb{F}^{k+1}$. Let $\boldsymbol{R} = [I_{\mathbb{F}^{k+1}}]_{E,D} = \begin{pmatrix} \boldsymbol{v} & \boldsymbol{u_1} & \cdots & \boldsymbol{u_k} \end{pmatrix}$. Then

$$\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R} = [I_{\mathbb{F}^{k+1}}]_{D,E}\,[L_{\boldsymbol{A}}]_E\,[I_{\mathbb{F}^{k+1}}]_{E,D} = [L_{\boldsymbol{A}}]_D = \begin{pmatrix} \lambda & b_1 & \cdots & b_k \\ 0 & & & \\ \vdots & & \boldsymbol{B} & \\ 0 & & & \end{pmatrix}$$

for some $b_1, \dots, b_k \in \mathbb{F}$ and $\boldsymbol{B} \in \mathcal{M}_{k\times k}(\mathbb{F})$. As $c_{\boldsymbol{A}}(x) = c_{\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}}(x) = (x-\lambda)c_{\boldsymbol{B}}(x)$, $c_{\boldsymbol{B}}(x)$ can also be factorized into linear factors over $\mathbb{F}$. By the inductive assumption, there exists an invertible matrix $\boldsymbol{Q} \in \mathcal{M}_{k\times k}(\mathbb{F})$ such that $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}$ is an upper triangular matrix.

Let $\boldsymbol{P} = \boldsymbol{R}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{Q} & \\ 0 & & & \end{pmatrix}$. Then

$$\begin{aligned}
\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{Q} & \\ 0 & & & \end{pmatrix}^{-1} \boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{Q} & \\ 0 & & & \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{Q}^{-1} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} \lambda & b_1 & \cdots & b_k \\ 0 & & & \\ \vdots & & \boldsymbol{B} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{Q} & \\ 0 & & & \end{pmatrix} \\
&= \begin{pmatrix} \lambda & (b_1 \ \cdots \ b_k)\boldsymbol{Q} \\ 0 & \\ \vdots & \boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} \\ 0 & \end{pmatrix}
\end{aligned}$$

which is an upper triangular matrix. (Use Lemma 11.2.2 for the block matrix multiplications preformed above.)

So by the mathematical induction, Part 1 of the theorem is true for all positive integer n.

For Part 2, we apply the result of Part 1 to $\boldsymbol{A} = [T]_C$, where C is an ordered basis for V. There exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}[T]_C\boldsymbol{P}$ is an upper triangular matrix. Then by Theorem 9.3.10, there exists an ordered basis B such that $[T]_B = \boldsymbol{D}$ is an upper triangular matrix.

(See Question 11.23 for an alternative proof of Part 2.)

Example 11.2.4 Let $\boldsymbol{A} = \begin{pmatrix} 8 & -6 & -2 & -2 \\ 4 & 3 & 1 & 1 \\ 8 & 6 & 10 & 2 \\ -4 & 1 & 3 & 11 \end{pmatrix}$ be a real matrix. Note that $c_{\boldsymbol{A}}(x) = (x-8)^4$.

The matrix $\boldsymbol{A}$ has an eigenvector $(0,0,1,-1)^{\mathrm{T}}$.

Extend $\{(0,0,1,-1)^{\mathrm{T}}\}$ to a basis $\{(0,0,1,-1)^{\mathrm{T}},\ (1,0,0,0)^{\mathrm{T}},\ (0,1,0,0)^{\mathrm{T}},\ (0,0,0,1)^{\mathrm{T}}\}$ for $\mathbb{R}^4$.

Let $\boldsymbol{S} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$. Then

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \begin{pmatrix} 8 & 8 & 6 & 2 \\ 0 & 8 & -6 & -2 \\ 0 & 4 & 3 & 1 \\ 0 & 4 & 7 & 13 \end{pmatrix}.$$

The matrix $\boldsymbol{B} = \begin{pmatrix} 8 & -6 & -2 \\ 4 & 3 & 1 \\ 4 & 7 & 13 \end{pmatrix}$ has an eigenvector $(2,1,-3)^{\mathrm{T}}$.

Extend $\{(2,1,-3)^{\mathrm{T}}\}$ to a basis $\{(2,1,-3)^{\mathrm{T}},\ (0,1,0)^{\mathrm{T}},\ (0,0,1)^{\mathrm{T}}\}$ for $\mathbb{R}^3$. Let $\boldsymbol{R} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$.

Then

$$\boldsymbol{R}^{-1}\boldsymbol{B}\boldsymbol{R} = \begin{pmatrix} 8 & -3 & -1 \\ 0 & 6 & 2 \\ 0 & -2 & 10 \end{pmatrix}.$$

The matrix $\boldsymbol{C} = \begin{pmatrix} 6 & 2 \\ -2 & 10 \end{pmatrix}$ has an eigenvector $(1,1)^{\mathrm{T}}$.

Extend $\{(1,1)^{\mathrm{T}}\}$ to a basis $\{(1,1)^{\mathrm{T}},\ (0,1)^{\mathrm{T}}\}$ for $\mathbb{R}^2$. Let $\boldsymbol{Q} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Then

$$\boldsymbol{Q}^{-1}\boldsymbol{C}\boldsymbol{Q} = \begin{pmatrix} 8 & 2 \\ 0 & 8 \end{pmatrix}.$$

Now, let us trace what we have done backward following the proof of Theorem 11.2.3.

Let $\boldsymbol{U} = \boldsymbol{R}\left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$\boldsymbol{Q}$}} \\ 0 & & \end{array}\right) = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -3 & 1 & 1 \end{pmatrix}$. Then

$$\boldsymbol{U}^{-1}\boldsymbol{B}\boldsymbol{U} = \left(\begin{array}{c|c} 8 & (-3\ \ -1)\boldsymbol{Q} \\ \hline 0 & \multirow{2}{*}{$\boldsymbol{Q}^{-1}\boldsymbol{C}\boldsymbol{Q}$} \\ 0 & \end{array}\right) = \begin{pmatrix} 8 & -4 & -1 \\ 0 & 8 & 2 \\ 0 & 0 & 8 \end{pmatrix}.$$

Let $\boldsymbol{P} = \boldsymbol{S}\left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & \boldsymbol{U} & \\ 0 & & & \end{array}\right) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -3 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & -3 & 1 & 1 \end{pmatrix}$. Then

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \left(\begin{array}{c|c} 8 & (8\ \ 6\ \ 2)\boldsymbol{U} \\ \hline 0 & \\ 0 & \boldsymbol{U}^{-1}\boldsymbol{B}\boldsymbol{U} \\ 0 & \end{array}\right) = \begin{pmatrix} 8 & 16 & 8 & 2 \\ 0 & 8 & -4 & -1 \\ 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

Remark 11.2.5 Theorem 11.2.3 is still true if we change "upper triangular matrix" to "lower triangular matrix".

Section 11.3 Invariant Subspaces

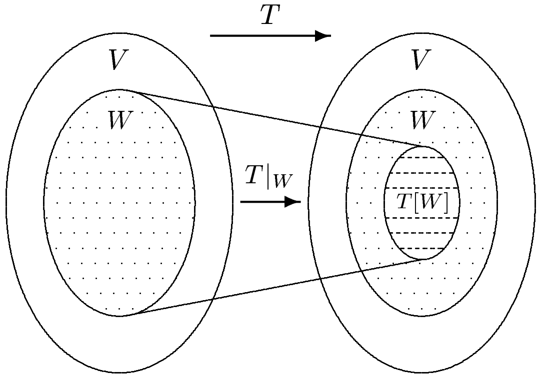
Definition 11.3.1 Let V be a vector space and $T : V \to V$ a linear operator. A subspace W of V is said to be T-*invariant* if $T(\boldsymbol{u})$ is contained in W for all $\boldsymbol{u} \in W$, i.e.

$$T[W] = \{T(\boldsymbol{u}) \mid \boldsymbol{u} \in W\} \subseteq W.$$

If W is a T-invariant subspace of V, the linear operator $T|_W : W \to W$ defined by

$$T|_W(\boldsymbol{u}) = T(\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \in W$$

is called the *restriction* of T on W.



Example 11.3.2

1. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator defined by $T((x, y, z)) = (y, -x, z)$ for $(x, y, z) \in \mathbb{R}^3$. Note that T is a rotation about the z-axis. (See Section 7.3.)

   (a) Let W_1 be the xy-plane in $\mathbb{R}^3$, i.e. $W_1 = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$. Since

   $$T((x, y, 0)) = (y, -x, 0) \in W_1 \quad \text{for all } (x, y, 0) \in W_1,$$

   W_1 is T-invariant. The restriction $T|_{W_1}$ of T on W_1 is equivalent a rotation defined on the xy-plane.

   (b) Let W_2 be the z-axis in $\mathbb{R}^3$, i.e. $W_2 = \{(0, 0, z) \mid z \in \mathbb{R}\}$. Since

   $$T((0, 0, z)) = (0, 0, z) \in W_2 \quad \text{for all } (0, 0, z) \in W_2,$$

   W_2 is T-invariant. The restriction $T|_{W_2}$ of T on W_2 is the identity operator.

   (c) Let W_3 be the yz-plane in $\mathbb{R}^3$, i.e. $W_3 = \{(0, y, z) \mid y, z \in \mathbb{R}\}$. It is not T-invariant. For example, $(0, 1, 1) \in W_3$ but $T((0, 1, 1)) = (1, 0, 1) \notin W_3$.

   Let $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ be the standard basis for $\mathbb{R}^3$, where $\boldsymbol{e_1} = (1, 0, 0)$, $\boldsymbol{e_2} = (0, 1, 0)$, $\boldsymbol{e_3} = (0, 0, 1)$, and let $C = \{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ and $D = \{\boldsymbol{e_3}\}$ which are bases for W_1 and W_2 respectively. Then

   $$\begin{aligned} [T]_E &= \Big([T(\boldsymbol{e_1})]_E \quad [T(\boldsymbol{e_2})]_E \quad [T(\boldsymbol{e_3})]_E \Big) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ [T|_{W_1}]_C &= \Big([T(\boldsymbol{e_1})]_C \quad [T(\boldsymbol{e_2})]_C \Big) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \\ [T|_{W_2}]_D &= \Big([T(\boldsymbol{e_3})]_D \Big) = (\,1\,) \end{aligned}$$

   and

   $$\begin{aligned} c_T(x) &= c_{[T]_E}(x) = (x-1)(x^2+1), \\ c_{T|_{W_1}}(x) &= c_{[T|_{W_1}]_C}(x) = x^2 + 1, \\ c_{T|_{W_2}}(x) &= c_{[T|_{W_2}]_D}(x) = x - 1. \end{aligned}$$

   Note that $\mathbb{R}^3 = W_1 \oplus W_2$, $[T]_E = \begin{pmatrix} [T|_{W_1}]_C & \boldsymbol{0} \\ \boldsymbol{0} & [T|_{W_2}]_D \end{pmatrix}$ and $c_T(x) = c_{T|_{W_1}}(x)\, c_{T|_{W_2}}(x)$. (See Discussion 11.3.12.)

2. Let T be a linear operator on a vector space V over a field $\mathbb{F}$. Suppose T has an eigenvector $\boldsymbol{v}$, i.e. v is a nonzero vector in V such that $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$ for some $\lambda \in \mathbb{F}$. Let $U = \operatorname{span}\{\boldsymbol{v}\}$. For any $\boldsymbol{u} \in U$, $\boldsymbol{u} = a\boldsymbol{v}$, for some $a \in \mathbb{F}$, and hence

   $$T(\boldsymbol{u}) = T(a\boldsymbol{v}) = aT(\boldsymbol{v}) = a\lambda\boldsymbol{v} \in U.$$

So U is T-invariant. Using $B = \{\boldsymbol{v}\}$ as a basis for U, we have $[T|_U]_B = \left([T(\boldsymbol{v})]_B \right) = (\,\lambda\,)$ and hence $c_{T|_U}(x) = c_{[T|_U]_B}(x) = x - \lambda$.

3. Let T be a linear operator on a vector space V over a field $\mathbb{F}$. Take any $\boldsymbol{u} \in V$. Define

$$W = \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u}),\, T^2(\boldsymbol{u}),\, T^3(\boldsymbol{u}),\, \dots\}.$$

For any $\boldsymbol{w} \in W$, $\boldsymbol{w} = a_0\boldsymbol{u} + a_1 T(\boldsymbol{u}) + \cdots + a_m T^m(\boldsymbol{u})$ for some $m \in \mathbb{N}$ and $a_0, a_1, \dots, a_m \in \mathbb{F}$. Then

$$\begin{aligned} T(\boldsymbol{w}) &= T(a_0\boldsymbol{u} + a_1T(\boldsymbol{u}) + \cdots + a_mT^m(\boldsymbol{u})) \\ &= a_0T(\boldsymbol{u}) + a_1T^2(\boldsymbol{u}) + \cdots + a_mT^{m+1}(\boldsymbol{u}) \in W. \end{aligned}$$

So W is T-invariant. This subspace W of V is called the *T-cyclic subspace of V generated by $\boldsymbol{u}$*.

4. Let T be a linear operator on $\mathbb{R}^4$ defined by

$$T((a,b,c,d)) = (4a+2b+2c+2d,\ a+2b+c,\ -a-b,\ -3a-2b-2c-d) \quad \text{for } (a,b,c,d) \in \mathbb{R}^4.$$

Take $\boldsymbol{u} = (1,-2,0,0)$. Then

$$\begin{aligned} T(\boldsymbol{u}) = T((1,-2,0,0)) &= (0,-3,1,1), \\ T^2(\boldsymbol{u}) = T(T(\boldsymbol{u})) = T((0,-3,1,1)) &= (-2,-5,3,3) \\ &= -2(1,-2,0,0) + 3(0,-3,1,1) \\ &= -2\boldsymbol{u} + 3T(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u})\}. \end{aligned}$$

Note that $T^3(\boldsymbol{u}) = T(T^2(\boldsymbol{u})) \in \operatorname{span}\{T(\boldsymbol{u}),\, T^2(\boldsymbol{u})\} \subseteq \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u})\}$. Repeating the process, we can show that $T^m(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u})\}$ for all $m \geq 2$. Thus the T-cyclic subspace of V generated by $\boldsymbol{u}$ is

$$W = \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u}),\, T^2(\boldsymbol{u}),\, T^3(\boldsymbol{u}),\, \dots\} = \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u})\}.$$

By the discussion in Part 3, W is a T-invariant subspace. (See also Theorem 11.3.10 and Example 11.3.11.1.)

Proposition 11.3.3 Let S and T be linear operators on V. Suppose W is a subspace of V which is both S-invariant and T-invariant. Then

1. W is $(S \circ T)$-invariant and $(S \circ T)|_W = S|_W \circ T|_W$;

2. W is $(S + T)$-invariant and $(S + T)|_W = S|_W + T|_W$; and

3. for any scalar c, W is (cT)-invariant and $(cT)|_W = c(T|_W)$.

Proof The proof is left as exercise. See Question 11.14.

Discussion 11.3.4 Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$. Suppose W is a T-invariant subspace of V with $\dim(W) \geq 1$. Let $\dim(V) = n$ and $\dim(W) = m$.

Take an ordered basis $C = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}\}$ for W. For each $j = 1, 2, \dots, m$, since $T(\boldsymbol{v_j}) \in W$,

$$T|_W(\boldsymbol{v_j}) = T(\boldsymbol{v_j}) = a_{1j}\boldsymbol{v_1} + a_{2j}\boldsymbol{v_2} + \cdots + a_{mj}\boldsymbol{v_m}$$

for some $a_{1j}, a_{2j}, \dots, a_{mj} \in \mathbb{F}$. Note that

$$[T|_W]_C = \Big([T|_W(\boldsymbol{v_1})]_B \quad [T|_W(\boldsymbol{v_2})]_B \quad \cdots \quad [T|_W(\boldsymbol{v_m})]_B \Big) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix}.$$

By Theorem 8.5.17, we can extend C to an ordered basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_m}, \boldsymbol{v_{m+1}}, \dots, \boldsymbol{v_n}\}$ for V. For each $j = m+1, m+2, \dots, n$,

$$T(\boldsymbol{v_j}) = a_{1j}\boldsymbol{v_1} + a_{2j}\boldsymbol{v_2} + \cdots + a_{mj}\boldsymbol{v_m} + a_{m+1,j}\boldsymbol{v_{m+1}} + \cdots + a_{nj}\boldsymbol{v_n}$$

for some $a_{1j}, a_{2j}, \dots, a_{nj} \in \mathbb{F}$. Then

$$\begin{aligned} [T]_B &= \Big([T(\boldsymbol{v_1})]_B \quad [T(\boldsymbol{v_2})]_B \quad \cdots \quad [T(\boldsymbol{v_m})]_B \quad [T(\boldsymbol{v_{m+1}})]_B \quad \cdots \quad [T(\boldsymbol{v_n})]_B \Big) \\ &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & a_{1,m+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2m} & a_{2,m+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} & a_{m,m+1} & \cdots & a_{mn} \\ 0 & 0 & \cdots & 0 & a_{m+1,m+1} & \cdots & a_{m+1,n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,m+1} & \cdots & a_{nn} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{A_1} & \boldsymbol{A_2} \\ \boldsymbol{0} & \boldsymbol{A_3} \end{pmatrix} \end{aligned}$$

where $\boldsymbol{A_1} = [T|_W]_C$, $\boldsymbol{A_2}$ is an $m \times (n-m)$ matrix and $\boldsymbol{A_3}$ is an $(n-m) \times (n-m)$ matrix.

Example 11.3.5 Consider the linear operator T on $\mathbb{R}^4$ in Example 11.3.2.4. Let $W = \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ where $\boldsymbol{v_1} = (1, -2, 0, 0)$ and $\boldsymbol{v_2} = (0, -3, 1, 1)$. From Example 11.3.2.4, we know that W is a T-invariant subspace of $\mathbb{R}^4$.

It is easy to check that $C = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ is a basis for W with $T|_W(\boldsymbol{v_1}) = T(\boldsymbol{v_1}) = \boldsymbol{v_2}$ and $T|_W(\boldsymbol{v_2}) = T(\boldsymbol{v_2}) = -2\boldsymbol{v_1} + 3\boldsymbol{v_2}$. Hence

$$[T|_W]_C = \Big([T|_W(\boldsymbol{v_1})]_C \quad [T|_W(\boldsymbol{v_2})]_C \Big) = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}.$$

Extend C to a basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}\}$ for $\mathbb{R}^4$ where $\boldsymbol{v_3} = (0,0,1,0)$ and $\boldsymbol{v_4} = (0,0,0,1)$. Then

$$T(\boldsymbol{v_3}) = (2,1,0,-2) = 2\boldsymbol{v_1} - \tfrac{5}{3}\boldsymbol{v_2} + \tfrac{5}{3}\boldsymbol{v_3} - \tfrac{1}{3}\boldsymbol{v_4}$$

and

$$T(\boldsymbol{v_3}) = (2,0,0,-1) = 2\boldsymbol{v_1} - \tfrac{4}{3}\boldsymbol{v_2} + \tfrac{4}{3}\boldsymbol{v_3} + \tfrac{1}{3}\boldsymbol{v_4}.$$

So

$$[T]_B = \left(\begin{array}{cc|cc} 0 & -2 & 2 & 2 \\ 1 & 3 & -\frac{5}{3} & -\frac{4}{3} \\ \hline 0 & 0 & \frac{5}{3} & \frac{4}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{3} \end{array}\right).$$

Take the standard basis $E = \{(1,0,0,0),\ (0,1,0,0),\ (0,0,1,0),\ (0,0,0,1)\}$ for $\mathbb{R}^4$. Let

$$\boldsymbol{A} = [T]_E = \begin{pmatrix} 4 & 2 & 2 & 2 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -3 & -2 & -2 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{P} = [I_{\mathbb{R}^4}]_{E,B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & -3 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = [T]_B = \begin{pmatrix} 0 & -2 & 2 & 2 \\ 1 & 3 & -\frac{5}{3} & -\frac{4}{3} \\ 0 & 0 & \frac{5}{3} & \frac{4}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}.$

Lemma 11.3.6 Let $\boldsymbol{D}$ be a quare matrix such that $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{pmatrix}$ where both $\boldsymbol{A}$ and $\boldsymbol{C}$ are square matrices. Then $\det(\boldsymbol{D}) = \det(\boldsymbol{A})\ \det(\boldsymbol{C})$.

Proof See Question 10.19.

Theorem 11.3.7 Let T be a linear operator on a finite dimensional vector space V. Suppose W is a T-invariant subspace of V with $\dim(W) \geq 1$. Then the characteristic polynomial of T is divisible by the characteristic polynomial of $T|_W$.

Proof Using the notation in Discussion 11.3.4,

$$\begin{aligned} c_T(x) = c_{[T]_B}(x) &= \det(x\boldsymbol{I_n} - [T]_B) \\ &= \begin{vmatrix} x\boldsymbol{I_m} - \boldsymbol{A_1} & -\boldsymbol{A_2} \\ \boldsymbol{0} & x\boldsymbol{I_{n-m}} - \boldsymbol{A_3} \end{vmatrix} \\ &= \det(x\boldsymbol{I_m} - \boldsymbol{A_1})\ \det(x\boldsymbol{I_{n-m}} - \boldsymbol{A_3}) \qquad \text{(see Lemma 11.3.6)} \\ &= c_{\boldsymbol{A_1}}(x)\, c_{\boldsymbol{A_3}}(x). \end{aligned}$$

Since $\boldsymbol{A_1} = [T|_W]_C$, $c_{\boldsymbol{A_1}}(x) = c_{T|_W}(x)$. Thus $c_T(x)$ is divisible by $c_{T|_W}(x)$.

Example 11.3.8 In Example 11.3.5, $c_T(x) = (x-1)^3(x-2)$ and $c_{T|_W}(x) = (x-1)(x-2)$. (Check them.) It is obvious that $c_T(x)$ is divisible by $c_{T|_W}(x)$.

Discussion 11.3.9 Given a linear operator T on a finite dimensional vector space V, by Discussion 11.3.4, a T-invariant subspace of V can be helpful in finding a basis B such that $[T]_B$ is in a simpler form. By Example 11.3.2.3, for any $\boldsymbol{u} \in V$, the T-cyclic subspace of V generated by $\boldsymbol{u}$ is T-invariant. This is a very useful invariant subspace. Actually, the subspace W used in Example 11.3.5 is a T-cyclic subspace, see Example 11.3.2.4.

Theorem 11.3.10 Let T be a linear operator on a vector space V over a field $\mathbb{F}$. Take a nonzero vector $\boldsymbol{u} \in V$. Suppose the T-cyclic subspace $W = \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), T^2(\boldsymbol{u}), \dots\}$ generated by $\boldsymbol{u}$ is finite dimensional.

1. The dimension of W is equal to the smallest positive integer k such that $T^k(\boldsymbol{u})$ is a linear combination of $\boldsymbol{u}$, $T(\boldsymbol{u})$, $\dots$, $T^{k-1}(\boldsymbol{u})$.

2. Suppose $\dim(W) = k$.

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

   (b) If $T^k(\boldsymbol{u}) = a_0\boldsymbol{u} + a_1T(\boldsymbol{u}) + \cdots + a_{k-1}T^{k-1}(\boldsymbol{u})$ where $a_0, a_1, \cdots, a_{k-1} \in \mathbb{F}$, then $c_{T|_W}(x) = -a_0 - a_1x - \cdots - a_{k-1}x^{k-1} + x^k$.

Proof Let k be the smallest positive integer such that $T^k(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$. Assume $T^m(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$ for $m \geq 1$. Then

$$T^{m+1}(\boldsymbol{u}) = T(T^m(\boldsymbol{u})) \in \operatorname{span}\{T(\boldsymbol{u}), T^2(\boldsymbol{u}), \dots, T^k(\boldsymbol{u})\} \subseteq \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}.$$

By mathematical induction, we have shown that $T^n(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$ for all positive integer n. Thus $W = \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$.

We claim that $\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$ is linearly independent: Assume the opposite, i.e. there exists $c_0, c_1, \dots, c_{k-1} \in \mathbb{F}$ such that $c_0\boldsymbol{u} + c_1T(\boldsymbol{u}) + \cdots + c_{k-1}T^{k-1}(\boldsymbol{u}) = \boldsymbol{0}$ and not all c_i's are zero. Let $j = \max\{i \mid 0 \leq i \leq k-1 \text{ and } c_i \neq 0\}$, i.e. $c_0\boldsymbol{u} + c_1T(\boldsymbol{u}) + \cdots + c_jT^j(\boldsymbol{u}) = \boldsymbol{0}$ and $c_j \neq 0$. Since $\boldsymbol{u}$ is a nonzero vector, $j > 0$. So we can write

$$T^j(\boldsymbol{u}) = -c_j^{-1}c_0\boldsymbol{u} - c_j^{-1}c_1T(\boldsymbol{u}) - \cdots - c_j^{-1}c_{j-1}T^{j-1}(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{j-1}(\boldsymbol{u})\}.$$

As $j < k$, it contradicts our choice of k.

Since $\{\boldsymbol{u}, T(\boldsymbol{u}), \dots, T^{k-1}(\boldsymbol{u})\}$ is linearly independent and it spans W, it is a basis for W and $\dim(W) = k$.

Finally, use $B = \{\boldsymbol{u}, T(\boldsymbol{u}), \ldots, T^{k-1}(\boldsymbol{u})\}$ as an ordered basis for W. Then

$$[T|_W]_B = \Big([T(\boldsymbol{u})]_B \quad [T^2(\boldsymbol{u})]_B \quad \cdots \quad [T^k(\boldsymbol{u})]_B \Big) = \begin{pmatrix} 0 & & & & & a_0 \\ 1 & 0 & & 0 & & a_1 \\ & 1 & 0 & & & a_2 \\ & & \ddots & \ddots & & \vdots \\ & 0 & & \ddots & 0 & a_{k-2} \\ & & & & 1 & a_{k-1} \end{pmatrix}$$

and hence by Question 10.23,

$$\begin{aligned} c_{T|_W}(x) = \det(xI_W - T|_W) &= \begin{vmatrix} x & & & & & -a_0 \\ -1 & x & & 0 & & -a_1 \\ & -1 & x & & & -a_2 \\ & & \ddots & \ddots & & \vdots \\ & 0 & & \ddots & x & -a_{k-2} \\ & & & & -1 & x - a_{k-1} \end{vmatrix} \\ &= -a_0 - a_1 x - \cdots - a_{k-1}x^{k-1} + x^k. \end{aligned}$$

Example 11.3.11 Consider the linear operator T on $\mathbb{R}^4$ in Example 11.3.2.4.

1. Let $\boldsymbol{u} = (1, -2, 0, 0)$ and $W = \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u}), T^2(\boldsymbol{u}), \ldots\}$. From Example 11.3.2.4, we have
$$T(\boldsymbol{u}) \notin \operatorname{span}\{\boldsymbol{u}\} \quad \text{and} \quad T^2(\boldsymbol{u}) = -2\boldsymbol{u} + 3T(\boldsymbol{u}) \in \operatorname{span}\{\boldsymbol{u}, T(\boldsymbol{u})\}.$$
By Theorem 11.3.10, $\dim(W) = 2$, $\{\boldsymbol{u}, T(\boldsymbol{u})\}$ is a basis for W and
$$c_{T|_W}(x) = -(-2) - 3x + x^2 = (x-1)(x-2).$$
(See also Example 11.3.5 and Example 11.3.8.)

2. Let $\boldsymbol{v} = (0, 0, 0, 1)$ and $W' = \operatorname{span}\{\boldsymbol{v}, T(\boldsymbol{v}), T^2(\boldsymbol{v}), \ldots\}$. Then
$$\begin{aligned} T(\boldsymbol{v}) = T((0,0,0,1)) &= (2, 0, 0, -1) \notin \operatorname{span}\{\boldsymbol{v}\} \\ T^2(\boldsymbol{v}) = T((2,0,0,-1)) &= (6, 2, -2, -5) \notin \operatorname{span}\{\boldsymbol{v}, T(\boldsymbol{v})\} \\ T^3(\boldsymbol{v}) = T((6,2,-2,-5)) &= (14, 8, -8, -13) \\ &= 2(0,0,0,1) - 5(2,0,0,-1) + 4(6,2,-2,-5) \\ &= 2\boldsymbol{v} - 5T(\boldsymbol{v}) + 4T^2(\boldsymbol{v}) \in \operatorname{span}\{\boldsymbol{v}, T(\boldsymbol{v}), T^2(\boldsymbol{v})\}. \end{aligned}$$
By Theorem 11.3.10, $\dim(W') = 3$, $\{\boldsymbol{v}, T(\boldsymbol{v}), T^2(\boldsymbol{v})\}$ is a basis for W' and
$$c_{T|_{W'}}(x) = -2 - (-5)x - 4x^2 + x^3 = (x-1)^2(x-2).$$

Discussion 11.3.12 Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$. Suppose

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

where W_1, W_2, ..., W_k are T-invariant subspaces of V with $\dim(W_t) = n_t \geq 1$ for $t = 1, 2, \ldots k$.

For each t, let $C_t = \{\boldsymbol{v_1^{(t)}}, \boldsymbol{v_2^{(t)}}, \ldots, \boldsymbol{v_{n_t}^{(t)}}\}$ be an ordered basis for W_t. As W_t is T-invariant, for $j = 1, 2, \ldots, n_t$, $T(\boldsymbol{v_j^{(t)}}) \in W_t$ and hence

$$T|_{W_t}(\boldsymbol{v_j^{(t)}}) = T(\boldsymbol{v_j^{(t)}}) = a_{1j}^{(t)}\boldsymbol{v_1^{(t)}} + a_{2j}^{(t)}\boldsymbol{v_2^{(t)}} + \cdots + a_{n_tj}^{(t)}\boldsymbol{v_{n_t}^{(t)}}$$

for some $a_{1j}^{(t)}, a_{2j}^{(t)}, \ldots, a_{n_tj}^{(t)} \in \mathbb{F}$. Thus

$$[T|_{W_t}]_{C_t} = \Big([T|_{W_t}(\boldsymbol{v_1^{(t)}})]_{C_t} \quad [T|_{W_t}(\boldsymbol{v_2^{(t)}})]_{C_t} \quad \cdots \quad [T|_{W_t}(\boldsymbol{v_{n_t}^{(t)}})]_{C_t} \Big) = \begin{pmatrix} a_{11}^{(t)} & a_{12}^{(t)} & \cdots & a_{1\,n_t}^{(t)} \\ a_{21}^{(t)} & a_{22}^{(t)} & \cdots & a_{2\,n_t}^{(t)} \\ \vdots & \vdots & & \vdots \\ a_{n_t1}^{(t)} & a_{n_t2}^{(t)} & \cdots & a_{n_tn_t}^{(t)} \end{pmatrix}.$$

By Theorem 8.6.7.1, we know that the set

$$\begin{aligned} B &= C_1 \cup C_2 \cup \cdots \cup C_k \\ &= \{\boldsymbol{v_1^{(1)}}, \boldsymbol{v_2^{(1)}}, \ldots, \boldsymbol{v_{n_1}^{(1)}}, \boldsymbol{v_1^{(2)}}, \boldsymbol{v_2^{(2)}}, \ldots, \boldsymbol{v_{n_2}^{(2)}}, \ldots\ldots, \boldsymbol{v_1^{(k)}}, \boldsymbol{v_2^{(k)}}, \ldots, \boldsymbol{v_{n_k}^{(k)}}\} \end{aligned}$$

is a basis for V.

Using B as an ordered basis with the order shown above,

$$\begin{aligned} [T]_B &= \Big([T(\boldsymbol{v_1^{(1)}})]_B \ \cdots \ [T(\boldsymbol{v_{n_1}^{(1)}})]_B \ [T(\boldsymbol{v_1^{(2)}})]_B \ \cdots \ [T(\boldsymbol{v_{n_2}^{(2)}})]_B \ \cdots \ [T(\boldsymbol{v_1^{(k)}})]_B \ \cdots \ [T(\boldsymbol{v_{n_k}^{(k)}})]_B \Big) \\ &= \begin{pmatrix} \boldsymbol{A_1} & \boldsymbol{0} & & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A_2} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{A_k} \end{pmatrix} \end{aligned}$$

where $\boldsymbol{A_t} = \left(a_{ij}^{(t)}\right)_{n_t \times n_t} = [T|_{W_t}]_{C_t}$ for $t = 1, 2, \ldots, k$. Furthermore,

$$c_T(x) = c_{\boldsymbol{A_1}}(x)\, c_{\boldsymbol{A_2}}(x) \cdots c_{\boldsymbol{A_k}}(x) = c_{T|_{W_1}}(x)\, c_{T|_{W_2}}(x) \cdots c_{T|_{W_k}}(x),$$

i.e. the characteristic polynomial of T is the product of the characteristic polynomials of $T|_{W_t}$ for $t = 1, 2, \ldots, k$.

Example 11.3.13 Consider the linear operator T on $\mathbb{R}^4$ in Example 11.3.2.4.

Let $W_1 = \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ and $W_2 = \operatorname{span}\{\boldsymbol{w_1}, \boldsymbol{w_2}\}$ where $\boldsymbol{v_1} = (1, -2, 0, 0)$, $\boldsymbol{v_2} = (0, -3, 1, 1)$, $\boldsymbol{w_1} = (0, 1, 1, -2)$ and $\boldsymbol{w_2} = (0, 0, 1, -1)$. We have the following observations:

1. $C_1 = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ and $C_2 = \{\boldsymbol{w_1}, \boldsymbol{w_2}\}$ are bases for W_1 and W_2 respectively.

2. $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{w_1}, \boldsymbol{w_2}\}$ is linearly independent. Hence $W_1 + W_2$ is a direct sum. (Why?) Also, since $\dim(\mathbb{R}^4) = 4 = \dim(W_1 \oplus W_2)$, we have $\mathbb{R}^4 = W_1 \oplus W_2$.

3. From Example 11.3.2.4, W_1 is T-invariant, $T(\boldsymbol{v_1}) = \boldsymbol{v_2}$ and $T(\boldsymbol{v_2}) = -2\boldsymbol{v_1} + 3\boldsymbol{v_2}$.

4. It can be shown that W_2 is also T-invariant, $T(\boldsymbol{w_1}) = 3\boldsymbol{w_1} - 4\boldsymbol{w_2}$ and $T(\boldsymbol{w_2}) = \boldsymbol{w_1} - \boldsymbol{w_2}$.

Then by Discussion 11.3.12, we have

$$[T|_{W_1}]_{C_1} = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}, \quad [T|_{W_2}]_{C_2} = \begin{pmatrix} 3 & 1 \\ -4 & -1 \end{pmatrix} \quad \text{and} \quad [T]_B = \left(\begin{array}{cc|cc} 0 & -2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ \hline 0 & 0 & 3 & 1 \\ 0 & 0 & -4 & -1 \end{array}\right).$$

Furthermore, $c_{T|_{W_1}}(x) = (x-1)(x-2)$, $c_{T|_{W_2}}(x) = (x-1)^2$ and $c_T(x) = (x-1)^3(x-2) = c_{T|_{W_1}}(x)\, c_{T|_{W_2}}(x)$.

Take the standard basis $E = \{(1,0,0,0),\ (0,1,0,0),\ (0,0,1,0),\ (0,0,0,1)\}$ for $\mathbb{R}^4$. Let

$$\boldsymbol{A} = [T]_E = \begin{pmatrix} 4 & 2 & 2 & 2 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -3 & -2 & -2 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{Q} = [I_{\mathbb{R}^4}]_{E,B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & -3 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -2 & -1 \end{pmatrix}.$$

Then $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = [T]_B = \begin{pmatrix} 0 & -2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & -4 & -1 \end{pmatrix}$.

Section 11.4 The Cayley-Hamilton Theorem

Notation 11.4.1 Let $\mathbb{F}$ be a field and let $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ where $a_0, a_1, \ldots, a_m \in \mathbb{F}$.

1. For a linear operator T on a vector space V over $\mathbb{F}$, we use $p(T)$ to denote the linear operator $a_0 I_V + a_1 T + \cdots + a_m T^m$ on V.

2. For an $n \times n$ matrix $\boldsymbol{A}$ over $\mathbb{F}$, we use $p(\boldsymbol{A})$ to denote the $n \times n$ matrix $a_0 \boldsymbol{I_n} + a_1 \boldsymbol{A} + \cdots + a_m \boldsymbol{A}^m$.

Lemma 11.4.2 Let T be a linear operator on a vector space V over a field $\mathbb{F}$ and $\boldsymbol{A}$ be an $n \times n$ matrix over $\mathbb{F}$. In the following, $p(x)$ and $q(x)$ are polynomials over $\mathbb{F}$.

1. Suppose V is finite dimensional where $\dim(V) = n \geq 1$. For any ordered basis B for V, $[p(T)]_B = p([T]_B)$.

2. If W is a T-invariant subspace of V, then W is also a $p(T)$-invariant subspace of V and $p(T)|_W = p(T|_W)$.

3. (a) If $r(x) = p(x) + q(x)$, then $r(T) = p(T) + q(T)$ and $r(\boldsymbol{A}) = p(\boldsymbol{A}) + q(\boldsymbol{A})$.

   (b) For $c \in \mathbb{F}$, if $s(x) = cp(x)$, then $s(T) = cp(T)$ and $s(\boldsymbol{A}) = cp(\boldsymbol{A})$.

   (c) If $u(x) = p(x)\,q(x)$, then $u(T) = p(T) \circ q(T) = q(T) \circ p(T)$ and $u(\boldsymbol{A}) = p(\boldsymbol{A})\,q(\boldsymbol{A}) = q(\boldsymbol{A})\,p(\boldsymbol{A})$.

Proof Let $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ where $a_0, a_1, \ldots, a_m \in \mathbb{F}$.

1. By Corollary 9.3.6 and Proposition 9.4.3,

$$\begin{aligned}
[p(T)]_B &= [a_0 I_V + a_1 T + a_2 T^2 + \cdots + a_m T^m]_B \\
&= a_0 \boldsymbol{I_n} + a_1 [T]_B + a_2([T]_B)^2 + \cdots + a_m([T]_B)^m \quad \text{(note that } [I_V]_B = \boldsymbol{I_n}) \\
&= p([T]_B).
\end{aligned}$$

2. By applying Proposition 11.3.3 repeatedly, W is a $p(T)$-invariant subspace of V and

$$\begin{aligned}
p(T)|_W &= (a_0 I_V + a_1 T + a_2 T^2 + \cdots + a_m T^m)|_W \\
&= a_0 I_W + a_1 T|_W + a_2 (T|_W)^2 + \cdots + a_m (T|_W)^m \quad \text{(note that } I_V|_W = I_W) \\
&= p(T|_W).
\end{aligned}$$

3. Parts (a) and (b) are obvious. Part (c) follows from the fact that $T^i \circ T^j = T^{i+j} = T^j \circ T^i$ and $\boldsymbol{A}^i \boldsymbol{A}^j = \boldsymbol{A}^{i+j} = \boldsymbol{A}^j \boldsymbol{A}^i$ for all i, j.

Example 11.4.3 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ be a real matrix.

Consider the linear operator $L_{\boldsymbol{A}} : \mathbb{R}^3 \to \mathbb{R}^3$ as defined in Example 9.1.4.1, i.e. $L_{\boldsymbol{A}}(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{R}^3$ where vectors in $\mathbb{R}^3$ are written as column vectors.

1. Find the matrix $p(\boldsymbol{A})$ and the linear operator $p(L_{\boldsymbol{A}})$ where $p(x) = -4 + 8x - 5x^2$.

   Solution We have

$$\begin{aligned}
p(\boldsymbol{A}) &= -4\boldsymbol{I_3} + 8\boldsymbol{A} - 5\boldsymbol{A}^2 \\
&= -4\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 8\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} - 5\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}^2 = \begin{pmatrix} 4 & 5 & 7 \\ -5 & -6 & -7 \\ -7 & -7 & -8 \end{pmatrix}
\end{aligned}$$

and for $\boldsymbol{u} = (x, y, z)^{\mathrm{T}} \in \mathbb{R}^3$,

$$p(L_{\boldsymbol{A}})(\boldsymbol{u}) = (-4I_{\mathbb{R}^3} + 8L_{\boldsymbol{A}} - 5(L_{\boldsymbol{A}})^2)(\boldsymbol{u}) = (-4\boldsymbol{I_3} + 8\boldsymbol{A} - 5\boldsymbol{A}^2)\boldsymbol{u}$$
$$= p(\boldsymbol{A})\,\boldsymbol{u} = \begin{pmatrix} 4 & 5 & 7 \\ -5 & -6 & -7 \\ -7 & -7 & -8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

2. Note that $c_{L_{\boldsymbol{A}}}(x) = c_{\boldsymbol{A}}(x) = (x-2)(x-1)^2$. Find the matrix $c_{\boldsymbol{A}}(\boldsymbol{A})$ and the linear operator $c_{L_{\boldsymbol{A}}}(L_{\boldsymbol{A}})$.

   Solution We have

   $$c_{\boldsymbol{A}}(\boldsymbol{A}) = (\boldsymbol{A} - 2\boldsymbol{I_3})(\boldsymbol{A} - \boldsymbol{I_3})^2 \quad \text{(by Lemma 11.4.2.3)}$$
   $$= \left[\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} - 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] \left[\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right]^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

   By Remark 11.1.8 and Example 9.2.4.3, $c_{L_{\boldsymbol{A}}}(x) = c_{\boldsymbol{A}}(x)$ and hence for $\boldsymbol{u} = (x, y, z)^{\mathrm{T}} \in \mathbb{R}^3$,

   $$c_{L_{\boldsymbol{A}}}(L_{\boldsymbol{A}})(\boldsymbol{x}) = c_{\boldsymbol{A}}(L_{\boldsymbol{A}})(\boldsymbol{u}) = (L_{\boldsymbol{A}} - 2I_{\mathbb{R}^3}) \circ (L_{\boldsymbol{A}} - I_{\mathbb{R}^3})^2(\boldsymbol{u}) \quad \text{(by Lemma 11.4.2.3)}$$
   $$= (\boldsymbol{A} - 2\boldsymbol{I_3})(\boldsymbol{A} - \boldsymbol{I_3})^2\boldsymbol{u}$$
   $$= c_{\boldsymbol{A}}(\boldsymbol{A})\,\boldsymbol{u} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

   Thus $c_{L_{\boldsymbol{A}}}(L_{\boldsymbol{A}}) = O_{\mathbb{R}^3}$, the zero operator on $\mathbb{R}^3$.

Theorem 11.4.4 (Cayley-Hamilton Theorem)

1. Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \geq 1$. Then $c_T(T) = O_V$ where O_V is the zero operator on V.

2. Let $\boldsymbol{A}$ be a square matrix. Then $c_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{0}$.

Proof Since Part 2 follows by applying Part 1 to the linear operator $T = L_{\boldsymbol{A}}$, we only need to prove Part 1. To prove that $c_T(T) = O_V$, we need to show that $c_T(T)(\boldsymbol{u}) = \boldsymbol{0}$ for all $\boldsymbol{u} \in V$.

Take any $\boldsymbol{u} \in V$. If $\boldsymbol{u} = \boldsymbol{0}$, then it is obvious that $c_T(T)(\boldsymbol{u}) = \boldsymbol{0}$. Suppose $\boldsymbol{u} \neq \boldsymbol{0}$. Let $W = \operatorname{span}\{\boldsymbol{u},\, T(\boldsymbol{u}),\, T^2(\boldsymbol{u}),\, \dots\}$. Note that W is a T-invariant subspace of V. Suppose $\dim(W) = k$. By Theorem 11.3.10, $B = \{\boldsymbol{u},\, T(\boldsymbol{u}),\, \dots, T^{k-1}(\boldsymbol{u})\}$ is a basis for W. Since $T^k(\boldsymbol{u}) \in W$, there exists $a_0, a_1, \dots, a_{k-1}$ such that

$$T^k(\boldsymbol{u}) = a_0\boldsymbol{u} + a_1T(\boldsymbol{u}) + \cdots + a_{k-1}T^{k-1}(\boldsymbol{u}). \tag{11.1}$$

By Theorem 11.3.10 again,

$$c_{T|_W}(x) = -a_0 - a_1x + \cdots - a_{k-1}x^{k-1} + x^k. \tag{11.2}$$

On the other hand, since W is a T-invariant subspace of V, by Theorem 11.3.7, the characteristic polynomial of T is divisible by the characteristic polynomial of $T|_W$, i.e.

$$c_T(x) = q(x)\, c_{T|_W}(x)$$

for some polynomial $q(x)$. Thus

$$\begin{aligned}
c_T(T)(\boldsymbol{u}) &= (q(T) \circ c_{T|_W}(T))(\boldsymbol{u}) && \text{(by Lemma 11.4.2.3)} \\
&= q(T)(c_{T|_W}(T)(\boldsymbol{u})) \\
&= q(T)((-a_0 I_V - a_1 T - \cdots - a_{k-1}T^{k-1} + T^k)(\boldsymbol{u})) && \text{(by (11.2))} \\
&= q(T)(-a_0\boldsymbol{u} - a_1 T(\boldsymbol{u}) - \cdots - a_{k-1}T^{k-1}(\boldsymbol{u}) + T^k(\boldsymbol{u})) \\
&= q(T)(\boldsymbol{0}) && \text{(by (11.1))} \\
&= \boldsymbol{0}.
\end{aligned}$$

Since $c_T(T)(\boldsymbol{u}) = 0$ for all $\boldsymbol{u} \in V$, $c_T(T) = O_V$.

Example 11.4.5 Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the linear operator defined by

$$T(a + bx + cx^2) = -2c + (a + 2b + c)x + (a + 3c)x^2 \quad \text{for } a + bx + cx^2 \in \mathcal{P}_2(\mathbb{R}).$$

Take the standard basis $B = \{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{R})$. Let

$$\boldsymbol{A} = [T]_B = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}.$$

Then

$$c_T(x) = c_{\boldsymbol{A}}(x) = \begin{vmatrix} x & 0 & 2 \\ -1 & x-2 & -1 \\ -1 & 0 & x-3 \end{vmatrix} = -4 + 8x - 5x^2 + x^3.$$

By the Cayley-Hamilton Theorem (Theorem 11.4.4), $c_T(T) = O_{\mathcal{P}_2(\mathbb{R})}$ and $c_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{0}$, i.e.

$$-4\, I_{\mathcal{P}_2(\mathbb{R})} + 8\,T - 5\,T^2 + T^3 = O_{\mathcal{P}_2(\mathbb{R})} \quad \text{and} \quad -4\boldsymbol{I} + 8\boldsymbol{A} - 5\boldsymbol{A}^2 + \boldsymbol{A}^3 = \boldsymbol{0}.$$

Section 11.5 Minimal Polynomials

Example 11.5.1 Consider the linear operator T on $\mathcal{P}_2(\mathbb{R})$ and the square matrix $\boldsymbol{A} = [T]_B$ defined in Example 11.4.5. By the Cayley-Hamilton Theorem (Theorem 11.4.4), $c_T(T) = O_{\mathcal{P}_2(\mathbb{R})}$ and $c_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{0}$.

Now, let $p(x) = 2 - 3x + x^2$. We have

$$\begin{aligned} p(\boldsymbol{A}) &= 2\boldsymbol{I_3} - 3\boldsymbol{A} + \boldsymbol{A}^2 \\ &= 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - 3\begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Thus $p(T) = O_{\mathcal{P}_2(\mathbb{R})}$. Note that $p(x)$ is a factor of $c_T(x)$, in fact, $c_T(x) = (x-2)p(x)$.

Definition 11.5.2 Let $\mathbb{F}$ be a field.

1. Let $p(x)$ be a polynomial of degree m over $\mathbb{F}$, i.e. $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ where $a_0, a_1, \dots, a_m \in \mathbb{F}$ and $a_m \neq 0$. If $a_m = 1$, then $p(x)$ is called a *monic polynomial*.

2. Let T be a linear operator on a finite dimensional vector space V over $\mathbb{F}$ where $\dim(V) \geq 1$. The *minimal polynomial* of T is the polynomial $m_T(x)$ over $\mathbb{F}$ such that

   (a) $m_T(x)$ is monic,
   (b) $m_T(T) = O_V$, and
   (c) if $p(x)$ is a nonzero polynomial over $\mathbb{F}$ such that $p(T) = O_V$, then the degree of $p(x)$ must be greater than or equal to the degree of $m_T(x)$.

   That is, the minimal polynomial $m_T(T)$ of T is the monic polynomial $p(x)$ of smallest degree such that $p(T) = O_V$.

   Since $c_T(T) = O_V$ by the Cayley-Hamilton Theorem, there exists at least one nonzero polynomial $p(x)$ such that $p(T) = O_V$. So $m_T(x)$ exists.

Remark 11.5.3 Let $\boldsymbol{A}$ be a square matrix. Similar to Definition 11.5.2.2, we can define the minimal polynomial of $\boldsymbol{A}$ accordingly, i.e. the *minimal polynomial* $m_{\boldsymbol{A}}(x)$ of $\boldsymbol{A}$ is the monic polynomial $p(x)$ of smallest degree such that $p(\boldsymbol{A}) = \boldsymbol{0}$.

Note that if T is a linear operator on a finite dimensional vector space V, where $\dim(V) \geq 1$, and B is an ordered basis for V, then $m_T(x) = m_{[T]_B}(x)$.

In this section, we mainly study the properties of minimal polynomials of linear operators. However, most results can be restated in terms of minimal polynomials of square matrices. See Question 11.33.

Example 11.5.4

1. For any finite dimensional vector space V with $\dim(V) \geq 1$, $m_{O_V}(x) = x$; and for the $n \times n$ zero matrix $\boldsymbol{0}$, $m_{\boldsymbol{0}}(x) = x$. (Why?)

2. In Example 11.4.3, $c_{L_{\boldsymbol{A}}}(x) = m_{L_{\boldsymbol{A}}}(x) = (x-2)(x-1)^2$.

3. In Example 11.4.5, $c_T(x) = -4 + 8x - 5x^2 + x^3 = (x-2)^2(x-1)$ while $m_T(x) = 2 - 3x + x^2 = (x-2)(x-1)$.

In general, it is not easy to find the minimal polynomial of a linear operator. In the mean time, we can only find them by trial-and-error. However, the next lemma will give us some idea to guess how the minimal polynomial looks like.

Lemma 11.5.5 Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$ where $\dim(V) \geq 1$.

1. Let $p(x)$ be a polynomial over $\mathbb{F}$. Then $p(T) = O_V$ if and only if $p(x)$ is divisible by the minimal polynomial of T.
2. If W is a T-invariant subspace of V with $\dim(W) \geq 1$, then the minimal polynomial of T is divisible by the minimal polynomial of $T|_W$.
3. Suppose λ is an eigenvalue of T such that $(x - \lambda)^r$ strictly divides $c_T(x)$, i.e. $c_T(x) = (x - \lambda)^r q(x)$ where $q(x)$ is a polynomial over $\mathbb{F}$ which is not divisible by $x - \lambda$. Then
$$m_T(x) = (x - \lambda)^s q_1(x)$$
where $1 \leq s \leq r$, $q_1(x)$ is a polynomial over $\mathbb{F}$ and $q_1(x)$ divides $q(x)$.

Proof

1. ($\Rightarrow$) Suppose $p(x)$ is a polynomial over $\mathbb{F}$ such that $p(T) = O_V$. By the Division Algorithm, we have polynomials $v(x)$ and $w(x)$ over $\mathbb{F}$ such that
$$p(x) = v(x)m_T(x) + w(x) \quad \text{and} \quad \deg(w(x)) < \deg(m_T(x)).$$
If $w(x) = 0$, then $p(x) = v(x)m_T(x)$ and hence $p(x)$ is divisible by $m_T(x)$.
Assume $w(x) \neq 0$. For any $\boldsymbol{u} \in V$,
$$\begin{aligned} \mathbf{0} = p(T)(\boldsymbol{u}) &= (v(T) \circ m_T(T) + w(T))(\boldsymbol{u}) \\ &= (v(T) \circ m_T(T))(\boldsymbol{u}) + w(T)(\boldsymbol{u}) \\ &= v(T)(m_T(T)(\boldsymbol{u})) + w(T)(\boldsymbol{u}) \\ &= v(T)(\mathbf{0}) + w(T)(\boldsymbol{u}) = \mathbf{0} + w(T)(\boldsymbol{u}) = w(T)(\boldsymbol{u}). \end{aligned}$$
Thus $w(T) = O_V$ and contracts that $m_T(x)$ is chosen to have the least degree.
($\Leftarrow$) Suppose $p(x) = t(x)m_T(x)$ for some polynomial $t(x)$ over $\mathbb{F}$. Then for all $\boldsymbol{u} \in V$,
$$p(T)(\boldsymbol{u}) = (t(T) \circ m_T(T))(\boldsymbol{u}) = t(T)(m_T(T)(\boldsymbol{u})) = t(T)(\mathbf{0}) = \mathbf{0}.$$
So $p(T) = O_V$.
2. Since $m_T(T) = O_V$, by Lemma 11.4.2.2, $m_T(T|_W) = m_T(T)|_W = O_V|_W = O_W$. By Part 1, $m_T(x)$ is divisible by $m_{T|_W}(x)$.

3. By the Cayley-Hamilton Theorem (Theorem 11.4.4), $c_T(T) = O_V$ and hence by Part 1, $c_T(x)$ is divisible by $m_T(x)$. Thus
$$m_T(x) = (x-\lambda)^s q_1(x)$$
where $0 \le s \le r$ and $q_1(x)$ divides $q(x)$. We only need to show that $s \ge 1$.

   Recall that λ is an eigenvalue of T. Take an eigenvector $\boldsymbol{v}$ associated with λ. Define $W_1 = \text{span}\{\boldsymbol{v}\}$. By Example 11.3.2.2, W_1 is a T-invariant subspace of V and $c_{T|_{W_1}}(x) = x - \lambda$. So
$$T|_{W_1} - \lambda I_{W_1} = O_{W_1}.$$
   The only monic polynomial with degree less than $x-\lambda$ is the constant polynomial $\mathit{1}(x) = 1$ and $\mathit{1}(T|_{W_1}) = I_{W_1} \ne O_{W_1}$. So $x - \lambda$ is the minimal polynomial of $T|_{W_1}$, i.e. $m_{T|_{W_1}}(x) = x - \lambda$. By Part 2, $m_T(x)$ is divisible by $m_{T|_{W_1}}(x) = x - \lambda$ and hence $s \ge 1$.

Example 11.5.6 Let $\boldsymbol{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ be a real matrix. Find the minimal polynomial of $\boldsymbol{A}$.

Solution The characteristic polynomial of $\boldsymbol{A}$ is $c_{\boldsymbol{A}}(x) = (x-2)^3$. By the matrix version of Lemma 11.5.5.3 (see Question 11.33), the minimal polynomial of $\boldsymbol{A}$ is $m_{\boldsymbol{A}}(x) = (x-2)^s$ where $s = 1$, 2 or 3. Note that s is the smallest positive integer such that $(\boldsymbol{A} - 2\boldsymbol{I})^s = \boldsymbol{0}$. As
$$\boldsymbol{A} - 2\boldsymbol{I} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \ne \boldsymbol{0} \qquad \text{and} \qquad (\boldsymbol{A} - 2\boldsymbol{I})^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}^2 = \boldsymbol{0},$$
$m_{\boldsymbol{A}}(x) = (x-2)^2$.

Theorem 11.5.7 Let T be a linear operator on a vector space V. Suppose W_1 and W_2 are T-invariant subspace of V. Then $W_1 + W_2$ is T-invariant and if W_1 and W_2 are finite dimensional with $\dim(W_1) \ge 1$ and $\dim(W_2) \ge 1$, $m_{T|_{W_1+W_2}}(x)$ is equal to the least common multiple of $m_{T|_{W_1}}(x)$ and $m_{T|_{W_2}}(x)$.

Proof The proof is left as an exercise. See Question 11.38.

Theorem 11.5.8 Let T be a linear operator on a finite dimensional vector space V where $\dim(V) \ge 1$. Suppose
$$c_T(x) = (x-\lambda_1)^{r_1}(x-\lambda_2)^{r_2}\cdots(x-\lambda_k)^{r_k}$$
where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T. Then
$$m_T(x) = (x-\lambda_1)^{s_1}(x-\lambda_2)^{s_2}\cdots(x-\lambda_k)^{s_k} \quad \text{where } 1 \le s_i \le r_i \text{ for all } i. \tag{11.3}$$
Define $K_{\lambda_i}(T) = \text{Ker}((T - \lambda_i I_V)^{s_i})$ for $i = 1, 2, \ldots, k$. Then
$$V = K_{\lambda_1}(T) \oplus K_{\lambda_2}(T) \oplus \cdots \oplus K_{\lambda_k}(T). \tag{11.4}$$

Furthermore, for each $i = 1, 2, \ldots, k$,

1. $E_{\lambda_i}(T) \subseteq K_{\lambda_i}(T)$,
2. $K_{\lambda_i}(T)$ is a T-invariant subspace of V.
3. $m_{T|_{K_{\lambda_i}(T)}}(x) = (x - \lambda_i)^{s_i}$,
4. $c_{T|_{K_{\lambda_i}(T)}}(x) = (x - \lambda_i)^{r_i}$,
5. $\dim(K_{\lambda_i}(T)) = r_i$.

Proof The formula of $m_T(x)$ in (11.3) follows easily from Lemma 11.5.5.3. The proof of (11.4) is left as exercise. (See Question 11.40.)

1. Take any $\boldsymbol{u} \in E_{\lambda_i}(T)$.
$$(T - \lambda_i I_V)(\boldsymbol{u}) = \boldsymbol{0} \quad \Rightarrow \quad (T - \lambda_i I_V)^{s_i}(\boldsymbol{u}) = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{u} \in K_{\lambda_i}(T).$$
So $E_{\lambda_i}(T) \subseteq K_{\lambda_i}(T)$.

2. Take any $\boldsymbol{u} \in K_{\lambda_i}(T)$. Then $(T - \lambda_i I_V)^{s_i}(\boldsymbol{u}) = \boldsymbol{0}$. Hence by Lemma 11.4.2.3,
$$\begin{aligned}(T - \lambda_i I_V)^{s_i}(T(\boldsymbol{u})) &= ((T - \lambda_i I_V)^{s_i} \circ T)(\boldsymbol{u}) \\ &= (T \circ (T - \lambda_i I_V)^{s_i})(\boldsymbol{u}) = T((T - \lambda_i I_V)^{s_i}(\boldsymbol{u})) = T(\boldsymbol{0}) = \boldsymbol{0}\end{aligned}$$
which implies $T(\boldsymbol{u}) \in K_{\lambda_i}(T)$. So $K_{\lambda_i}(T)$ is T-invariant.

3. Let $p(x) = (x - \lambda_i)^{s_i}$. For any $\boldsymbol{u} \in K_{\lambda_i}(T)$,
$$p(T|_{K_{\lambda_i}(T)})(\boldsymbol{u}) = p(T)|_{K_{\lambda_i}(T)}(\boldsymbol{u}) = p(T)(\boldsymbol{u}) = (T - \lambda_i I_V)^{s_i}(\boldsymbol{u}) = \boldsymbol{0}.$$
It means $p(T|_{K_{\lambda_i}(T)}) = O_{K_{\lambda_i}(T)}$ and hence by Lemma 11.5.5.1, $p(x)$ is divisible by $m_{T|_{K_{\lambda_i}(T)}}(x)$. Write $m_{T|_{K_{\lambda_i}(T)}}(x) = (x - \lambda_i)^{t_i}$, where $1 \le t_i \le s_i$, for each i.
Applying Theorem 11.5.7 to (11.4),
$$\begin{aligned}(x - \lambda_1)^{s_1}(x - \lambda_2)^{s_2} \cdots (x - \lambda_k)^{s_k} = m_T(x) &= \mathrm{lcm}\{(x - \lambda_1)^{t_1},\, (x - \lambda_2)^{t_2},\, \ldots,\, (x - \lambda_k)^{t_k}\} \\ &= (x - \lambda_1)^{t_1}(x - \lambda_2)^{t_2} \cdots (x - \lambda_k)^{t_k}.\end{aligned}$$
So $t_i = s_i$ for all i.

4. Since $m_{T|_{K_{\lambda_i}(T)}}(x) = (x - \lambda_i)^{s_i}$, by Lemma 11.5.5.3, $c_{T|_{K_{\lambda_i}(T)}}(x)$ cannot have other factor $x - \mu$ for $\mu \ne \lambda_i$. So $c_{T|_{K_{\lambda_i}(T)}}(x) = (x - \lambda_i)^{u_i}$ for $u_i \ge s_i$.
Applying Discussion 11.3.12 to (11.4),
$$(x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k} = c_T(x) = (x - \lambda_1)^{u_1}(x - \lambda_2)^{u_2} \cdots (x - \lambda_k)^{u_k}.$$
So $u_i = r_i$ for all i.

5. By Remark 11.1.8, $\dim(K_{\lambda_i}(T)) = \deg(c_{T|_{K_{\lambda_i}(T)}}(x)) = r_i$

Example 11.5.9

1. Consider the linear operator $L_{\boldsymbol{A}}$ in Example 11.3.5 and Example 11.4.3. We know that $m_{L_{\boldsymbol{A}}}(x) = c_{L_{\boldsymbol{A}}}(x) = (x-2)(x-1)^2$.

$$\begin{aligned}\begin{pmatrix}x\\y\\z\end{pmatrix} \in K_2(L_{\boldsymbol{A}}) \quad &\Leftrightarrow \quad \begin{pmatrix}1-2 & 0 & -1\\ 0 & 1-2 & 1\\ 1 & 1 & 2-2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\\ &\Leftrightarrow \quad \begin{pmatrix}x\\y\\z\end{pmatrix} = t\begin{pmatrix}-1\\1\\1\end{pmatrix} \quad \text{where } t \in \mathbb{R}\end{aligned}$$

and

$$\begin{aligned}\begin{pmatrix}x\\y\\z\end{pmatrix} \in K_1(L_{\boldsymbol{A}}) \quad &\Leftrightarrow \quad \begin{pmatrix}1-1 & 0 & -1\\ 0 & 1-1 & 1\\ 1 & 1 & 2-1\end{pmatrix}^2\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\\ &\Leftrightarrow \quad \begin{pmatrix}-1 & -1 & -1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\\ &\Leftrightarrow \quad \begin{pmatrix}x\\y\\z\end{pmatrix} = s\begin{pmatrix}-1\\0\\1\end{pmatrix} + t\begin{pmatrix}-1\\1\\0\end{pmatrix} \quad \text{where } s, t \in \mathbb{R}.\end{aligned}$$

So $K_2(L_{\boldsymbol{A}}) = \text{span}\{(-1,1,1)^{\text{T}}\}$ and $K_1(L_{\boldsymbol{A}}) = \text{span}\{(-1,0,1)^{\text{T}}, (-1,1,0)^{\text{T}}\}$. Note that $E_2(L_{\boldsymbol{A}}) = K_2(L_{\boldsymbol{A}})$ but $E_1(L_{\boldsymbol{A}}) \subsetneqq K_1(L_{\boldsymbol{A}})$.
With $B = \{(-1,1,1)^{\text{T}}, (-1,0,1)^{\text{T}}, (-1,1,0)^{\text{T}}\}$, $[L_{\boldsymbol{A}}]_B = \left(\begin{array}{c|cc}2 & 0 & 0\\ \hline 0 & 1 & 0\\ 0 & 1 & 1\end{array}\right)$. (See Theorem 11.5.8.)

2. Consider the linear operator T in Example 11.4.5. We know that $m_T(x) = (x-2)(x-1)$ and $c_T(x) = (x-2)^2(x-1)$.

$$\begin{aligned}a + bx + cx^2 \in K_2(T) \quad &\Leftrightarrow \quad \begin{pmatrix}0-2 & 0 & -2\\ 1 & 2-2 & 1\\ 1 & 0 & 3-2\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}\\ &\Leftrightarrow \quad \begin{pmatrix}a\\b\\c\end{pmatrix} = s\begin{pmatrix}-1\\0\\1\end{pmatrix} + t\begin{pmatrix}0\\1\\0\end{pmatrix} \quad \text{where } t, s \in \mathbb{R}\end{aligned}$$

and

$$a+bx+cx^2 \in K_1(T) \quad \Leftrightarrow \quad \begin{pmatrix} 0-1 & 0 & -2 \\ 1 & 2-1 & 1 \\ 1 & 0 & 3-1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix} = t \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \quad \text{where } t \in \mathbb{R}.$$

So $K_2(T) = \operatorname{span}\{-1+x^2,\, x\}$ and $K_1(T) = \operatorname{span}\{-2+x+x^2\}$. Note that $E_2(T) = K_2(T)$ and $E_1(T) = K_1(T)$.

With $B = \{-1+x^2,\, x,\, -2+x+x^2\}$, $[T]_B = \left(\begin{array}{cc|c} 2 & 0 & 0 \\ 0 & 2 & 0 \\ \hline 0 & 0 & 1 \end{array}\right)$. (See Theorem 11.5.8.)

Theorem 11.5.10 Let T be a linear operator on a finite dimensional vector space V, where $\dim(V) \geq 1$, such that $c_T(x) = (x-\lambda_1)^{r_1}(x-\lambda_2)^{r_2} \cdots (x-\lambda_k)^{r_k}$ where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T. (Note that $r_1 + r_2 + \cdots + r_k = \dim(V)$.)

The following are equivalent:

1. T is diagonalizable.
2. $m_T(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$.
3. $\dim(E_{\lambda_i}(T)) = r_i$ for $i = 1, 2, \ldots, k$.
4. $V = E_{\lambda_1}(T) \oplus E_{\lambda_2}(T) \oplus \cdots \oplus E_{\lambda_k}(T)$.

Proof

(1 ⇒ 2) Suppose T is diagonalizable. By Theorem 11.1.11, V has a basis B consisting of eigenvectors of T. For every $\boldsymbol{u} \in B$, $T(\boldsymbol{u}) = \lambda_i \boldsymbol{u}$ for some λ_i. Thus $(T - \lambda_i I_V)(\boldsymbol{u}) = \boldsymbol{0}$ and hence by Lemma 11.4.2.3,

$$\begin{aligned}
&((T-\lambda_1 I_V) \circ (T-\lambda_2 I_V) \circ \cdots \circ (T-\lambda_k I_V))(\boldsymbol{u}) \\
&= ((T-\lambda_1 I_V) \circ \cdots \circ (T-\lambda_{i-1} I_V) \circ (T-\lambda_{i+1} I_V) \circ \cdots \circ (T-\lambda_k I_V) \circ (T-\lambda_i I_V))(\boldsymbol{u}) \\
&= ((T-\lambda_1 I_V) \circ \cdots \circ (T-\lambda_{i-1} I_V) \circ (T-\lambda_{i+1} I_V) \circ \cdots \circ (T-\lambda_k I_V))((T-\lambda_i I_V)(\boldsymbol{u})) \\
&= ((T-\lambda_1 I_V) \circ \cdots \circ (T-\lambda_{i-1} I_V) \circ (T-\lambda_{i+1} I_V) \circ \cdots \circ (T-\lambda_k I_V))(\boldsymbol{0}) \\
&= \boldsymbol{0}.
\end{aligned}$$

This implies $(T-\lambda_1 I_V) \circ (T-\lambda_2 I_V) \circ \cdots \circ (T-\lambda_k I_V) = O_V$. By Lemma 11.5.5, $m_T(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$.

(2 ⇒ 3) Suppose $m_T(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$. For every i, $K_{\lambda_i}(T) = E_{\lambda_i}(T)$ and hence by Theorem 11.5.8, $\dim(E_{\lambda_i}(T)) = \dim(K_{\lambda_i}(T)) = r_i$.

(3 ⇒ 4) Suppose $\dim(E_{\lambda_i}(T)) = r_i$ for all i. For every i, by Theorem 11.5.8, $E_{\lambda_i}(T) = K_{\lambda_i}(T)$ and hence $V = K_{\lambda_1}(T) \oplus K_{\lambda_2}(T) \oplus \cdots \oplus K_{\lambda_k}(T) = E_{\lambda_1}(T) \oplus E_{\lambda_2}(T) \oplus \cdots \oplus E_{\lambda_k}(T)$.

(4 ⇒ 1) For each i, let B_{λ_i} be a basis for $E_{\lambda_i}(T)$ (every vector in B_{λ_i} is an eigenvector of T associated with the eigenvalue λ_i). As $V = E_{\lambda_1}(T) \oplus E_{\lambda_2}(T) \oplus \cdots \oplus E_{\lambda_k}(T)$, by Theorem 8.6.7.1, $B = B_{\lambda_1} \cup B_{\lambda_2} \cup \cdots \cup B_{\lambda_k}$ is a basis for V and every vector in B is an eigenvector of T. Hence by Theorem 11.1.11, T is diagonalizable.

Corollary 11.5.11 Let T be a linear operator on a finite dimensional vector space V and let W be a T-invariant subspace of V with $\dim(W) \geq 1$. If T is diagonalizable, then $T|_W$ is also diagonalizable.

Proof Since T is diagonalizable, by Theorem 11.5.10, $m_T(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_k)$ where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T. By Lemma 11.5.5.2, $m_T(x)$ is divisible by $m_{T|_W}(x)$ and hence $m_{T|_W}(x) = (x - \lambda_{i_1})(x - \lambda_{i_2}) \cdots (x - \lambda_{i_s})$ for some distinct eigenvalues $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_s} \in \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$. By Theorem 11.5.10, $T|_W$ is diagonalizable.

Section 11.6 Jordan Canonical Forms

Example 11.6.1 Let $T : \mathbb{C}^3 \to \mathbb{C}^3$ be the linear operator defined by

$$T((a, b, c)) = ((1 + \mathrm{i})a + b,\ \mathrm{i}b + c,\ -a - b - (1 - \mathrm{i})c) \quad \text{for } (a, b, c) \in \mathbb{C}^3.$$

Using the standard basis $E = \{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)\}$,

$$[T]_E = \begin{pmatrix} 1 + \mathrm{i} & 1 & 0 \\ 0 & \mathrm{i} & 1 \\ -1 & -1 & -(1 - \mathrm{i}) \end{pmatrix}.$$

Note that

$$c_T(x) = \begin{vmatrix} x - (1 + \mathrm{i}) & -1 & 0 \\ 0 & x - \mathrm{i} & -1 \\ 1 & 1 & x + (1 - \mathrm{i}) \end{vmatrix} = (x - \mathrm{i})^3$$

and hence $m_T(x) = (x - \mathrm{i})^s$ where $1 \leq s \leq 3$. Let $Q = T - \mathrm{i}I_V$, i.e.

$$Q((a, b, c)) = (a + b,\ c,\ -a - b - c) \quad \text{for } (a, b, c) \in \mathbb{C}^3.$$

Then $c_Q(x) = x^3$ and $m_Q(x) = x^s$.

It is easy to check that $Q \neq O_{\mathbb{C}^3}$ and $Q^2 \neq O_{\mathbb{C}^3}$ but $Q^3 = O_{\mathbb{C}^3}$. So $s = 3$, i.e. $m_Q(x) = x^3$ and $m_T(x) = (x - \mathrm{i})^3$.

Take $\boldsymbol{v} = (1, 0, 0)$. Then $Q(\boldsymbol{v}) = (1, 0, -1)$, $Q^2(\boldsymbol{v}) = Q((1, 0, -1)) = (1, -1, 0)$ and $Q^3(\boldsymbol{v}) = Q((1, -1, 0)) = (0, 0, 0)$. Using $B = \{Q^2(\boldsymbol{v}),\ Q(\boldsymbol{v}),\ \boldsymbol{v}\}$ as an ordered basis for $\mathbb{C}^3$,

$$
\begin{aligned}
[T]_B &= \Big([T(Q^2(\boldsymbol{v}))]_B \quad [T(Q(\boldsymbol{v}))]_B \quad [T(\boldsymbol{v})]_B \Big) \\
&= \Big([Q^3(\boldsymbol{v}) + \mathrm{i}Q^2(\boldsymbol{v})]_B \quad [Q^2(\boldsymbol{v}) + \mathrm{i}Q(\boldsymbol{v})]_B \quad [Q(\boldsymbol{v}) + \mathrm{i}\boldsymbol{v}]_B \Big) \quad \text{(because } T = Q + \mathrm{i}I_V) \\
&= \Big([\mathrm{i}Q^2(\boldsymbol{v})]_B \quad [Q^2(\boldsymbol{v}) + \mathrm{i}Q(\boldsymbol{v})]_B \quad [Q(\boldsymbol{v}) + \mathrm{i}\boldsymbol{v}]_B \Big) = \begin{pmatrix} \mathrm{i} & 1 & 0 \\ 0 & \mathrm{i} & 1 \\ 0 & 0 & \mathrm{i} \end{pmatrix}.
\end{aligned}
$$

Definition 11.6.2 Let λ be a scalar. The $t \times t$ matrix

$$
\boldsymbol{J_t}(\lambda) = \begin{pmatrix} \lambda & 1 & & & \mathbf{0} \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & \mathbf{0} & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}
$$

is called a *Jordan block* of order t associated with λ. For example,

$$
\boldsymbol{J_1}(\lambda) = \begin{pmatrix} \lambda \end{pmatrix}, \quad \boldsymbol{J_2}(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \boldsymbol{J_3}(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \boldsymbol{J_4}(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.
$$

Lemma 11.6.3 Given a Jordan block $\boldsymbol{J} = \boldsymbol{J_t}(\lambda)$, $c_{\boldsymbol{J}}(x) = m_{\boldsymbol{J}}(x) = (x - \lambda)^t$.

Proof Since $\boldsymbol{J}$ is a $t \times t$ triangular matrix with diagonal entries all equal to λ, $c_{\boldsymbol{J}}(x) = (x - \lambda)^t$. Then by Theorem 11.5.8, $m_{\boldsymbol{J}}(x) = (x - \lambda)^s$ for $1 \le s \le t$. By direct computations, we get $(\boldsymbol{J} - \lambda\boldsymbol{I})^r \neq \mathbf{0}$, for $r = 1, 2, \ldots, t-1$, while $(\boldsymbol{J} - \lambda\boldsymbol{I})^t = \mathbf{0}$. (Check it.) So $m_{\boldsymbol{J}}(x) = (x - \lambda)^t$.

Theorem 11.6.4 Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$ where $\dim(V) \ge 1$. Suppose the characteristic polynomial of T can be factorized into linear factors over $\mathbb{F}$. Then there exists an ordered basis B for V such that $[T]_B = \boldsymbol{J}$ with

$$
\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J_{t_1}}(\lambda_1) & & & & \mathbf{0} \\ & \boldsymbol{J_{t_2}}(\lambda_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & \mathbf{0} & & & \boldsymbol{J_{t_m}}(\lambda_m) \end{pmatrix} \tag{11.5}
$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvalues of T. (Note that $\lambda_1, \lambda_2, \ldots, \lambda_m$ are not necessarily distinct.)

Proof We omit the proof because it is too technical. (See Question 11.53 and Question 11.54.)

Remark 11.6.5 Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Suppose the characteristic polynomial of $\boldsymbol{A}$ can be factorized into linear factors over $\mathbb{F}$. Apply Theorem 11.6.4 to $T = L_{\boldsymbol{A}}$. We can find an invertible matrix $\boldsymbol{P} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$.

Definition 11.6.6 For a linear operator T of a finite dimensional vector space V, if there exists an ordered basis B for V such that $[T]_B = \boldsymbol{J}$ where $\boldsymbol{J}$ is a square matrix of the form stated in (11.5), we say that T *has a Jordan canonical form* $\boldsymbol{J}$ and $\boldsymbol{J}$ is *a Jordan canonical form for* T.

Similarly, for a square matrix $\boldsymbol{A}$, if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$, we say that $\boldsymbol{A}$ *has a Jordan canonical form* $\boldsymbol{J}$ and $\boldsymbol{J}$ is *a Jordan canonical form for* $\boldsymbol{A}$.

Example 11.6.7 Consider the matrix $\boldsymbol{A}$ and the linear operator $L_{\boldsymbol{A}}$ in Example 11.3.5 and Example 11.5.9.1. If we choose an ordered basis $B = \{(-1,1,1)^{\mathrm{T}},\ (-1,1,0)^{\mathrm{T}},\ (-1,0,1)\}$. Then

$$[L_{\boldsymbol{A}}]_B = \left(\begin{array}{c|cc} 2 & 0 & 0 \\ \hline 0 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right) = \begin{pmatrix} \boldsymbol{J_1}(2) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_2}(1) \end{pmatrix}.$$

which is a Jordan canonical form for $L_{\boldsymbol{A}}$.

Let $\boldsymbol{P} = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{J_1}(2) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_2}(1) \end{pmatrix}$ which is a Jordan canonical form for $\boldsymbol{A}$.

Remark 11.6.8 Jordan canonical forms are not unique. But two Jordan canonical forms for a linear operator (or a matrix) have the same collection of Jordan blocks but in different orders. (Actually, two matrices in Jordan canonical forms are similar if and only if they have the same collection of Jordan blocks.)

In Example 11.6.7, the following are all the possible Jordan canonical forms for $\boldsymbol{A}$:

$$\begin{pmatrix} \boldsymbol{J_1}(2) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_2}(1) \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{J_2}(1) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_1}(2) \end{pmatrix}.$$

Usually, we say that the Jordan canonical form is unique up to the ordering of the Jordan blocks.

Theorem 11.6.9 Suppose a linear operator T on a finite dimensional space V (respectively, a square matrix $\boldsymbol{A}$) has a Jordan canonical form

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J_{t_1}}(\lambda_1) & & & \boldsymbol{0} \\ & \boldsymbol{J_{t_2}}(\lambda_2) & & \\ & & \ddots & \\ & & & \ddots & \\ & \boldsymbol{0} & & & \boldsymbol{J_{t_m}}(\lambda_m) \end{pmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvalues of T (or $\boldsymbol{A}$). (Note that $\lambda_1, \lambda_2, \ldots, \lambda_m$ are not necessarily distinct.)

1. $c_T(x)$ (or $c_{\boldsymbol{A}}(x)$) $= (x-\lambda_1)^{t_1}(x-\lambda_2)^{t_2}\cdots(x-\lambda_m)^{t_m}$.

2. $m_T(x)$ (or $m_{\boldsymbol{A}}(x)$) is the least common multiple of $(x-\lambda_1)^{t_1}$, $(x-\lambda_2)^{t_2}$, $\dots$, $(x-\lambda_m)^{t_m}$, i.e. if $\lambda'_1, \lambda'_2, \dots, \lambda'_k$ are all the distinct eigenvalues of T (respectively, $\boldsymbol{A}$), then $m_T(x)$ (or $m_{\boldsymbol{A}}(x)$) $= (x-\lambda'_1)^{s_1}(x-\lambda'_2)^{s_2}\cdots(x-\lambda'_k)^{s_k}$ where for each i, s_i is the order of the largest Jordan block associated with λ'_i in the matrix $\boldsymbol{J}$.

3. For every eigenvalue λ of T (or $\boldsymbol{A}$), the dimension of the eigenspace $E_\lambda(T)$ (or $E_\lambda(\boldsymbol{A})$) is equal to the total number of Jordan blocks associated with λ in the matrix $\boldsymbol{J}$.

Proof In the following, we prove the linear operation part of the theorem. For the matrix part of the theorem, we only need to apply our arguments to the linear operator $L_{\boldsymbol{A}}$.

Suppose $B = \{\boldsymbol{v_1^{(1)}}, \boldsymbol{v_2^{(1)}}, \dots, \boldsymbol{v_{t_1}^{(1)}}, \boldsymbol{v_1^{(2)}}, \boldsymbol{v_2^{(2)}}, \dots, \boldsymbol{v_{t_2}^{(2)}}, \dots, \boldsymbol{v_1^{(m)}}, \boldsymbol{v_2^{(m)}}, \dots, \boldsymbol{v_{t_m}^{(m)}}\}$ is an ordered basis for V such that $[T]_B = \boldsymbol{J}$.

1. Since $x\boldsymbol{I}-\boldsymbol{J}$ is an upper triangular matrix,
$$\begin{aligned} c_T(x) &= \det(x\boldsymbol{I}-\boldsymbol{J}) \\ &= \text{the product of the diagonal entries of } x\boldsymbol{I}-\boldsymbol{J} \\ &= (x-\lambda_1)^{t_1}(x-\lambda_2)^{t_2}\cdots(x-\lambda_m)^{t_m}. \end{aligned}$$

2. Let $W_i = \operatorname{span}(C_i)$ where $C_i = \{\boldsymbol{v_1^{(i)}}, \boldsymbol{v_2^{(i)}}, \dots, \boldsymbol{v_{t_i}^{(i)}}\}$ for $i = 1, 2, \dots m$. Then
$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_m.$$
For $i = 1, 2, \dots m$, $[T|_{W_i}]_{C_i} = \boldsymbol{J_{t_i}}(\lambda_i)$ and by Lemma 11.6.3,
$$m_{T|_{W_i}}(x) = m_{\boldsymbol{J_{t_i}}(\lambda_i)}(x) = (x-\lambda_i)^{t_i}.$$
By Theorem 11.5.7, $m_T(x)$ is the least common multiple of $(x-\lambda_1)^{t_1}$, $(x-\lambda_2)^{t_2}$, $\dots$, $(x-\lambda_m)^{t_m}$.

3. Let $\boldsymbol{u} \in V$ with $[\boldsymbol{u}]_B = \begin{pmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_m} \end{pmatrix}$ where $\boldsymbol{x_i} \in \mathbb{F}^{t_i}$ is a column vector for $i = 1, 2, \dots, m$.

Then
$$\begin{aligned} & \boldsymbol{u} \in E_\lambda(T) \\ \Leftrightarrow\ & \begin{pmatrix} \boldsymbol{J_{t_1}}(\lambda_1)-\lambda\boldsymbol{I_{t_1}} & & & \boldsymbol{0} \\ & \boldsymbol{J_{t_2}}(\lambda_2)-\lambda\boldsymbol{I_{t_2}} & & \\ & & \ddots & \\ \boldsymbol{0} & & & \boldsymbol{J_{t_m}}(\lambda_m)-\lambda\boldsymbol{I_{t_m}} \end{pmatrix} \begin{pmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{pmatrix} \end{aligned}$$

$$\Leftrightarrow \quad (\boldsymbol{J_{t_i}}(\lambda_i) - \lambda \boldsymbol{I_{t_i}})\boldsymbol{x_i} = \boldsymbol{0} \quad \text{for } i = 1, 2, \ldots, m$$

$$\Leftrightarrow \quad \boldsymbol{x_i} = \begin{cases} \alpha_i(1, 0, \ldots, 0)^{\mathrm{T}}, \text{ for some } \alpha_i \in \mathbb{F}, & \text{if } \lambda_i = \lambda \\ (0, 0, \ldots, 0)^{\mathrm{T}} & \text{if } \lambda_i \neq \lambda \end{cases} \quad \text{for } i = 1, 2, \ldots, m.$$

Thus

$$E_\lambda(T) = \text{span}\{\, \boldsymbol{v_1^{(i)}} \mid i = 1, 2, \ldots, m \text{ and } \lambda_i = \lambda\}$$

and $\dim(E_\lambda(T))$ = the number of Jordan blocks associated with λ in $\boldsymbol{J}$.

Example 11.6.10

1. Suppose a complex square matrix $\boldsymbol{A}$ has a Jordan canonical form given by

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J_3}(\mathrm{i}) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_2}(4) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J_2}(4) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J_1}(4) \end{pmatrix} = \begin{pmatrix} \mathrm{i} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{i} & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

Then $c_{\boldsymbol{A}} = (x - \mathrm{i})^3(x - 4)^2(x - 4)^2(x - 4) = (x - \mathrm{i})^3(x - 4)^5$, $m_{\boldsymbol{A}} = (x - \mathrm{i})^3(x - 4)^2$, $\dim(E_{\mathrm{i}}(\boldsymbol{A})) = 1$ and $\dim(E_4(\boldsymbol{A})) = 3$.

2. Let $\boldsymbol{A}$ be a real square matrix such that

$$c_{\boldsymbol{A}}(x) = (x - 1)^3(x - 2)^2 \quad \text{and} \quad m_{\boldsymbol{A}}(x) = (x - 1)^2(x - 2).$$

Find a Jordan canonical form for $\boldsymbol{A}$.

Solution Let $\boldsymbol{J}$ be a Jordan canonical form for $\boldsymbol{A}$. Since $c_{\boldsymbol{A}}(x) = (x-1)^3(x-2)^2$, along the diagonal of $\boldsymbol{J}$, there are three 1's and two 2's. As $m_{\boldsymbol{A}}(x) = (x-1)^2(x-2)$, the largest Jordan block associated with 1 has order 2 and the largest Jordan block associated with 2 has order 1. So $\boldsymbol{J}$ must be similar to

$$\begin{pmatrix} \boldsymbol{J_2}(1) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J_1}(1) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J_1}(2) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J_1}(2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

Remark 11.6.11 The simplest form that a square matrix can be reduced to (or similar to) is a Jordan canonical form. However, in practice, we seldom use Jordan canonical forms because they are very sensitive to computational errors. For example, a Jordan canonical form for the matrix $\begin{pmatrix} 0 & 1 \\ \varepsilon & 0 \end{pmatrix}$ is $\begin{pmatrix} \sqrt{\varepsilon} & 0 \\ 0 & -\sqrt{\varepsilon} \end{pmatrix}$ if $\varepsilon \neq 0$; and it is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ if $\varepsilon = 0$. So small truncation errors during computation may end up with dramatic differences in Jordan canonical forms. So mathematicians working in numerical analysis prefer some other canonical forms though not so simple but more stable in computation.

Exercise 11

Question 11.1 to Question 11.6 are exercises for Section 11.1.

1. For each of the following linear operators T on V, (i) determine whether T is diagonalizable; and (ii) if T is diagonalizable, find an ordered basis B for V such that $[T]_B$ is a diagonal matrix.

   (a) $V = \mathbb{F}_2^2$ and $T((x, y)) = (x + y,\, y)$ for $(x, y) \in V$.

   (b) $V = \mathcal{P}_2(\mathbb{R})$ and $T(a + bx + cx^2) = (a + b) + (b - a)x + (c - a)x^2$ for $a + bx + cx^2 \in V$.

   (c) $V = \mathcal{P}_2(\mathbb{C})$ and $T(a + bx + cx^2) = (a + b) + (b - a)x + (c - a)x^2$ for $a + bx + cx^2 \in V$.

   (d) $V = \mathcal{P}_n(\mathbb{R})$ and $T(p(x)) = \dfrac{d(p(x))}{dx}$ for $p(x)V$.

   (e) $V = \mathcal{P}_n(\mathbb{R})$ and $T(p(x)) = \dfrac{d(xp(x))}{dx}$ for $p(x) \in V$.

2. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ be a real matrix.

   (a) Find an invertible 2×2 real matrix $\boldsymbol{P}$ so that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

   (b) Define a linear operator T on $\mathcal{M}_{2\times 2}(\mathbb{R})$ such that $T(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X}$ for $\boldsymbol{X} \in \mathcal{M}_{2\times 2}(\mathbb{R})$.

      (i) Let $C = \{\boldsymbol{E_{11}}, \boldsymbol{E_{21}}, \boldsymbol{E_{12}}, \boldsymbol{E_{22}}\}$. Find $[T]_C$.

      (ii) Find an ordered basis B for $\mathcal{M}_{2\times 2}(\mathbb{R})$ so that $[T]_B$ is a diagonal matrix.

3. Let S and T be linear operator on the same finite dimensional vector space. Suppose S is diagonalizable.

   (a) If every eigenvalue of S is an eigenvalue of T, must T be diagonalizable?

   (b) If every eigenvector of S is an eigenvector of T, must T be diagonalizable?

4. (a) Prove that two similar square matrices have the same characteristic polynomial.

(b) Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{F})$ where $\mathbb{F}$ is a field.

(i) Show that the following two $2n \times 2n$ matrices

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{AB} & \boldsymbol{0}_{n\times n} \\ \boldsymbol{B} & \boldsymbol{0}_{n\times n} \end{pmatrix} \quad \text{and} \quad \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} \\ \boldsymbol{B} & \boldsymbol{BA} \end{pmatrix}$$

are similar.

(ii) Hence, or otherwise, prove that $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same characteristic polynomial.

(c) Restate the result of Part (b)(ii) in terms of linear operators.

5. Let V be a finite dimensional vector space over a field $\mathbb{F}$ and T a linear operator on V such that $T^2 = I_V$.

(a) If $1 + 1 \neq 0$ in $\mathbb{F}$, prove that T is diagonalizable. (Hint: See Question 9.18(b)(i).)

(b) Give an example of T such that $T^2 = I_V$ but T is not diagonalizable.

6. Let T be a linear operator on a vector space V. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T. Show that $E_{\lambda_1}(T) + E_{\lambda_2}(T) + \cdots + E_{\lambda_k}(T)$ is a direct sum.

Question 11.7 to Question 11.11 are exercises for Section 11.2.

7. Let $\boldsymbol{A} = \begin{pmatrix} 3 & -3 & 4 \\ 1 & 1 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ be a real matrix.

Following the procedure of Example 11.2.4, find an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is an upper triangular matrix.

8. (a) Let $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & -1 \\ 1 & 0 & 3 \end{pmatrix}$ and $\boldsymbol{Q} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix}$ be real matrices.

Verify that $\boldsymbol{Q}^{-1}\boldsymbol{BQ}$ is an upper triangular matrix.

(b) Let $\boldsymbol{A} = \begin{pmatrix} 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{pmatrix}$ and $\boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ be real matrices.

Compute $\boldsymbol{R}^{-1}\boldsymbol{AR}$.

Hence, or otherwise, find an invertible 4×4 real matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is an upper triangular matrix.

9. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -2 & -2 \\ 2 & 1 & 0 & 0 \\ -1 & -2 & -1 & 0 \\ 3 & 2 & -2 & -3 \end{pmatrix}$ be a real matrix and $\boldsymbol{v_1} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}$, $\boldsymbol{v_2} = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$ vectors in $\mathbb{R}^4$.

   (a) Verify that $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are eigenvectors of $\boldsymbol{A}$.

   (b) Extend $\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ to an ordered basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{w_1}, \boldsymbol{w_2}\}$ for $\mathbb{R}^4$.

   (c) Using the vectors $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$ obtained in Part (b), let $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{v_1} & \boldsymbol{v_2} & \boldsymbol{w_1} & \boldsymbol{w_2} \end{pmatrix}$. Compute $\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}$.

   (d) Find an invertible real matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is an upper triangular matrix.

10. (a) Let $\boldsymbol{B} = (b_{ij})_{n\times n}$ and $\boldsymbol{C} = (c_{ij})_{n\times n}$ be upper triangular matrices . Show that $\boldsymbol{BC}$ is an upper triangular matrix. What are the diagonal entries of $\boldsymbol{BC}$?

    (b) Let $\boldsymbol{A}$ be an $n \times n$ complex matrix. Prove that if μ is an eigenvalue of $\boldsymbol{A}^2$, then $\sqrt{\mu}$ or $-\sqrt{\mu}$ is an eigenvalue of $\boldsymbol{A}$.

11. A linear operator T on a finite dimensional vector space V is called *triangularizable* if there exists an ordered basis B for V such that $[T]_B$ is a triangular matrix.

    (a) Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$. Prove that T is triangularizable if and only if the characteristic polynomial of T can be factorized into linear factors over $\mathbb{F}$.

    (b) For each of the following linear operators T on V, determine whether T is triangularizable.

       (i) $V = \mathbb{F}_2^2$ and $T((x, y)) = (x + y,\ y)$ for all $(x, y) \in V$.

       (ii) $V = \mathcal{P}_2(\mathbb{R})$ and $T(a+bx+cx^2) = (a+b)+(b-a)x+(c-a)x^2$ for all $a+bx+cx^2 \in V$.

       (iii) $V = \mathcal{M}_{n\times n}(\mathbb{C})$ and $T(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X}$ for $\boldsymbol{X} \in V$, where $\boldsymbol{A}$ is a complex $n \times n$ matrix.

Question 11.12 to Question 11.24 are exercises for Section 11.3.

12. For each of the following linear operators T on V and subspaces W of V, determine whether W is T-invariant.

    (a) $V = \mathbb{C}^3$, $W = \{(a,\ b,\ a + b) \mid a, b \in \mathbb{C}\}$ and $T((a, b, c)) = (b + \mathrm{i}c,\ c + \mathrm{i}a,\ a + \mathrm{i}b)$ for $(a, b, c) \in V$.

    (b) $V = \mathbb{C}^3$, $W = \mathrm{span}\{(1, 1, 1)\}$ and $T((a, b, c)) = (b+\mathrm{i}c,\ c+\mathrm{i}a,\ a+\mathrm{i}b)$ for $(a, b, c) \in V$.

    (c) $V = \mathbb{C}^3$, $W = \{(x, y, z) \in \mathbb{C}^3 \mid x+y+z = 0\}$ and $T((a, b, c)) = (b+\mathrm{i}c,\ c+\mathrm{i}a,\ a+\mathrm{i}b)$ for $(a, b, c) \in V$.

(d) $V = \mathcal{P}(\mathbb{R})$, $W = \mathcal{P}_3(\mathbb{R})$ and $T(p(x)) = \dfrac{dp(x)}{dx}$ for $p(x) \in V$.

(e) $V = \mathcal{P}(\mathbb{R})$, $W = \mathcal{P}_3(\mathbb{R})$ and $T(p(x)) = \displaystyle\int_0^x p(t)dt$ for $p(x) \in V$.

(f) $V = \mathcal{M}_{2\times 2}(\mathbb{R})$, $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\}$ and $T(\boldsymbol{X}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \boldsymbol{X}$ for $\boldsymbol{X} \in V$.

(g) $V = \mathcal{M}_{2\times 2}(\mathbb{R})$, $W = \{\boldsymbol{A} \in V \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\}$ and $T(\boldsymbol{X}) = -\boldsymbol{X}^{\mathrm{T}}$ for $\boldsymbol{X} \in V$.

13. Consider the linear operator T defined in Parts (a), (b), (c) of Question 11.12.

    (i) Compute $c_T(x)$.

    (ii) For each of Parts (a), (b), (c) of Question 11.12, if W is T-invariant, compute $c_{T|_W}(x)$ and verify that $c_T(x)$ is divisible by $c_{T|_W}(x)$.

14. Prove Proposition 11.3.3:

    Let S and T be linear operators on V. Suppose W is a subspace of V which is both S-invariant and T-invariant. Show that

    (a) W is $(S \circ T)$-invariant and $(S \circ T)|_W = S|_W \circ T|_W$;

    (b) W is $(S + T)$-invariant and $(S + T)|_W = S|_W + T|_W$; and

    (c) for any scalar c, W is (cT)-invariant and $(cT)|_W = c(T|_W)$.

15. Let V be a real vector space which has a basis $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}\}$. Suppose T is a linear operator on V such that

$$[T]_B = \begin{pmatrix} 1 & -1 & 1 & 1 \\ 2 & 1 & 2 & -1 \\ 1 & -1 & 3 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

    Let $W = \operatorname{span}\{\boldsymbol{v_4}, T(\boldsymbol{v_4}), T^2(\boldsymbol{v_4}), \dots\}$, the T-cyclic subspace generated by $\boldsymbol{v_4}$.

    (a) Compute $T(\boldsymbol{v_4})$ and $T^2(\boldsymbol{v_4})$.

    (b) What is the dimension of W?

    (c) Write down the characteristic polynomial of $T|_W$.

16. Let T be a linear operator on a vector space V. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are two eigenvectors of T associated with eigenvalues λ and μ, respectively, where $\lambda \neq \mu$. Let $W = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ and $B = \{\boldsymbol{w}, T(\boldsymbol{w})\}$ where $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$.

    (a) Show that W is a T-invariant subspace of V.

(b) Prove that B is a basis for W.

(c) Find $c_{T|_W}(x)$ and $[T|_W]_B$.

17. Let S be the shift operator on $\mathbb{R}^{\mathbb{N}}$, i.e. $S\left((a_n)_{n\in\mathbb{N}}\right) = (a_{n+1})_{n\in\mathbb{N}}$ for $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, and let

$$W = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid a_{n+3} = 2a_{n+2} + a_{n+1} - 2a_n \text{ for } n = 1, 2, 3, \dots\}.$$

(a) Is W an S-invariant subspace of $\mathbb{R}^{\mathbb{N}}$?

(b) Find $\dim(W)$.

(c) Take $\boldsymbol{b} = (b_n)_{n\in\mathbb{N}} \in W$ with $b_1 = 0$, $b_2 = 0$, $b_3 = 1$. Prove that $\{\boldsymbol{b},\ S(\boldsymbol{b}),\ S^2(\boldsymbol{b})\}$ is a basis for W.

(d) Find the characteristic polynomial of $S|_W$.

(e) Find a basis B for W such that $[S|_W]_B$ is a diagonal matrix.

18. Use the shift operator S in Question 11.17. Let

$$W' = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid a_{n+2} = 2a_{n+1} - a_n \text{ for } n = 1, 2, 3, \dots\}.$$

Show that $S|_{W'}$ is not diagonalizable.

19. Let $\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ where $\mathbb{F}$ is a field. Define T to be a linear operator on $\mathcal{M}_{n\times n}(\mathbb{F})$ such that $T(\boldsymbol{X}) = \boldsymbol{AX}$ for $\boldsymbol{X} \in \mathcal{M}_{n\times n}(\mathbb{F})$.

Let $\boldsymbol{E_{ij}}$, $i, j = 1, 2, \dots, n$, be the $n \times n$ matrices as defined in Example 8.4.6.7. For $q = 1, 2, \dots, n$, define $W_q = \operatorname{span}(B_q)$ where $B_q = \{\boldsymbol{E_{1q}},\ \boldsymbol{E_{2q}},\ \dots,\ \boldsymbol{E_{nq}}\}$.

(a) Is $\mathcal{M}_{n\times n}(\mathbb{F}) = W_1 \oplus W_2 \oplus \cdots \oplus W_n$?

(b) Prove that W_q is T-invariant. What is $[T|_{W_q}]_{B_q}$?

(c) Let $B = \{\boldsymbol{E_{11}},\ \boldsymbol{E_{21}},\ \dots,\ \boldsymbol{E_{n1}}, \boldsymbol{E_{12}},\ \boldsymbol{E_{22}},\ \dots,\ \boldsymbol{E_{n2}},\ \dots,\ \boldsymbol{E_{1n}},\ \boldsymbol{E_{2n}},\ \dots,\ \boldsymbol{E_{nn}}\}$. Write down $[T]_B$ and hence prove that $c_T(x) = [c_{\boldsymbol{A}}(x)]^n$.

(d) Prove that T is diagonalizable if and only if $\boldsymbol{A}$ is diagonalizable.

20. Let W_1 and W_2 be subspaces of a vector space V such that $V = W_1 \oplus W_2$ and let $P : V \to V$ be the projection on W_1 along W_2 defined in Question 9.6. For a linear operator T on V, prove that $T \circ P = P \circ T$ if and only if both W_1 and W_2 are T-invariant. (Hint: You need the results proved in Question 9.6.)

21. Let T be a linear operator on a vector space V and W a T-invariant subspace of V.

(a) For $\boldsymbol{u}, \boldsymbol{v} \in V$, show that if $W + \boldsymbol{u} = W + \boldsymbol{v}$, then $W + T(\boldsymbol{u}) = W + T(\boldsymbol{v})$.

(b) Define a mapping $T/W : V/W \to V/W$ such that

$$T/W(W + \boldsymbol{u}) = W + T(\boldsymbol{u}) \quad \text{for } W + \boldsymbol{u} \in V/W.$$

(By Part (a), T/W is well-define.)

(i) Show that T/W is a linear operator on V/W.

(ii) Show that $c_T(x) = c_{T|_W}(x)\, c_{T/W}(x)$. (Hint: Follow Discussion 11.3.4.)

(iii) Let $\mathcal{U}$ be a T/W-invariant subspace of V/W and $U = \{\boldsymbol{u} \in V \mid W + \boldsymbol{u} \in \mathcal{U}\}$. Show that U is a T-invariant subspace of V.

22. Let T be a linear operator on a finite dimensional vector space V where $\dim(V) = n$. Prove that T is triangularizable if and only if there exist T-invariant subspaces W_1, W_2, $\ldots$, W_n of V such that $W_1 \subseteq W_2 \subseteq \cdots \subseteq W_n = V$ and $\dim(W_j) = j$ for $j = 1, 2, \ldots, n$. (See Question 11.11 for the definition of "triangularizable".)

23. Reprove Theorem 11.2.3.2 without using the matrix arguments:

    Let T be a linear operator on a finite dimensional vector space V over $\mathbb{F}$ where $\dim(V) \geq 1$. If the characteristic polynomial of T can be factorized into linear factors over $\mathbb{F}$, by using results of Question 11.21 and Question 11.22, prove that T is triangularizable.

24. (a) Let S and T be linear operators on a vector space V such that $S \circ T = T \circ S$. Prove that $E_\lambda(T)$ is S-invariant where λ is an eigenvalue of T.

    (b) Let S and T be linear operators on a finite dimensional vector space V, where $\dim(V) \geq 1$, over a field $\mathbb{F}$ such that $S \circ T = T \circ S$. If the characteristic polynomials of S and T can both be factorized into linear factors over $\mathbb{F}$, prove that there exists an ordered basis B for V such that both $[S]_B$ and $[T]_B$ are upper triangular matrices.

    (c) Restate the result in Part (b) using square matrices.

Question 11.25 to Question 11.32 are exercises for Section 11.4.

25. Let T be a linear operator on a vector space V over a field $\mathbb{F}$ and $p(x)$ a polynomial over $\mathbb{F}$. Prove that $\mathrm{Ker}(p(T))$ and $\mathrm{R}(p(T))$ are T-invariant.

26. Let $\mathbb{F}$ be a field and $p(x)$, $q(x)$ two polynomials over $\mathbb{F}$. The *greatest common divisor* (or the *highest common factor*) of $p(x)$ and $q(x)$, denoted by $\gcd(p(x), q(x))$, is the monic polynomial of highest degree which divides both $p(x)$ and $q(x)$. The following is an algorithm called the *Euclidean Algorithm* which is used to find $\gcd(p(x), q(x))$:

    Assume $\deg(p(x)) \geq \deg(q(x))$.

    Step 1: Let $r_0(x) = p(x)$ and $r_1(x) = q(x)$. Set $t = 1$.

Step 2: Divide $r_{t-1}(x)$ by $r_t(x)$ to get the reminder $r_{t+1}(x)$, i.e. $r_{t+1}(x)$ is the polynomial over $\mathbb{F}$ satisfying

$$r_{t-1}(x) = q_t(x)\, r_t(x) + r_{t+1}(x) \quad \text{and} \quad \deg(r_{t+1}(x)) < \deg(r_t(x))$$

for some polynomial $q_t(x)$ over $\mathbb{F}$.

Step 3: If $r_{t+1}(x) \neq 0$, increase the value of t by 1 and goto Step 2.

Step 4: Now, $r_{t+1}(x) = 0$. Let c be the coefficient of the term of highest degree in $r_t(x)$. Then $\gcd(p(x),\, q(x)) = c^{-1} r_t(x)$.

(a) For each of the following cases, use the Euclidean Algorithm to find $\gcd(p(x),\, q(x))$.

(i) $\mathbb{F} = \mathbb{R}$, $p(x) = x^5 - 3x^4 + 2x^3 - 2x^2 - 2x + 1$ and $q(x) = x^4 - x^3 + 7x^2 - 2x$.

(ii) $\mathbb{F} = \mathbb{F}_2$, $p(x) = x^5 + x^4 + 1$ and $q(x) = x^5 + x^2 + x + 1$.

(b) Prove that the polynomial $c^{-1} r_t(x)$ in Step 4 is the greatest common divisor of $p(x)$ and $q(x)$.

(c) Prove that there exist polynomials $a(x)$ and $b(x)$ over $\mathbb{F}$ such that

$$a(x)\, p(x) + b(x)\, q(x) = \gcd(p(x),\, q(x)).$$

27. Let T be a linear operator on a vector space V over a field $\mathbb{F}$ and let $p(x)$, $q(x)$ be polynomials over $\mathbb{F}$ such that $\gcd(p(x),\, q(x)) = 1$.

(a) Prove that $\mathrm{Ker}(q(T)) \subseteq \mathrm{R}(p(T))$.

(b) Prove that $\mathrm{Ker}(p(T)) \cap \mathrm{Ker}(q(T)) = \{\mathbf{0}\}$.

(Hint: Use the result of Question 11.26(c).)

28. The following is a famous "wrong proof" of the Cayley-Hamilton Theorem:

- Let $\boldsymbol{A}$ be a square matrix.
  As $c_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A})$, $c_{\boldsymbol{A}}(\boldsymbol{A}) = \det(\boldsymbol{A}\boldsymbol{I} - \boldsymbol{A}) = \det(\boldsymbol{A} - \boldsymbol{A}) = \det(\mathbf{0}) = 0$.

Which of the equalities in the argument above is wrong? Why?

29. Define $p(x) = \det(x\boldsymbol{B} - \boldsymbol{C})$ where $\boldsymbol{B}$ and $\boldsymbol{C}$ are two $n \times n$ matrices. Suppose there exists an $n \times n$ matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}$. Prove that $p(\boldsymbol{A}) = \mathbf{0}$.

30. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ be a real matrix.

(a) Find the characteristic polynomial of $\boldsymbol{A}$.

(b) Find a real polynomial $p(x)$ such that $\boldsymbol{A}^{-1} = p(\boldsymbol{A})$.

31. Let $\boldsymbol{A}$ be a nonzero $n \times n$ matrix over a field $\mathbb{F}$. Prove that if $\boldsymbol{A}$ is invertible, then there is a polynomial $p(x)$ over $\mathbb{F}$ such that $\boldsymbol{A}^{-1} = p(\boldsymbol{A})$.

32. Consider the linear operator $T = L_{\boldsymbol{A}}$ on $\mathbb{R}^4$ with

$$\boldsymbol{A} = \begin{pmatrix} 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{pmatrix}.$$

Define $W_1 = \mathrm{Ker}(p(T))$ and $W_2 = \mathrm{Ker}(q(T))$ where $p(x) = (x-1)^2$ and $q(x) = (x-2)^2$.

(a) Compute $c_T(x)$ and show that the eigenvalues of T are 1 and 2.

(b) Find a basis $\{\boldsymbol{w_1}, \boldsymbol{w_2}\}$ for W_1 and a basis $\{\boldsymbol{w_3}, \boldsymbol{w_4}\}$ for W_2.

(c) Is $\mathbb{R}^4 = W_1 \oplus W_2$? (Hint: Use the result of Question 11.27(b).)

(d) Using $B = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{w_3}, \boldsymbol{w_4}\}$ as an ordered basis for $\mathbb{R}^4$, compute $[T]_B$.

Question 11.33 to Question 11.40 are exercises for Section 11.5.

33. Restate the results in Lemma 11.5.5 (Parts 1 and 3), Theorem 11.5.8 and Theorem 11.5.10 using square matrices.

(Hint: For Theorem 11.5.8, follow Discussion 11.3.12 to relate the direct sum of T-invariant subspaces with square matrices.)

34. Let $\boldsymbol{A} = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$ be a real matrix.

(a) Find the minimal polynomial of $\boldsymbol{A}$.

(b) By the result of (a), determine if $\boldsymbol{A}$ is diagonalizable.

35. Let $\boldsymbol{A}$ be a complex square matrix of order 3 such that

$$(\boldsymbol{A} - \boldsymbol{I})(\boldsymbol{A} + \mathrm{i}\boldsymbol{I}) = \boldsymbol{0}. \tag{11.6}$$

(a) List all possible answers for $m_{\boldsymbol{A}}(x)$.

(b) List all possible answers for $c_{\boldsymbol{A}}(x)$.

(c) Is $\boldsymbol{A}$ invertible? Justify your answer.

(d) Is $\boldsymbol{A}$ diagonalizable? Justify your answer.

(e) Find all complex 3×3 matrices that satisfy the equation (11.6).

36. (a) For $n \geq 2$, factorize $x^n - x$ into linear factors over $\mathbb{C}$.

    (b) Let $\boldsymbol{A}$ be a complex square matrix such that $\boldsymbol{A}^n = \boldsymbol{A}$ for some $n \geq 2$. Prove that $\boldsymbol{A}$ is diagonalizable.

37. (a) Prove that similar square matrices have the same minimal polynomial.

    (b) Suppose two square matrices have the same minimal polynomial, is it true that they are similar?

38. Prove Theorem 11.5.7:

    Let T be a linear operator on a vector space V. Suppose W_1 and W_2 are T-invariant subspace of V.

    (a) Show that $W_1 + W_2$ is T-invariant.

    (b) If W_1 and W_2 are finite dimensional with $\dim(W_1) \geq 1$ and $\dim(W_2) \geq 1$, prove that $m_{T|_{W_1+W_2}}(x)$ is equal to the least common multiple of $m_{T|_{W_1}}(x)$ and $m_{T|_{W_2}}(x)$, i.e.

$$m_{T|_{W_1+W_2}}(x) = \frac{m_{T|_{W_1}}(x)\, m_{T|_{W_2}}(x)}{\gcd(m_{T|_{W_1}}(x),\, m_{T|_{W_2}}(x))}.$$

39. (a) Let S and T be two diagonalizable linear operators on a finite dimensional vector space V. Show that there exists an ordered basis B for V such that $[S]_B$ and $[T]_B$ are diagonal matrices if and only if $S \circ T = T \circ S$.

    (b) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two diagonalizable matrices in $\mathcal{M}_{n\times n}(\mathbb{F})$. Show that there exists an invertible matrix $\boldsymbol{P} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ and $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$ are diagonal matrices if and only if $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$.

40. Let T be a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$ where $\dim(V) \geq 1$.

    (a) Suppose $m_T(x) = p(x)\,q(x)$ where $p(x)$ and $q(x)$ are polynomials over $\mathbb{F}$ such that $\deg(p(x)) \geq 1$, $\deg(q(x)) \geq 1$ and $\gcd(p(x),\, q(x)) = 1$.

       (i) Prove that $\mathrm{R}(p(T)) = \mathrm{Ker}(q(T))$.

       (ii) Prove that $V = \mathrm{Ker}(p(T)) \oplus \mathrm{Ker}(q(T))$.

       (Hint: Use the results of Question 11.27.)

    (b) Complete the proof of Theorem 11.5.8:

       Prove that $V = K_{\lambda_1}(T) \oplus K_{\lambda_2}(T) \oplus \cdots \oplus K_{\lambda_k}(T)$ where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all the (distinct) eigenvalues of T.

Question 11.41 to Question 11.54 are exercises for Section 11.6.

41. For each of the characteristic polynomials $c_{\boldsymbol{A}}(x)$ of a real matrix $\boldsymbol{A}$ below, (i) list all possible (non-similar) Jordan canonical forms for $\boldsymbol{A}$; and (ii) for each possible Jordan canonical form, write down the minimal polynomial $m_{\boldsymbol{A}}(x)$ of $\boldsymbol{A}$.

    (a) $c_{\boldsymbol{A}}(x) = (x-1)(x-2)(x-3)(x-4)$.

    (b) $c_{\boldsymbol{A}}(x) = (x-1)^2(x-2)^2$.

    (c) $c_{\boldsymbol{A}}(x) = (x-1)^4$.

42. Let V be a real vector space of dimension 5 and let T be a linear operator on V. Suppose that the minimal polynomial of T is

$$m_T(x) = (x-1)(x-2)^2.$$

    (i) List all possible (non-similar) Jordan canonical forms for T.

    (ii) For each possible Jordan canonical form, write down the characteristic polynomial $c_T(x)$ of T.

43. Let $\boldsymbol{A}$ be a complex square matrix of order 3 such that

$$(\boldsymbol{A} - \boldsymbol{I})^2(\boldsymbol{A} + \mathrm{i}\boldsymbol{I}) = \boldsymbol{0}. \tag{11.7}$$

    (a) List all possible answers for $m_{\boldsymbol{A}}(x)$.

    (b) List all possible answers for $c_{\boldsymbol{A}}(x)$.

    (c) Find all complex 3×3 matrices that satisfy the equation (11.7).

44. Let T be a linear operator on a finite dimensional real vector space V with an ordered basis $C = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_{11}}\}$ such that

$$[T]_C = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

    (a) Find the characteristic polynomial and the minimal polynomial of T.

    (b) For each $1 \le i \le 11$, express $T(\boldsymbol{v_i})$ as a linear combination of the vectors in C.

(c) Find the eigenvalues of T and bases for the corresponding eigenspaces.

(d) For each eigenvalue λ of T, write down a basis for $K_\lambda(T)$.

45. Let T be a linear operator on a real vector space such that T has a Jordan canonical form

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J_3}(2) & & \boldsymbol{0} \\ & \boldsymbol{J_2}(2) & \\ \boldsymbol{0} & & \boldsymbol{J_2}(3) \end{pmatrix}.$$

(a) Write down $c_T(x)$ and $m_T(x)$.

(b) Write down $\text{nullity}(T - 2I_V)$ and find $\text{nullity}((T - 2I_V)^2)$.

46. Let $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 2 & 2 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$ be a real matrix.

(a) Compute the characteristic polynomial of $\boldsymbol{A}$ and find all eigenvalues.

(b) For each eigenvalue λ obtained in Part (a), determine the dimension of the eigenspace associated with λ.

(c) Find a Jordan canonical form for $\boldsymbol{A}$.

(d) Write down the minimal polynomial of $\boldsymbol{A}$.

47. Let $\boldsymbol{E_{ij}}$ be the $n \times n$ matrix as defined in Example 8.4.6.7.

(a) Compute $\boldsymbol{E_{ij}}^2$.

(b) Find the characteristic polynomial and the minimal polynomial of $\boldsymbol{E_{ij}}$.

(c) Write down a Jordan canonical form for $\boldsymbol{E_{ij}}$.

48. (a) Let $\boldsymbol{J} = \boldsymbol{J_t}(\lambda)$ where $\lambda \neq 0$.

(i) Compute the inverse of $\boldsymbol{J}$.

(ii) Find the minimal polynomial of $\boldsymbol{J}^{-1}$.

(iii) Write down a Jordan canonical form for $\boldsymbol{J}^{-1}$.

(b) Suppose a square matrix $\boldsymbol{A}$ has a Jordan canonical form

$$\begin{pmatrix} \boldsymbol{J_{t_1}}(\lambda_1) & & & \boldsymbol{0} \\ & \boldsymbol{J_{t_2}}(\lambda_2) & & \\ & & \ddots & \\ & & & \ddots & \\ & \boldsymbol{0} & & \boldsymbol{J_{t_m}}(\lambda_m) \end{pmatrix}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are nonzero. Write down a Jordan canonical form for $\boldsymbol{A}^{-1}$.

49. Let T be a linear operator on a finite dimensional vector space V. Define $Q = T - \lambda I_V$ where λ is a scalar. Suppose $Q^2 = O_V$ and $Q \neq O_V$. Let $W = \text{Ker}(Q)$ and let $\{W + \boldsymbol{v_1}, W + \boldsymbol{v_2}, \ldots, W + \boldsymbol{v_m}\}$ be a basis for V/W.

    (a) Show that $Q(\boldsymbol{v_1}), Q(\boldsymbol{v_2}), \ldots, Q(\boldsymbol{v_m})$ are linearly independent vectors in W.

    (b) Let $\{Q(\boldsymbol{v_1}), Q(\boldsymbol{v_2}), \ldots, Q(\boldsymbol{v_m}), \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ be a basis for W. Explain why $B = \{Q(\boldsymbol{v_1}), \boldsymbol{v_1}, Q(\boldsymbol{v_2}), \boldsymbol{v_2}, \ldots, Q(\boldsymbol{v_m}), \boldsymbol{v_m}, \boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ is a basis for V. Write down $[Q]_B$ and $[T]_B$ using B as an ordered basis.

    (This question is a particular case of Question 11.53.)

50. (a) Let V be a finite dimensional vector space over a field $\mathbb{F}$.

    Suppose T is a linear operator on V such that $[T]_B = \boldsymbol{J_n}(\lambda)$ where $n = \dim(V)$, $\lambda \in \mathbb{F}$ and $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$ is an order basis for V.

    Let $C = \{\boldsymbol{v_n}, \boldsymbol{v_{n-1}}, \ldots, \boldsymbol{v_1}\}$. Write down $[T]_C$.

    (b) Prove that every complex square matrix is similar to its transpose.

51. (a) For $s \geq 1$, find the nullity of $(\boldsymbol{J_t}(\lambda) - \lambda \boldsymbol{I_t})^s$.

    (b) Let T be a linear operator on a finite dimensional vector space V such that T has a Jordan form

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J_{t_1}}(\lambda) & & & \boldsymbol{0} \\ & \boldsymbol{J_{t_2}}(\lambda) & & \\ & & \ddots & \\ & & & \ddots & \\ \boldsymbol{0} & & & & \boldsymbol{J_{t_m}}(\lambda) \end{pmatrix}$$

    where $t_1 + t_2 + \cdots + t_m = \dim(V)$.

    (i) Show that for $s \geq 1$, the nullity of $(\boldsymbol{J} - \lambda \boldsymbol{I_n})^s$ is equal to $\sum_{i=1}^{m} \min\{s, t_i\}$.

    (ii) Suppose $\text{nullity}((T - \lambda I_V)^s) - \text{nullity}((T - \lambda I_V)^{s-1}) = r$ where $s \geq 2$. What can you say about the Jordan blocks in $\boldsymbol{J}$ based on the value of r? (Compare your answer with the results in Question 11.53.)

52. Let T be a real linear operator on a finite dimensional vector space V such that

    (i) $c_T(x) = (x+1)^9(x-2)^7$;

    (ii) $m_T(x) = (x+1)^4(x-2)^3$;

    (iii) $\text{nullity}(T + I_V) = 4$, $\text{nullity}((T + I_V)^2) = 7$, $\text{nullity}((T + I_V)^3) = 8$; and

    (iv) $\text{nullity}(T - 2I_V) = 3$, $\text{nullity}((T - 2I_V)^2) = 5$.

    Find a Jordan canonical form for T. (Hint: You may need the answer to Question 11.51(b)(ii).)

53. (This question is the preliminary of the proof of Theorem 11.6.4 in Question 11.54. You can only use results discussed before Theorem 11.6.4 to do this question.)

    Let T be a linear operator on a finite dimensional vector space V. Suppose λ is an eigenvalue of T. Define $Q = T - \lambda I_V$ and $N_t = \mathrm{Ker}(Q^t)$ for $t \geq 0$.

    (a) Suppose $\boldsymbol{v} \in N_t - N_{t-1}$ where $t \geq 1$. Let $B = \{Q^{t-1}(\boldsymbol{v}), \ldots, Q(\boldsymbol{v}), \boldsymbol{v}\}$ and $W = \mathrm{span}(B)$.

        (i) Prove that W is T-invariant.

        (ii) Prove that B is a basis for W.

        (iii) Using B as an ordered basis, write down $[T|_W]_B$.

    (b) For $1 \leq t \leq s$, let $C_t \subseteq N_t - N_{t-1}$ such that $\{N_{t-1} + \boldsymbol{v} \mid \boldsymbol{v} \in C_t\}$ is a basis for N_t/N_{t-1}. (We also assume that for any $\boldsymbol{u}, \boldsymbol{v} \in C_t$, if $\boldsymbol{u} \neq \boldsymbol{v}$, then $N_{t-1} + \boldsymbol{u} \neq N_{t-1} + \boldsymbol{v}$.)

        (i) Prove that for $t \geq 2$, if $C_t = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$, then $N_{t-2} + Q(\boldsymbol{v_1})$, $N_{t-2} + Q(\boldsymbol{v_2})$, $\ldots$, $N_{t-2} + Q(\boldsymbol{v_k})$ are linearly independent vectors in N_{t-1}/N_{t-2}.

        (ii) Prove that $C_1 \cup C_2 \cup \cdots \cup C_s$ is a basis for N_s.

    (c) For $s \geq 1$, show that there exists an ordered basis D for N_s such that $[T|_{N_s}]_D$ is a Jordan canonical form for $T|_{N_s}$. (Hint: Construct sets C_t in part (b) so that $\{Q(\boldsymbol{v}) \mid \boldsymbol{v} \in C_t\} \subseteq C_{t-1}$ for $t \geq 2$. Then rearrange the vectors in $C_1 \cup C_2 \cup \cdots \cup C_s$ to form a suitable ordered basis D for V.)

54. (This question is a proof of Theorem 11.6.4. You can only use results of Question 11.53 and results discussed before Theorem 11.6.4 to do this question.)

    Let T be a linear operator on a finite dimensional space V over a field $\mathbb{F}$. Suppose the characteristic polynomial of T can be factorized into linear factors over $\mathbb{F}$. Use the results of Theorem 11.5.8 and Question 11.53 to show that there exists an ordered basis B for V such that $[T]_B$ is a Jordan canonical form for T.

Chapter 12

Inner Product Spaces

In this chapter, we only study real and complex vector spaces.

Section 12.1 Inner Products

Discussion 12.1.1 In Chapter 5, we use dot product to define lengths, distances and angles in $\mathbb{R}^n$. For general vector spaces, we need an abstract version of "dot product" so that we can generalize the works we have done in Chapter 5. In order to do so, we first require that the field used must have some built-in measurement and ordering. Since not all fields are suitable, we only study vector spaces over $\mathbb{R}$ and $\mathbb{C}$ in this chapter.

Notation 12.1.2 For a complex number $c = a + b\mathrm{i}$, where $a, b \in \mathbb{R}$, we use $\overline{c}$ to denote the *complex conjugate* of c, i.e. $\overline{c} = a - b\mathrm{i}$. In particular, if c is real, then $\overline{c} = c$.

Let $\boldsymbol{A} = (a_{ij})_{m\times n}$ be a complex matrix. We use $\overline{\boldsymbol{A}}$ to denote the *conjugate* of $\boldsymbol{A}$, i.e. $\overline{\boldsymbol{A}} = (\overline{a}_{ij})_{m\times n}$. Furthermore, the *conjugate transpose* of $\boldsymbol{A}$ is denoted by

$$\boldsymbol{A}^* = \overline{\boldsymbol{A}}^{\mathrm{T}} = (\overline{a}_{ji})_{n\times m}.$$

In particular, if $\boldsymbol{A}$ is a real matrix, then $\overline{\boldsymbol{A}} = \boldsymbol{A}$ and $\boldsymbol{A}^* = \boldsymbol{A}^{\mathrm{T}}$.

Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{m\times n}(\mathbb{C})$, $\boldsymbol{C} \in \mathcal{M}_{n\times p}(\mathbb{C})$ and $c \in \mathbb{C}$. Then

$$\begin{aligned}
(\boldsymbol{A}+\boldsymbol{B})^* &= (\,\overline{\boldsymbol{A}+\boldsymbol{B}}\,)^{\mathrm{T}} = (\,\overline{\boldsymbol{A}}+\overline{\boldsymbol{B}}\,)^{\mathrm{T}} = \overline{\boldsymbol{A}}^{\mathrm{T}} + \overline{\boldsymbol{B}}^{\mathrm{T}} = \boldsymbol{A}^* + \boldsymbol{B}^*;\\
(\boldsymbol{A}\boldsymbol{C})^* &= (\,\overline{\boldsymbol{A}\boldsymbol{C}}\,)^{\mathrm{T}} = (\,\overline{\boldsymbol{A}}\,\overline{\boldsymbol{C}}\,)^{\mathrm{T}} = \overline{\boldsymbol{C}}^{\mathrm{T}}\overline{\boldsymbol{A}}^{\mathrm{T}} = \boldsymbol{C}^*\boldsymbol{A}^*;\\
(c\boldsymbol{A})^* &= (\,\overline{c\boldsymbol{A}}\,)^{\mathrm{T}} = (\,\overline{c}\,\overline{\boldsymbol{A}}\,)^{\mathrm{T}} = \overline{c}\,\overline{\boldsymbol{A}}^{\mathrm{T}} = \overline{c}\,\boldsymbol{A}^*.
\end{aligned}$$

Definition 12.1.3 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and let V be a vector space over $\mathbb{F}$. An *inner product* on V is a mapping which assigns to each ordered pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ a scalar $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \in \mathbb{F}$ such that it satisfies the following axioms:

(IP1) For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle}$. (This axiom implies $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \in \mathbb{R}$ for all $\boldsymbol{u} \in V$.)

(IP2) For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $\langle \boldsymbol{u} + \boldsymbol{v},\ \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$.

(IP3) For all $c \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle c\boldsymbol{u}, \boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

(IP4) $\langle \boldsymbol{0}, \boldsymbol{0} \rangle = 0$ and, for all nonzero $\boldsymbol{u} \in V$, $\langle \boldsymbol{u}, \boldsymbol{u} \rangle > 0$.

(Compare the axioms with Theorem 5.1.5.)

Remark 12.1.4

1. If $\mathbb{F} = \mathbb{R}$, we can rewrite (IP1) as:

   (IP1') For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$.

   For this case, we say that the inner product is *symmetric*.

2. By (IP1) and (IP2), for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$,

$$\langle \boldsymbol{w},\ \boldsymbol{u} + \boldsymbol{v} \rangle = \overline{\langle \boldsymbol{u} + \boldsymbol{v},\ \boldsymbol{w} \rangle} = \overline{\langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle} = \overline{\langle \boldsymbol{u}, \boldsymbol{w} \rangle} + \overline{\langle \boldsymbol{v}, \boldsymbol{w} \rangle} = \langle \boldsymbol{w}, \boldsymbol{u} \rangle + \langle \boldsymbol{w}, \boldsymbol{v} \rangle.$$

3. By (IP1) and (IP3), for all $c \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\langle \boldsymbol{u}, c\boldsymbol{v} \rangle = \overline{\langle c\boldsymbol{v}, \boldsymbol{u} \rangle} = \overline{c\langle \boldsymbol{v}, \boldsymbol{u} \rangle} = \overline{c}\,\overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle} = \overline{c}\,\langle \boldsymbol{u}, \boldsymbol{v} \rangle.$$

   When $\mathbb{F} = \mathbb{R}$, we have $\langle \boldsymbol{u}, c\boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

4. By (IP2), for all $\boldsymbol{u} \in V$,

$$\langle \boldsymbol{0}, \boldsymbol{u} \rangle = \langle \boldsymbol{0} + \boldsymbol{0},\ \boldsymbol{u} \rangle = \langle \boldsymbol{0}, \boldsymbol{u} \rangle + \langle \boldsymbol{0}, \boldsymbol{u} \rangle \quad \Rightarrow \quad \langle \boldsymbol{0}, \boldsymbol{u} \rangle = 0$$

   and then by (IP1), $\langle \boldsymbol{u}, \boldsymbol{0} \rangle = \overline{\langle \boldsymbol{0}, \boldsymbol{u} \rangle} = \overline{0} = 0$.

Definition 12.1.5 A vector space V, over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, equipped with an inner product is called an *inner product space*.

If $\mathbb{F} = \mathbb{R}$, V is called a *real inner product space*; and if $\mathbb{F} = \mathbb{C}$, V is called a *complex inner product space*.

Example 12.1.6

1. For all $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, define

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}.$$

   Note that $\langle\ ,\ \rangle$ is actually the dot product defined in Chapter 5. By Theorem 5.1.5, $\langle\ ,\ \rangle$

is an inner product on $\mathbb{R}^n$.

This inner product is also called the *usual inner product* or the *Euclidean inner product* on $\mathbb{R}^n$. Furthermore, the *Euclidean n-space* is usually referred to the vector space $\mathbb{R}^n$ equipped with this inner product.

2. For all $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{C}^n$, define

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n = \boldsymbol{u}\boldsymbol{v}^*$$

Then $\langle\ ,\ \rangle$ is an inner product on $\mathbb{C}^n$ which is called the *usual inner product* on $\mathbb{C}^n$.

3. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. For $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{m\times n}(\mathbb{F})$, define $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^*)$. (See Definition 8.1.10 and Proposition 8.1.11 for the definition and properties of the trace function tr.)

If $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$, then

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^*) = \mathrm{tr}\left(\left(\sum_{k=1}^{n} a_{ik}\overline{b}_{jk} \right)_{m\times m} \right) = \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}\overline{b}_{ik}.$$

It is obvious that $\langle\ ,\ \rangle$ is an inner product on $\mathcal{M}_{m\times n}(\mathbb{F})$.

4. Determine which of the followings are inner products on $\mathbb{R}^2$.

(a) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1^2 + u_2^2 + v_1^2 + v_2^2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(b) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1v_1 + 2u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(c) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1v_1 - 2u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(d) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{3}{2}u_1v_1 - \frac{1}{2}u_1v_2 + \frac{1}{2}u_2v_1 + \frac{3}{2}u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

(e) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{3}{2}u_1v_1 - \frac{1}{2}u_1v_2 - \frac{1}{2}u_2v_1 + \frac{3}{2}u_2v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.

Solution

(a) It does not satisfy (IP2) and (IP3). Hence it is not an inner product.

(b) It is an inner product. (Check it.)

(c) It does not satisfy (IP4). Hence it is not an inner product.

(d) It does not satisfy (IP1). Hence it is not an inner product.

(e) $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \left(\frac{u_1+u_2}{\sqrt{2}}\right)\left(\frac{v_1+v_2}{\sqrt{2}}\right) + 2\left(\frac{u_1-u_2}{\sqrt{2}}\right)\left(\frac{v_1-v_2}{\sqrt{2}}\right)$ for all $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$. Same as (b), it is an inner product.

5. Let $[a, b]$, with $a < b$, be a closed interval on the real line. Consider the vector space $C([a, b])$ defined in Example 8.3.6.5. Then

$$\langle f, g \rangle = \tfrac{1}{b-a}\int_a^b f(t)g(t)dt \quad \text{for } f, g \in C([a, b])$$

is an inner product on $C[a, b]$). (We leave the verification as an exercise. See Question 12.5.)

6. Let V be the set of all real infinite sequences $(a_n)_{n\in\mathbb{N}}$ such that $\sum_{n=1}^{\infty} a_n^2$ converges. Define

$$\langle (a_n)_{n\in\mathbb{N}},\ (b_n)_{n\in\mathbb{N}} \rangle = \sum_{n=1}^{\infty} a_n b_n \quad \text{for } (a_n)_{n\in\mathbb{N}},\ (b_n)_{n\in\mathbb{N}} \in V.$$

   Then V is a real inner product space. (We leave the verification as an exercise. See Question 12.6.)

   This inner product space is called the *l_2-space*. It is an example of a class of inner product spaces known as *Hilbert spaces*.

Section 12.2 Norms and Distances

Discussion 12.2.1 One of the important uses of an inner product is that we can use it to measure the length of a vector and the distance between two vectors.

Definition 12.2.2 Let V be an inner product space.

1. For $\boldsymbol{u} \in V$, the *norm* (or *length*) of $\boldsymbol{u}$ is defined to be $||\boldsymbol{u}|| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$.

   In particular, vectors of norm 1 are called *unit vectors*.

2. For $\boldsymbol{u}, \boldsymbol{v} \in V$, the *distance* between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}||$.

Example 12.2.3

1. Let $\mathbb{R}^n$ be equipped with the usual inner product. For $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$,

$$||\boldsymbol{u}|| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle} = \sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

   and for $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$,

$$\begin{aligned} d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| &= \sqrt{\langle \boldsymbol{u} - \boldsymbol{v},\ \boldsymbol{u} - \boldsymbol{v} \rangle} \\ &= \sqrt{(\boldsymbol{u} - \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v})} = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}. \end{aligned}$$

   (See Section 5.1.)

2. Let $\mathbb{C}^n$ be equipped with the usual inner product. For $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{C}^n$,

$$||\boldsymbol{u}|| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle} = \sqrt{|u_1|^2 + |u_2|^2 + \cdots + |u_n|^2}$$

   and for $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$,

$$d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = \sqrt{\langle \boldsymbol{u} - \boldsymbol{v},\ \boldsymbol{u} - \boldsymbol{v} \rangle} = \sqrt{|u_1 - v_1|^2 + |u_2 - v_2|^2 + \cdots + |u_n - v_n|^2}.$$

   (For $c = a + b\mathrm{i} \in \mathbb{C}$ where $a, b \in \mathbb{R}$, $|c| = \sqrt{c\,\bar{c}} = \sqrt{a^2 + b^2}$ is called the *modulus* of c.)

3. We compare the usual inner product on $\mathbb{R}^2$ with the inner products defined in Parts (b) and (e) of Example 12.1.6.4.

The usual inner product	Example 12.1.6.4(b)	Example 12.1.6.4(e)
$\langle\,(1,0),\,(0,1)\,\rangle = 0$	$\langle\,(1,0),\,(0,1))\,\rangle = 0$	$\langle\,(1,0),\,(0,1)\,\rangle = -\frac{1}{2}$
$\|\|(1,0)\|\| = 1$	$\|\|(1,0)\|\| = 1$	$\|\|(1,0)\|\| = \sqrt{\frac{3}{2}}$
$\|\|(0,1)\|\| = 1$	$\|\|(0,1)\|\| = \sqrt{2}$	$\|\|(0,1)\|\| = \sqrt{\frac{3}{2}}$
$d((1,0),\,(0,1))$ $= \|\|(1,-1)\|\| = \sqrt{2}$	$d((1,0),\,(0,1))$ $= \|\|(1,-1)\|\| = \sqrt{3}$	$d((1,0),\,(0,1))$ $= \|\|(1,-1)\|\| = 2$

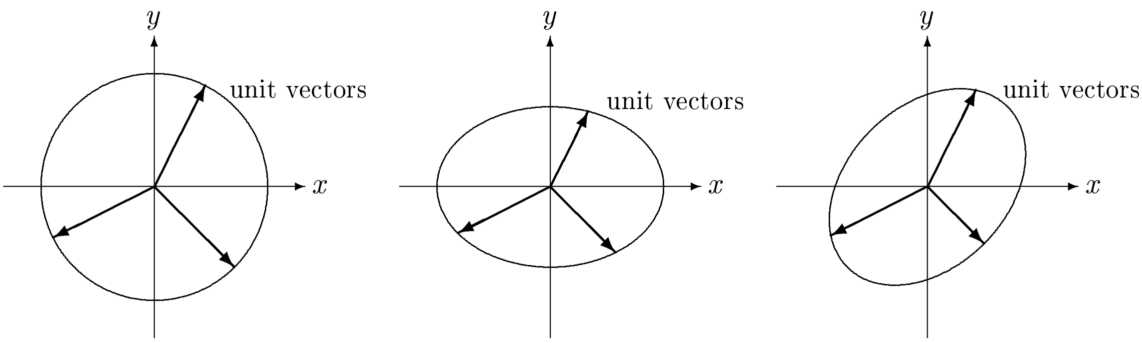
Using the usual inner product, for $\boldsymbol{u} = (x, y) \in \mathbb{R}^2$,

$$\boldsymbol{u} = 1 \quad \Leftrightarrow \quad x^2 + y^2 = 1.$$

Thus, the set of all unit vectors forms the circle $x^2 + y^2 = 1$.

Similarly, using the inner product in Example 12.1.6.4(b), the set of all unit vectors forms the ellipse $x^2 + 2y^2 = 1$; and

using the inner product in Example 12.1.6.4(e), the set of all unit vectors forms the ellipse $3x^2 - 2xy + 3y^2 = 2$.



The usual inner product Example 12.1.6.4(b) Example 12.1.6.4(e)

4. Let $\mathcal{M}_{m\times n}(\mathbb{F})$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, be equipped with the inner product defined in Example 12.1.6.3. For $\boldsymbol{A} = (a_{ij})$, $\boldsymbol{B} = (b_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$,

$$||\boldsymbol{A}|| = \sqrt{\langle \boldsymbol{A}, \boldsymbol{A}\rangle} = \sqrt{\sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}\bar{a}_{ik}} = \sqrt{\sum_{i=1}^{m}\sum_{k=1}^{n} |a_{ik}|^2}$$

and

$$d(\boldsymbol{A},\,\boldsymbol{B}) = ||\boldsymbol{A} - \boldsymbol{B}|| = \sqrt{\sum_{i=1}^{m}\sum_{k=1}^{n} |a_{ik} - b_{ik}|^2}.$$

5. Let $[a, b]$, with $a < b$, be a closed interval on the real line. Suppose the vector space $C([a, b])$ is equipped with the inner product defined in Example 12.1.6.5. For $f, g \in C([a, b])$,
$$||f|| = \sqrt{\langle f, f\rangle} = \sqrt{\tfrac{1}{b-a}\int_a^b [f(t)]^2 dt}$$
and
$$d(f, g) = ||f - g|| = \sqrt{\tfrac{1}{b-a}\int_a^b [f(t) - g(t)]^2 dt}$$
.

Theorem 12.2.4 Let V be an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

1. $||\mathbf{0}|| = 0$ and, for any nonzero $\boldsymbol{u} \in V$, $||\boldsymbol{u}|| > 0$.
2. For any $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $||c\boldsymbol{u}|| = |c|\,||\boldsymbol{u}||$.
3. **(Cauchy-Schwarz Inequality)** For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| \le ||\boldsymbol{u}||\,||\boldsymbol{v}||$.
   The equality holds if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent, i.e. $\boldsymbol{u} = a\boldsymbol{v}$ for some $a \in \mathbb{F}$ or $\boldsymbol{v} = b\boldsymbol{u}$ for some $b \in \mathbb{F}$.
4. **(Triangle Inequality)** For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $||\boldsymbol{u} + \boldsymbol{v}|| \le ||\boldsymbol{u}|| + ||\boldsymbol{v}||$.

Proof We only prove the Cauchy-Schwarz Inequality:

If $\boldsymbol{v} = \mathbf{0}$, then $|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| = 0 = ||\boldsymbol{u}||\,||\boldsymbol{v}||$. So the inequality is trivially true for this case.

Now, assume $\boldsymbol{v} \ne 0$. Define
$$\boldsymbol{w} = \boldsymbol{u} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\,\boldsymbol{v}.$$
Then
$$\begin{aligned}
0 \le ||\boldsymbol{w}||^2 = \langle \boldsymbol{w}, \boldsymbol{w}\rangle &= \left\langle \boldsymbol{u} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\,\boldsymbol{v},\ \boldsymbol{u} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\,\boldsymbol{v}\right\rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u}\rangle - \frac{\overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}}{||\boldsymbol{v}||^2}\langle \boldsymbol{u}, \boldsymbol{v}\rangle - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\langle \boldsymbol{v}, \boldsymbol{u}\rangle + \left(\frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\right)\left(\overline{\frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}}\right)\langle \boldsymbol{v}, \boldsymbol{v}\rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u}\rangle - \frac{\overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}}{||\boldsymbol{v}||^2}\langle \boldsymbol{u}, \boldsymbol{v}\rangle - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{||\boldsymbol{v}||^2}\overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle} + \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle\,\overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}}{||\boldsymbol{v}||^2} \\
&= ||\boldsymbol{u}||^2 - \frac{|\langle \boldsymbol{u}, \boldsymbol{v}\rangle|^2}{||\boldsymbol{v}||^2}. \qquad (12.1)
\end{aligned}$$

So $\dfrac{|\langle \boldsymbol{u}, \boldsymbol{v}\rangle|^2}{||\boldsymbol{v}||^2} \le ||\boldsymbol{u}||^2$ and hence $|\langle \boldsymbol{u}, \boldsymbol{v}\rangle|^2 \le ||\boldsymbol{u}||^2||\boldsymbol{v}||^2$.

(Proofs of the other parts are left as exercises. See Question 12.12.)

Example 12.2.5

1. **(Cauchy-Schwarz Inequality for Real Numbers)** For any real numbers $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_n$, prove that

$$(x_1y_1 + x_2y_2 + \cdots + x_ny_n)^2 \leq (x_1^2 + x_2^2 + \cdots + x_n^2)(y_1^2 + y_2^2 + \cdots + y_n^2).$$

   Solution Use $\mathbb{R}^n$ equipped with the usual inner product. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. Then $||\boldsymbol{x}||^2 = x_1^2 + x_2^2 + \cdots + x_n^2$, $||\boldsymbol{y}||^2 = y_1^2 + y_2^2 + \cdots + y_n^2$ and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$. The inequality follows by Theorem 12.2.4.3.

2. **(Cauchy-Schwarz Inequality for Continuous Functions)** For any $f, g \in C([a, b])$ where $[a, b]$, with $a < b$, is a closed interval on the real line, prove that

$$\left(\int_a^b f(t)g(t)dt\right)^2 \leq \left(\int_a^b f(t)^2 dt\right)\left(\int_a^b g(t)^2 dt\right).$$

   Solution The inequality follows by applying Theorem 12.2.4.3 to the inner product defined in Example 12.1.6.5.

Section 12.3 Orthogonal and Orthonormal Bases

Discussion 12.3.1 In Section 5.1, we use the dot product to define the angle between two vectors in $\mathbb{R}^n$. Given a real inner product space V, we can define the angle between two vectors in V in the same way, i.e. the *angle* between $\boldsymbol{u}, \boldsymbol{v} \in V$ is

$$\cos^{-1}\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

In particular, $\boldsymbol{u}$ and $\boldsymbol{v}$ are *perpendicular* to each other if and only if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. Note that by the Cauchy-Schwarz Inequality (Theorem 12.2.4.3), we have $-1 \leq \dfrac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{||\boldsymbol{u}||\,||\boldsymbol{v}||} \leq 1$ and hence the angle is well-defined.

Although this definition of "angles" does not work for complex inner product spaces, the concept of "perpendicular" can still be defined accordingly.

In the following, we restate some important results in Chapter 5 using inner product spaces.

Definition 12.3.2 Let V be an inner product space.

1. Two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$ are said to be *orthogonal* to each other if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.

2. Let W be a subspace of V. A vector $\boldsymbol{u}$ is said to be *orthogonal* (or *perpendicular*) to W if $\boldsymbol{u}$ is orthogonal to all vectors in W.

3. A subset B of V is called *orthogonal* if the vectors in B are pairwise orthogonal.

   If B is an orthogonal set and it is a basis for V, then B is called an *orthogonal basis* for V.

4. A subset B of V is called *orthonormal* if B is orthogonal and all vectors in B are unit vectors.

   If B is an orthonormal set and it is a basis for V, then B is called an *orthonormal basis* for V.

(See Definition 5.2.1, Definition 5.2.5 and Definition 5.2.10.)

Lemma 12.3.3 Let V be an inner product space over $\mathbb{F}$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

1. Let $W = \text{span}(B)$ where $B \subseteq V$. For $\boldsymbol{u} \in V$, $\boldsymbol{u}$ is orthogonal to W if and only if $\boldsymbol{u}$ is orthogonal to every vectors in B.

2. If B is an orthogonal set of nonzero vectors from V, then B is always linearly independent.

3. Suppose V is finite dimensional where $\dim(V) \geq 1$. Let B be an ordered orthonormal basis for V. Then
$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = (\boldsymbol{u})_B \, ((\boldsymbol{v})_B)^* = ([\boldsymbol{u}]_B)^{\mathrm{T}} \, \overline{[\boldsymbol{v}]}_B.$$
   Furthermore, if $\mathbb{F} = \mathbb{R}$, then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = (\boldsymbol{u})_B \cdot (\boldsymbol{v})_B$.

Proof

1. ($\Leftarrow$) It is obvious.

   ($\Rightarrow$) Suppose $\boldsymbol{u} \in V$ is orthogonal to every vectors in B. Take any $\boldsymbol{w} \in W$. Since $W = \text{span}(B)$, we can write $\boldsymbol{w} = a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + \cdots + a_k\boldsymbol{v_k}$ for some $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k} \in B$ and $a_1, a_2, \ldots, a_k \in \mathbb{F}$. Then
$$\begin{aligned} \langle \boldsymbol{u}, \boldsymbol{w} \rangle &= \langle \boldsymbol{u}, a_1\boldsymbol{v_1} \rangle + \langle \boldsymbol{u}, a_2\boldsymbol{v_2} \rangle + \cdots + \langle \boldsymbol{u}, a_k\boldsymbol{v_k} \rangle \\ &= \bar{a}_1 \langle \boldsymbol{u}, \boldsymbol{v_1} \rangle + \bar{a}_2 \langle \boldsymbol{u}, \boldsymbol{v_2} \rangle + \cdots + \bar{a}_k \langle \boldsymbol{u}, \boldsymbol{v_k} \rangle \\ &= \bar{a}_1 0 + \bar{a}_2 0 + \cdots + \bar{a}_k 0 = 0. \end{aligned}$$
   So $\boldsymbol{u}$ is orthogonal to all vector in W and hence is orthogonal to W.

2. Take any finite subset $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ of B. Consider the vector equation
$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k} = \boldsymbol{0} \tag{12.2}$$

Since B is orthogonal, $\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = 0$ whenever $i \neq j$. For each $j = 1, 2, \ldots, k$, (12.2) implies

$$\begin{aligned} & \langle c_1 \boldsymbol{v_1} + c_2 \boldsymbol{v_2} + \cdots + c_k \boldsymbol{v_k},\ \boldsymbol{v_j} \rangle = \langle 0, \boldsymbol{v_j} \rangle = 0 \\ \Rightarrow \quad & c_1 \langle \boldsymbol{v_1}, \boldsymbol{v_j} \rangle + c_2 \langle \boldsymbol{v_2}, \boldsymbol{v_j} \rangle + \cdots + c_k \langle \boldsymbol{v_k}, \boldsymbol{v_j} \rangle = 0 \\ \Rightarrow \quad & c_j \langle \boldsymbol{v_j}, \boldsymbol{v_j} \rangle = 0 \\ \Rightarrow \quad & c_j = 0 \quad \text{(because vectors in } B \text{ are nonzero vectors).} \end{aligned}$$

Since (12.2) has only the trivial solution, $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ is linearly independent.

As all finite subsets of B are linearly independent, B is linearly independent.

3. Let $B = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_n}\}$. Note that for $i, j \in \{1, 2, \ldots, n\}$,

$$\langle \boldsymbol{w_i}, \boldsymbol{w_j} \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Take any $\boldsymbol{u}, \boldsymbol{v} \in V$, say, $\boldsymbol{u} = a_1 \boldsymbol{w_1} + a_2 \boldsymbol{w_2} + \cdots + a_n \boldsymbol{w_n}$ and $\boldsymbol{v} = b_1 \boldsymbol{w_1} + b_2 \boldsymbol{w_2} + \cdots + b_n \boldsymbol{w_n}$ where $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{F}$. Then

$$\begin{aligned} \langle \boldsymbol{u}, \boldsymbol{v} \rangle &= \langle a_1 \boldsymbol{w_1} + a_2 \boldsymbol{w_2} + \cdots + a_n \boldsymbol{w_n},\ b_1 \boldsymbol{w_1} + b_2 \boldsymbol{w_2} + \cdots + b_n \boldsymbol{w_n} \rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle a_i \boldsymbol{w_i},\ b_j \boldsymbol{w_j} \rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{b}_j \langle \boldsymbol{w_i}, \boldsymbol{w_j} \rangle \\ &= a_1 \overline{b}_1 + a_2 \overline{b}_2 + \cdots + a_n \overline{b}_n \\ &= (\boldsymbol{u})_B \left((\boldsymbol{v})_B \right)^* \\ &= ([\boldsymbol{u}]_B)^{\mathrm{T}}\, \overline{[\boldsymbol{v}]}_B. \end{aligned}$$

When $\mathbb{F} = \mathbb{R}$, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = (\boldsymbol{u})_B \cdot (\boldsymbol{v})_B$.

Remark 12.3.4

1. Suppose V is a finite dimensional inner product space. By Theorem 8.5.13 and Lemma 12.3.3.2, if we know the dimension of V, to determine whether a set B of nonzero vectors from V is an orthogonal (respectively, orthonormal) basis for V, we only need to check:

   (i) B is orthogonal (respectively, orthonormal); and

   (ii) $|B| = \dim(V)$.

   (See Remark 5.2.6.)

2. By Lemma 12.3.3.3, a finite dimensional real inner product space is essentially the same as the Euclidean space.

Example 12.3.5

1. Suppose $\mathbb{C}^3$ is equipped with the usual inner product. Let $W = \text{span}\{(1,1,1),\ (1,\mathrm{i},\mathrm{i})\}$. For any $(x,y,z) \in \mathbb{C}^3$, by Lemma 12.3.3.1,

$$
\begin{aligned}
(x,y,z) \text{ is orthogonal to } W \quad &\Leftrightarrow \quad \begin{cases} \langle (x,y,z),\ (1,1,1)\rangle = 0 \\ \langle (x,y,z),\ (1,\mathrm{i},\mathrm{i})\rangle = 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} x + \ y + \ z = 0 \\ x - \mathrm{i}y - \mathrm{i}z = 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} x = 0 \\ y = -t \qquad \text{for } t \in \mathbb{C}. \\ z = t \end{cases}
\end{aligned}
$$

   So $(0,-t,t)$, $t \in \mathbb{C}$, are all the vectors orthogonal to W.

2. Suppose $\mathcal{M}_{n\times n}(\mathbb{R})$ is equipped with the inner product defined in Example 12.1.6.3. Let $W = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A} \text{ is symmetric}\}$.

   Take any skew symmetric matrix $\boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$, i.e. $\boldsymbol{B}^{\mathrm{T}} = -\boldsymbol{B}$. For any $\boldsymbol{A} \in W$, i.e. $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$, by Proposition 8.1.11, we have

$$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}) = \mathrm{tr}((\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}})^{\mathrm{T}}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}) = \mathrm{tr}(-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) = -\mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) = -\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}).$$

   This implies that $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}) = 0$. So $\boldsymbol{B}$ is orthogonal to W.

3. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Using the usual inner product, the standard basis $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ for $\mathbb{F}^n$ is an orthonormal basis.

4. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Using the inner product defined in Example 12.1.6.3, the standard basis $\{\boldsymbol{E_{ij}} \mid 1 \le i \le m \text{ and } 1 \le j \le n\}$ for $\mathcal{M}_{m\times n}(\mathbb{F})$ is an orthonormal basis.

5. Suppose $\mathcal{P}_2(\mathbb{R})$ is equipped with an inner product such that

$$\langle p(x), q(x)\rangle = \tfrac{1}{2}\int_{-1}^{1} p(t)q(t)dt \quad \text{for } p(x), q(x) \in \mathcal{P}_2(\mathbb{R}).$$

   Consider the standard basis $\{1, x, x^2\}$ for $\mathcal{P}_2(\mathbb{R})$. Then

$$\langle 1, x\rangle = \tfrac{1}{2}\int_{-1}^{1} t dt = 0, \quad \langle x, x^2\rangle = \tfrac{1}{2}\int_{-1}^{1} t^3 dt = 0 \quad \text{and} \quad \langle 1, x^2\rangle = \tfrac{1}{2}\int_{-1}^{1} t^2 dt = \tfrac{1}{3},$$

   i.e. x is orthogonal to both 1 and x^2 but 1 and x^2 are not orthogonal to each other. The standard basis $\{1, x, x^2\}$ is not an orthogonal basis.

Theorem 12.3.6 Let V be a finite dimensional inner product space. If $B = \{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \dots,\, \boldsymbol{w_n}\}$ is an orthonormal basis for V, then for any vector $\boldsymbol{u} \in V$,

$$\boldsymbol{u} = \langle \boldsymbol{u}, \boldsymbol{w_1} \rangle\, \boldsymbol{w_1} + \langle \boldsymbol{u}, \boldsymbol{w_2} \rangle\, \boldsymbol{w_2} + \cdots + \langle \boldsymbol{u}, \boldsymbol{w_n} \rangle\, \boldsymbol{w_n},$$

i.e. using B as an ordered basis, $(\boldsymbol{u})_B = (\langle \boldsymbol{u}, \boldsymbol{w_1} \rangle,\, \langle \boldsymbol{u}, \boldsymbol{w_2} \rangle,\, \dots,\, \langle \boldsymbol{u}, \boldsymbol{w_n} \rangle)$.

Proof The proof follows the same argument as the proof for Theorem 5.2.8.

Theorem 12.3.7 (Gram-Schmidt Process) Suppose $\{\boldsymbol{u_1},\, \boldsymbol{u_2},\, \dots,\, \boldsymbol{u_n}\}$ is a basis for a finite dimensional inner product space V. Let

$$\begin{aligned}
\boldsymbol{v_1} &= \boldsymbol{u_1},\\
\boldsymbol{v_2} &= \boldsymbol{u_2} - \frac{\langle \boldsymbol{u_2}, \boldsymbol{v_1} \rangle}{\langle \boldsymbol{v_1}, \boldsymbol{v_1} \rangle}\, \boldsymbol{v_1},\\
\boldsymbol{v_3} &= \boldsymbol{u_3} - \frac{\langle \boldsymbol{u_3}, \boldsymbol{v_1} \rangle}{\langle \boldsymbol{v_1}, \boldsymbol{v_1} \rangle}\, \boldsymbol{v_1} - \frac{\langle \boldsymbol{u_3}, \boldsymbol{v_2} \rangle}{\langle \boldsymbol{v_2}, \boldsymbol{v_2} \rangle}\, \boldsymbol{v_2},\\
&\ \ \vdots\\
\boldsymbol{v_n} &= \boldsymbol{u_n} - \frac{\langle \boldsymbol{u_n}, \boldsymbol{v_1} \rangle}{\langle \boldsymbol{v_1}, \boldsymbol{v_1} \rangle}\, \boldsymbol{v_1} - \frac{\langle \boldsymbol{u_n}, \boldsymbol{v_2} \rangle}{\langle \boldsymbol{v_2}, \boldsymbol{v_2} \rangle}\, \boldsymbol{v_2} - \cdots - \frac{\langle \boldsymbol{u_n}, \boldsymbol{v_{n-1}} \rangle}{\langle \boldsymbol{v_{n-1}}, \boldsymbol{v_{n-1}} \rangle}\, \boldsymbol{v_{n-1}}.
\end{aligned}$$

Then $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \dots,\, \boldsymbol{v_n}\}$ is an orthogonal basis for V. Furthermore, let

$$\boldsymbol{w_1} = \frac{1}{||\boldsymbol{v_1}||}\, \boldsymbol{v_1}, \quad \boldsymbol{w_2} = \frac{1}{||\boldsymbol{v_2}||}\, \boldsymbol{v_2}, \quad \dots, \quad \boldsymbol{w_k} = \frac{1}{||\boldsymbol{v_n}||}\, \boldsymbol{v_n}$$

Then $\{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \dots,\, \boldsymbol{w_n}\}$ is an orthonormal basis for V.

(The process of converting an orthogonal set to an orthonormal set by multiplying each vector $\boldsymbol{w}$ by $\frac{1}{||\boldsymbol{w}||}$ is called *normalizing*.)

Proof For each i, write $\boldsymbol{v_i} = \boldsymbol{u_i} - \boldsymbol{x_i}$ where $\boldsymbol{x_i} = \displaystyle\sum_{j=1}^{i-1} \frac{\langle \boldsymbol{u_i}, \boldsymbol{v_j} \rangle}{\langle \boldsymbol{v_j}, \boldsymbol{v_j} \rangle}\, \boldsymbol{v_j} \in \operatorname{span}\{\boldsymbol{u_1}, \dots, \boldsymbol{u_{i-1}}\}$. Since $\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_i}$ are linearly independent, $\boldsymbol{v_i} \neq \boldsymbol{0}$. Thus by Remark 12.3.4.1, we only need to show that $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \dots,\, \boldsymbol{v_n}\}$ is orthogonal. We prove by using mathematical induction.

It is obvious that $\{\boldsymbol{v_1}\}$ is an orthogonal set.

Assume that $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \dots,\, \boldsymbol{v_{k-1}}\}$ is an orthogonal set. For $i \in \{1, 2, \dots, k-1\}$,

$$\begin{aligned}
\langle \boldsymbol{v_k}, \boldsymbol{v_i} \rangle &= \left\langle \boldsymbol{u_k} - \sum_{j=1}^{k-1} \frac{\langle \boldsymbol{u_k}, \boldsymbol{v_j} \rangle}{\langle \boldsymbol{v_j}, \boldsymbol{v_j} \rangle}\, \boldsymbol{v_j}\ ,\ \boldsymbol{v_i} \right\rangle\\
&= \langle \boldsymbol{u_k}, \boldsymbol{v_i} \rangle - \sum_{j=1}^{k-1} \frac{\langle \boldsymbol{u_k}, \boldsymbol{v_i} \rangle}{\langle \boldsymbol{v_j}, \boldsymbol{v_j} \rangle}\, \langle \boldsymbol{v_j}, \boldsymbol{v_i} \rangle\\
&= \langle \boldsymbol{u_k}, \boldsymbol{v_i} \rangle - \frac{\langle \boldsymbol{u_k}, \boldsymbol{v_i} \rangle}{\langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle}\, \langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle \qquad (\text{ because for } 1 \leq j \leq k-1,\ \langle \boldsymbol{v_j}, \boldsymbol{v_i} \rangle = 0 \text{ if } i \neq j\ )\\
&= 0.
\end{aligned}$$

So $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \ldots,\, \boldsymbol{v_k}\}$ is orthogonal.

By mathematical induction, $\{\boldsymbol{v_1},\, \boldsymbol{v_2},\, \ldots,\, \boldsymbol{v_n}\}$ is orthogonal.

Finally, $\{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \ldots,\, \boldsymbol{w_n}\}$ is orthonormal because

$$||\boldsymbol{w_i}|| = \left|\left| \frac{1}{||\boldsymbol{v_i}||}\,\boldsymbol{v_i} \right|\right| = \frac{1}{||\boldsymbol{v_i}||}\,||\boldsymbol{v_i}|| = 1 \quad \text{for all } i$$

and

$$\langle \boldsymbol{w_i}, \boldsymbol{w_j} \rangle = \left\langle \frac{1}{||\boldsymbol{v_i}||}\,\boldsymbol{v_i},\ \frac{1}{||\boldsymbol{v_j}||}\,\boldsymbol{v_j} \right\rangle = \frac{1}{||\boldsymbol{v_i}||\,||\boldsymbol{v_j}||}\,\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = 0 \quad \text{if } i \neq j.$$

Example 12.3.8 Consider the vector space $\mathcal{P}_2(\mathbb{R})$ equipped with the inner product defined in Example 12.3.5.5. Start with the standard basis $\{1, x, x^2\}$. By the Gram-Schmidt Process,

$$\begin{aligned}
p_1(x) &= 1, \\
p_2(x) &= x - \frac{\langle x,\, p_1(x)\rangle}{\langle p_1(x),\, p_1(x)\rangle} p_1(x) = x, \\
p_3(x) &= x^2 - \frac{\langle x^2,\, p_1(x)\rangle}{\langle p_1(x),\, p_1(x)\rangle} p_1(x) - \frac{\langle x^2,\, p_2(x)\rangle}{\langle p_2(x),\, p_2(x)\rangle} p_2(x) = -\frac{1}{3} + x^2
\end{aligned}$$

form an orthogonal basis for $\mathcal{P}_2(\mathbb{R})$. Then

$$\left\{ \frac{1}{||p_1(x)||} p_1(x),\ \frac{1}{||p_2(x)||} p_2(x),\ \frac{1}{||p_3(x)||} p_3(x) \right\} = \left\{ 1,\ \sqrt{3}\,x,\ \frac{\sqrt{5}}{2}(-1 + 3x^2) \right\}$$

is an orthonormal basis.

Section 12.4 Orthogonal Complements and Orthogonal Projections

Definition 12.4.1 Let V be an inner product space and W a subspace of V. The *orthogonal complement* of W is defined to be the set

$$\begin{aligned}
W^{\perp} &= \{\boldsymbol{v} \in V \mid \boldsymbol{v} \text{ is orthogonal to } W\} \\
&= \{\boldsymbol{v} \in V \mid \langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0 \text{ for all } \boldsymbol{u} \in W\} \subseteq V.
\end{aligned}$$

Example 12.4.2

1. In Example 12.3.5.1, $W^{\perp} = \{(0, -t, t) \mid t \in \mathbb{C}\} = \operatorname{span}\{(0, -1, 1)\}$ which is also a subspace of $\mathbb{C}^3$.

2. For any inner product space V, $V^{\perp} = \{\mathbf{0}\}$ and $\{\mathbf{0}\}^{\perp} = V$.

Theorem 12.4.3 Let V be an inner product space and W a subspace of V.

1. $W^{\perp}$ is a subspace of V.
2. $W \cap W^{\perp} = \{\mathbf{0}\}$, i.e. $W + W^{\perp}$ is a direct sum.
3. If W is finite dimensional, then $V = W \oplus W^{\perp}$.
4. If V is finite dimensional, then $\dim(V) = \dim(W) + \dim(W^{\perp})$.

Proof

1. (S1) Since $\langle \mathbf{0}, \boldsymbol{u} \rangle = 0$ for all $\boldsymbol{u} \in W$, $\mathbf{0} \in W^{\perp}$.

   (S2) Take any $\boldsymbol{v}, \boldsymbol{w} \in W^{\perp}$, i.e. $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = \langle \boldsymbol{w}, \boldsymbol{u} \rangle = 0$ for all $\boldsymbol{u} \in W$. Then

   $$\langle \boldsymbol{v} + \boldsymbol{w},\ \boldsymbol{u} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle + \langle \boldsymbol{w}, \boldsymbol{u} \rangle = 0 + 0 = 0 \quad \text{for all } \boldsymbol{u} \in W.$$

   So $\boldsymbol{v} + \boldsymbol{w} \in W^{\perp}$.

   (S3) Take any $\boldsymbol{v} \in W^{\perp}$, i.e. $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0$ for all $\boldsymbol{u} \in W$. For any scalar c,

   $$\langle c\boldsymbol{v}, \boldsymbol{u} \rangle = c\langle \boldsymbol{v}, \boldsymbol{u} \rangle = c0 = 0 \quad \text{for all } \boldsymbol{u} \in W.$$

   So $c\boldsymbol{v} \in W^{\perp}$.

   Since $W^{\perp}$ is a subset of V satisfying (S1), (S2) and (S3), $W^{\perp}$ is a subspace of V.

2. If $\boldsymbol{v} \in W \cap W^{\perp}$, then $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0$ and hence by (IP4), $\boldsymbol{v} = \mathbf{0}$. Thus $W \cap W^{\perp} = \{\mathbf{0}\}$. By Theorem 8.6.5, $W + W^{\perp}$ is a direct sum.

3. If $W = \{\mathbf{0}\}$, then $W \oplus W^{\perp} = W^{\perp} = V$.

   Suppose W is not a zero space. Let $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ be an orthonormal basis for W. For any $\boldsymbol{u} \in V$, define

   $$\boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{w_1} \rangle\, \boldsymbol{w_1} + \langle \boldsymbol{u}, \boldsymbol{w_2} \rangle\, \boldsymbol{w_2} + \cdots + \langle \boldsymbol{u}, \boldsymbol{w_k} \rangle\, \boldsymbol{w_k} \quad \text{and} \quad \boldsymbol{v'} = \boldsymbol{u} - \boldsymbol{v}.$$

   We have $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{v'}$ where $\boldsymbol{v} \in W$ and since for $i = 1, 2, \ldots, k$,

   $$\begin{aligned} \langle \boldsymbol{v'}, \boldsymbol{w_i} \rangle &= \left\langle \boldsymbol{u} - \sum_{j=1}^{k} \langle \boldsymbol{u}, \boldsymbol{w_j} \rangle\, \boldsymbol{w_j}\ ,\ \boldsymbol{w_i} \right\rangle \\ &= \langle \boldsymbol{u}, \boldsymbol{w_i} \rangle - \sum_{j=1}^{k} \langle \boldsymbol{u}, \boldsymbol{w_j} \rangle\, \langle \boldsymbol{w_j}, \boldsymbol{w_i} \rangle \\ &= \langle \boldsymbol{u}, \boldsymbol{w_i} \rangle - \langle \boldsymbol{u}, \boldsymbol{w_i} \rangle = 0, \end{aligned}$$

   by Lemma 12.3.3.1, $\boldsymbol{v'} \in W^{\perp}$. So we have shown that $V = W + W^{\perp}$ and by Part 2, $V = W \oplus W^{\perp}$.

4. If V is finite dimensional, by Part 3 and Theorem 8.6.7.2, $\dim(V) = \dim(W) + \dim(W^\perp)$.

Example 12.4.4

1. In Example 12.3.5.1 and Example 12.4.2.1, $\dim(W) = 2$ and $\dim(W^\perp) = 1$. Hence $\dim(W) + \dim(W^\perp) = 3 = \dim(\mathbb{C}^3)$.

2. Suppose $\mathcal{M}_{n\times n}(\mathbb{R})$ is equipped with the inner product defined in Example 12.1.6.3. Let $W_1 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A} \text{ is symmetric}\}$ and $W_2 = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A} \text{ is skew symmetric}\}$. By Example 12.3.5.2, we have $W_2 \subseteq W_1^\perp$.

   By Example 8.6.6.2, we have $\mathcal{M}_{n\times n}(\mathbb{R}) = W_1 \oplus W_2$ and hence

$$\begin{aligned}\dim(W_2) &= \dim(\mathcal{M}_{n\times n}(\mathbb{R})) - \dim(W_1) && \text{(by Theorem 8.6.7.2)}\\ &= \dim(W_1^\perp) && \text{(by Theorem 12.4.3.4).}\end{aligned}$$

   So Theorem 8.5.15, $W_2 = W_1^\perp$.

Remark 12.4.5 Theorem 12.4.3.3 is not always true when W is infinite dimensional.

Consider the l_2-space V defined in Example 12.1.6.6. For $i = 1, 2, 3, \dots$, define $\boldsymbol{e_i} \in V$ to be the real infinite sequence such that the ith term of the sequence is 1 and all other terms are 0. Let $W = \operatorname{span}\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \dots\}$ which is an infinite dimensional subspace of V. Note that $W \neq V$, (see Question 8.20).

Let $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}} \in W^\perp$, i.e. $\langle \boldsymbol{a}, \boldsymbol{e_i}\rangle = 0$ for all i. For each i, by the definition of the inner product of the l_2-space,

$$\langle \boldsymbol{a}, \boldsymbol{e_i}\rangle = \sum_{n=1}^{\infty} a_n(\text{the } n\text{th term of } \boldsymbol{e_i}) = a_i.$$

This means $a_i = 0$ for all i and hence $\boldsymbol{a}$ is the zero sequence.

So we have shown that $W^\perp = \{\boldsymbol{0}\}$. In this case, $W \oplus W^\perp = W \neq V$.

Theorem 12.4.6 Let V be an inner product space and W a subspace of V.

1. $W \subseteq (W^\perp)^\perp$.

2. If W is finite dimensional, then $W = (W^\perp)^\perp$.

Proof

1. Take any $\boldsymbol{u} \in W$. Then

$$\begin{aligned}\boldsymbol{u} \in W \quad &\Rightarrow \quad \langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0 \ \text{ for all } \boldsymbol{v} \in W^\perp\\ &\Rightarrow \quad \boldsymbol{u} \in (W^\perp)^\perp.\end{aligned}$$

   Thus $W \subseteq (W^\perp)^\perp$.

2. Take any $\boldsymbol{v} \in (W^{\perp})^{\perp}$. Since W is finite dimensional, by Theorem 12.4.3.3, $V = W \oplus W^{\perp}$. Write $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{w'}$ where $\boldsymbol{w} \in W$ and $\boldsymbol{w'} \in W^{\perp}$. By Part 1, $\boldsymbol{w} \in (W^{\perp})^{\perp}$ and hence $\boldsymbol{w'} = \boldsymbol{v} - \boldsymbol{w} \in (W^{\perp})^{\perp}$. But this means $\boldsymbol{w'} \in W^{\perp} \cap (W^{\perp})^{\perp}$. By Theorem 12.4.3.2, $W^{\perp} \cap (W^{\perp})^{\perp} = \{\boldsymbol{0}\}$. Thus $\boldsymbol{v} - \boldsymbol{w} = \boldsymbol{w'} = \boldsymbol{0}$, i.e. $\boldsymbol{v} = \boldsymbol{w} \in W$.

   So we have shown that $(W^{\perp})^{\perp} \subseteq W$. Together with Part 1, $(W^{\perp})^{\perp} = W$.

Remark 12.4.7 Theorem 12.4.6.2 is not always true when W is infinite dimensional.

Use W defined in Remark 12.4.5. Since $W^{\perp} = \{\boldsymbol{0}\}$, $(W^{\perp})^{\perp} = \{\boldsymbol{0}\}^{\perp} = V$. We still have $W \subseteq (W^{\perp})^{\perp}$ but $W \neq (W^{\perp})^{\perp}$.

Definition 12.4.8 Let V be an inner product space and W a subspace of V such that $V = W \oplus W^{\perp}$. (By Theorem 12.4.3.3, we know that $V = W \oplus W^{\perp}$ is always true if W is finite dimensional.) Every $\boldsymbol{u} \in V$ can be uniquely expressed as

$$\boldsymbol{u} = \boldsymbol{w} + \boldsymbol{w'} \quad \text{where } \boldsymbol{w} \in W \text{ and } \boldsymbol{w'} \in W^{\perp}.$$

The vector $\boldsymbol{w}$ is called the *orthogonal projection* of $\boldsymbol{u}$ onto W and is denoted by $\mathrm{Proj}_W(\boldsymbol{u})$.

Proposition 12.4.9 The mapping $\mathrm{Proj}_W : V \to V$ is a linear operator and is called the *orthogonal projection* of V onto W.

Proof This is only a particular case of Question 9.6.

Example 12.4.10

1. In Example 12.3.5.1, $W = \mathrm{span}\{(1, 1, 1), (1, \mathrm{i}, \mathrm{i})\} = \{(a, b, b) \mid a, b \in \mathbb{C}\}$ and $W^{\perp} = \{(0, -t, t) \mid t \in \mathbb{C}\}$.

   For any $(x, y, z) \in \mathbb{C}$,

$$(x, y, z) = \left(x,\ \tfrac{y+z}{2},\ \tfrac{y+z}{2}\right) + \left(0,\ -\left(\tfrac{z-y}{2}\right),\ \tfrac{z-y}{2}\right)$$

   where $\left(x,\ \tfrac{y+z}{2},\ \tfrac{y+z}{2}\right) \in W$ and $\left(0,\ -\left(\tfrac{z-y}{2}\right),\ \tfrac{z-y}{2}\right) \in W^{\perp}$. So

$$\mathrm{Proj}_W((x, y, z)) = \left(x,\ \tfrac{y+z}{2},\ \tfrac{y+z}{2}\right).$$

2. Suppose $\mathcal{M}_{n\times n}(\mathbb{R})$ is equipped with the inner product defined in Example 12.1.6.3. Let $W = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A} \text{ is symmetric}\}$.

   By Example 12.4.4.2, $W^{\perp} = \{\boldsymbol{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) \mid \boldsymbol{A} \text{ is skew symmetric}\}$. For each $\boldsymbol{B} \in \mathcal{M}_{n\times n}(\mathbb{R})$, by Example 8.6.6.2, we can write

$$\boldsymbol{B} = \tfrac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) + \tfrac{1}{2}(\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}})$$

   where $\frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) \in W$ and $\frac{1}{2}(\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}) \in W^{\perp}$. So $\mathrm{Proj}_W(\boldsymbol{B}) = \frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}})$.

Theorem 12.4.11 Let V be an inner product space and W a finite dimensional subspace of V. If $B = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ is an orthonormal basis for W, then for any vector $\boldsymbol{u} \in V$,

$$\mathrm{Proj}_W(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{w_1} \rangle \boldsymbol{w_1} + \langle \boldsymbol{u}, \boldsymbol{w_2} \rangle \boldsymbol{w_2} + \cdots + \langle \boldsymbol{u}, \boldsymbol{w_k} \rangle \boldsymbol{w_k}$$

and

$$\begin{aligned} \mathrm{Proj}_{W^\perp}(\boldsymbol{u}) &= \boldsymbol{u} - \mathrm{Proj}_W(\boldsymbol{u}) \\ &= \boldsymbol{u} - \langle \boldsymbol{u}, \boldsymbol{w_1} \rangle \boldsymbol{w_1} - \langle \boldsymbol{u}, \boldsymbol{w_2} \rangle \boldsymbol{w_2} - \cdots - \langle \boldsymbol{u}, \boldsymbol{w_k} \rangle \boldsymbol{w_k}. \end{aligned}$$

Proof The proof follows the same argument as the proof for Theorem 5.2.15.

Example 12.4.12

1. In Example 12.4.10.1, $W = \{(a, b, b) \mid a, b \in \mathbb{C}\} = \mathrm{span}\{(1, 0, 0), (0, 1, 1)\} \subseteq \mathbb{C}^3$. To compute $\mathrm{Proj}_W((x, y, z))$, for $(x, y, z) \in \mathbb{C}^3$, by using Theorem 12.4.11, we first need an orthonormal basis for W. In this case, $\{(1, 0, 0), \frac{1}{\sqrt{2}}(0, 1, 1)\}$ is an orthonormal basis for W. Thus

$$\begin{aligned} \mathrm{Proj}_W((x, y, z)) &= \langle (x, y, z), (1, 0, 0) \rangle (1, 0, 0) + \left\langle (x, y, z), \tfrac{1}{\sqrt{2}}(0, 1, 1) \right\rangle \tfrac{1}{\sqrt{2}}(0, 1, 1) \\ &= x(1, 0, 0) + \tfrac{y+z}{2}(0, 1, 1) = \left(x, \tfrac{y+z}{2}, \tfrac{y+z}{2}\right) \end{aligned}$$

which gives us the same formula as in Example 12.4.10.1.

2. Let $C([-1, 1])$ be equipped with the inner product defined by

$$\langle f, g \rangle = \tfrac{1}{2} \int_{-1}^{1} f(t)g(t)dt \quad \text{for } f, g \in C([-1, 1]).$$

Since real polynomials can be regarded as functions in $C([-1, 1])$, $W_1 = \mathcal{P}_1(\mathbb{R})$ and $W_2 = \mathcal{P}_2(\mathbb{R})$ can be regarded as a subspace of $C([-1, 1])$.

(a) Using the orthonormal basis $\left\{1, \sqrt{3}\,x\right\}$ for W_1, for any $f \in C([-1, 1])$,

$$\mathrm{Proj}_{W_1}(f) = \left(\tfrac{1}{2} \int_{-1}^{1} f(t)d(t)\right) + \left(\tfrac{3}{2} \int_{-1}^{1} tf(t)dt\right)x.$$

In particular, if $f(x) = e^x$ for $x \in [-1, 1]$, then

$$\mathrm{Proj}_{W_1}(f) = \tfrac{1}{2}\left(e - \tfrac{1}{e}\right) + \tfrac{3}{e}\,x.$$

(b) Using the orthonormal basis $\left\{1, \sqrt{3}\,x, \frac{\sqrt{5}}{2}(-1 + 3x^2)\right\}$ for W_2 (see Example 12.3.8), for any $f \in C([-1, 1])$,

$$\mathrm{Proj}_{W_2}(f) = \left(\tfrac{1}{2} \int_{-1}^{1} f(t)d(t)\right) + \left(\tfrac{3}{2} \int_{-1}^{1} tf(t)dt\right)x + \left(\tfrac{5}{8} \int_{-1}^{1} (-1 + 3t^2)f(t)dt\right)(-1 + 3x^2).$$

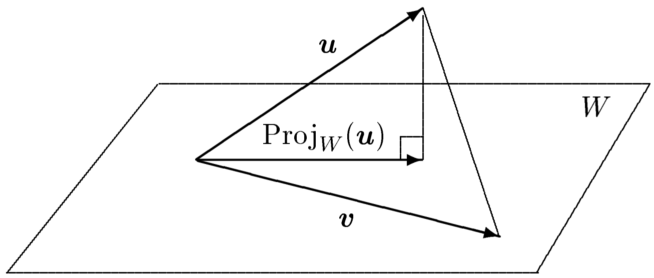
In particular, if $f(x) = e^x$ for $x \in [-1, 1]$, then

$$\mathrm{Proj}_{W_2}(f) = \tfrac{3}{4}\left(-e + \tfrac{11}{e}\right) + \tfrac{3}{e}\,x + \tfrac{15}{4}\left(e - \tfrac{7}{e}\right)x^2.$$

Theorem 12.4.13 (Best Approximation) Let V be an inner product space and W a subspace of V such that $V = W \oplus W^{\perp}$. Then for any $\boldsymbol{u} \in V$,

$$d(\boldsymbol{u}, \operatorname{Proj}_W(\boldsymbol{u})) \le d(\boldsymbol{u}, \boldsymbol{w}) \quad \text{for all } \boldsymbol{w} \in W,$$

i.e. $\operatorname{Proj}_W(\boldsymbol{u})$ is the *best approximation* of $\boldsymbol{u}$ in W.



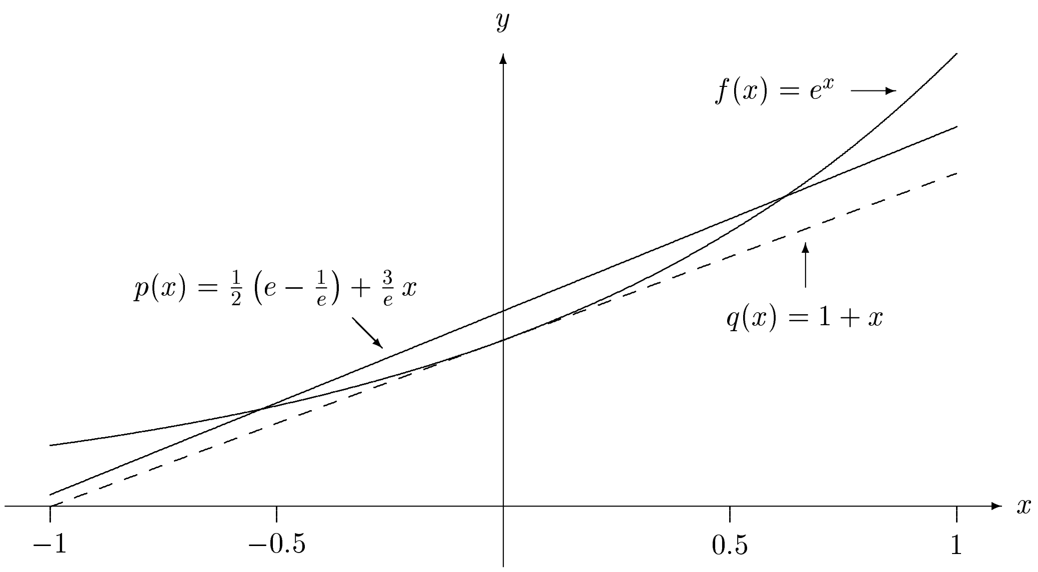
Proof The proof follows the same argument as the proof for Theorem 5.3.2.

Example 12.4.14 Let $C([-1,1])$ be equipped with the inner product defined by

$$\langle f, g\rangle = \tfrac{1}{2}\int_{-1}^{1} f(t)g(t)dt \quad \text{for } f, g \in C([-1,1]).$$

By Example 12.4.12.2(a), the best approximation of the exponential function $f(x) = e^x$ in $\mathcal{P}_1(\mathbb{R})$ is $p(x) = \frac{1}{2}\left(e - \frac{1}{e}\right) + \frac{3}{e}x$.

In the following diagram, we compare $p(x)$ with the approximation $q(x) = 1 + x$ computed using Taylor's expansion of e^x about $x = 0$.



Section 12.5 Adjoints of Linear Operators

Definition 12.5.1 Let V be an inner product space and let T be a linear operator on V. A linear operator T^* on V is called the *adjoint* of T if

$$\langle T(\boldsymbol{u}),\, \boldsymbol{v}\rangle = \langle \boldsymbol{u},\, T^*(\boldsymbol{v})\rangle \quad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in V. \tag{12.3}$$

In Theorem 12.5.4.1, we shall learn that the adjoint of a linear operator is unique if it exists. Thus we always use T^* to denote the adjoint of T.

Please note that the (classical) adjoint of a matrix defined in Definition 2.5.24 is a completely different concept.

Remark 12.5.2 Let V be an inner product space and let T be a linear operator on V such that T^* exists. Then for all $\boldsymbol{u}, \boldsymbol{v} \in V$,

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

Example 12.5.3

1. Let V be an inner product space. For all $\boldsymbol{u}, \boldsymbol{v} \in V$,

   $$\langle I_V(\boldsymbol{u}),\, \boldsymbol{v}\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{u},\, I_V(\boldsymbol{v})\rangle \quad \text{and} \quad \langle O_V(\boldsymbol{u}),\, \boldsymbol{v}\rangle = 0 = \langle \boldsymbol{u},\, O_V(\boldsymbol{v})\rangle.$$

   Both I_V and O_V are the adjoints of themselves, i.e $I_V^* = I_V$ and $O_V^* = O_V$.

2. Let $\mathbb{R}^n$ be equipped with the usual inner product. Given any $n \times n$ real matrix $\boldsymbol{A}$, let $L_{\boldsymbol{A}}$ be the linear operator on $\mathbb{R}^n$ as defined in Example 9.1.4.1, i.e. $L_{\boldsymbol{A}}(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for $\boldsymbol{u} \in \mathbb{R}^n$. (In here, all vectors are expressed as column vectors.) For any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

   $$\langle L_{\boldsymbol{A}}(\boldsymbol{u}),\, \boldsymbol{v}\rangle = \langle \boldsymbol{A}\boldsymbol{u},\, \boldsymbol{v}\rangle = (\boldsymbol{A}\boldsymbol{u})^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} = \langle \boldsymbol{u},\, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\rangle = \langle \boldsymbol{u},\, L_{\boldsymbol{A}^{\mathrm{T}}}(\boldsymbol{v})\rangle.$$

   Thus the adjoint of $L_{\boldsymbol{A}}$ is $L_{\boldsymbol{A}^{\mathrm{T}}}$, i.e. $L_{\boldsymbol{A}}^* = L_{\boldsymbol{A}^{\mathrm{T}}}$.

3. Let $\mathbb{C}^n$ be equipped with the usual inner product. Following the same arguments as above, $L_{\boldsymbol{A}}^* = L_{\boldsymbol{A}^*}$ for any $n \times n$ complex matrix $\boldsymbol{A}$.

4. Consider the l_2-space V defined in Example 12.1.6.6. Let S be the shift operator on V as defined in Example 11.1.3.2, i.e.

   $$S((a_n)_{n\in\mathbb{N}}) = (a_{n+1})_{n\in\mathbb{N}} \quad \text{for } (a_n)_{n\in\mathbb{N}} \in V.$$

   Define $T : V \to V$ such that for $(a_n)_{n\in\mathbb{N}} \in V$, $T((a_n)_{n\in\mathbb{N}}) = (a'_n)_{n\in\mathbb{N}}$ where

   $$a'_n = \begin{cases} 0 & \text{if } n = 1 \\ a_{n-1} & \text{if } n > 1. \end{cases}$$

Then for all $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}} \in V$,

$$\begin{aligned}\langle (a_n)_{n\in\mathbb{N}},\, T((b_n)_{n\in\mathbb{N}})\rangle = \sum_{n=1}^{\infty} a_n b'_n &= \sum_{n=2}^{\infty} a_n b_{n-1} \\ &= \sum_{n=1}^{\infty} a_{n+1} b_n = \langle S((a_n)_{n\in\mathbb{N}}),\, (b_n)_{n\in\mathbb{N}}\rangle.\end{aligned}$$

Thus T is the adjoint of S.

Theorem 12.5.4 Let V be an inner product space and let T be a linear operator on V.

1. The adjoint of T is unique if it exists.
2. Suppose V is finite dimensional where $\dim(V) \geq 1$.
   (a) T^* always exists.
   (b) If B is an ordered orthonormal basis for V, then $[T^*]_B = ([T]_B)^*$.
   (c) $\operatorname{rank}(T) = \operatorname{rank}(T^*)$ and $\operatorname{nullity}(T) = \operatorname{nullity}(T^*)$.

Proof

1. Suppose there are two adjoints T_1 and T_2 of T. To show that $T_1 = T_2$, we need to show that $T_1(\boldsymbol{v}) = T_2(\boldsymbol{v})$ for all $\boldsymbol{v} \in V$. By (12.3), for all $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\langle \boldsymbol{u},\, T_1(\boldsymbol{v})\rangle = \langle T(\boldsymbol{u}),\, \boldsymbol{v}\rangle = \langle \boldsymbol{u},\, T_2(\boldsymbol{v})\rangle \quad \Rightarrow \quad \langle \boldsymbol{u},\, T_1(\boldsymbol{v}) - T_2(\boldsymbol{v})\rangle = 0.$$

   Substituting $\boldsymbol{u}$ by $T_1(\boldsymbol{v}) - T_2(\boldsymbol{v})$, we have

$$\langle T_1(\boldsymbol{v}) - T_2(\boldsymbol{v}),\, T_1(\boldsymbol{v}) - T_2(\boldsymbol{v})\rangle = 0 \quad \text{for all } \boldsymbol{v} \in V.$$

   But by (IP4), $T_1(\boldsymbol{v}) - T_2(\boldsymbol{v}) = \mathbf{0}$ and hence $T_1(\boldsymbol{v}) = T_2(\boldsymbol{v})$ for all $\boldsymbol{v} \in V$.

2. Suppose V is finite dimensional. Let $B = \{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \ldots,\, \boldsymbol{w_n}\}$ be an ordered orthonormal basis for V. Define an linear operator $T' : V \to V$ such that $[T']_B = ([T]_B)^*$, i.e. for every $a_1\boldsymbol{w_1} + a_2\boldsymbol{w_2} + \cdots + a_n\boldsymbol{w_n} \in V$,

$$T'(a_1\boldsymbol{w_1}+a_2\boldsymbol{w_2}+\cdots+a_n\boldsymbol{w_n}) = b_1\boldsymbol{w_1}+b_2\boldsymbol{w_2}+\cdots+b_n\boldsymbol{w_n} \quad \text{where} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = ([T]_B)^* \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

   By Lemma 12.3.3.3, for all $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\begin{aligned}\langle T(\boldsymbol{u}),\, \boldsymbol{v}\rangle = ([T(\boldsymbol{u})]_B)^{\mathrm{T}}\, \overline{[\boldsymbol{v}]}_B &= ([T]_B\, [\boldsymbol{u}]_B)^{\mathrm{T}}\, \overline{[\boldsymbol{v}]}_B \\ &= ([\boldsymbol{u}]_B)^{\mathrm{T}}\, ([T]_B)^{\mathrm{T}}\, \overline{[\boldsymbol{v}]}_B \\ &= ([\boldsymbol{u}]_B)^{\mathrm{T}}\, \overline{([T]_B)^*\, [\boldsymbol{v}]_B} = ([\boldsymbol{u}]_B)^{\mathrm{T}}\, \overline{[T'(\boldsymbol{v})]}_B = \langle \boldsymbol{u},\, T'(\boldsymbol{v})\rangle.\end{aligned}$$

Thus T' is the adjoint of T, i.e. T^* $(=T')$ exists and $[T^*]_B = ([T]_B)^*$.

For any complex (or real) matrices $\boldsymbol{A}$, $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^*)$ (see Question 12.1). So

$$\operatorname{rank}(T) = \operatorname{rank}([T]_B) = \operatorname{rank}(([T]_B)^*) = \operatorname{rank}([T^*]_B) = \operatorname{rank}(T^*)$$

(see Question 12.35 for a proof of $\operatorname{rank}(T) = \operatorname{rank}(T^*)$ without using matrices) and

$$\operatorname{nullity}(T) = \dim(V) - \operatorname{rank}(T) = \dim(V) - \operatorname{rank}(T^*) = \operatorname{nullity}(T^*).$$

Example 12.5.5 Let $\mathbb{R}^3$ be equipped with the usual inner product and let T be the linear operator on $\mathbb{R}^3$ defined by

$$T((x,y,z)) = (x,\ x+y,\ x+y+z) \quad \text{for } (x,y,z) \in \mathbb{R}^3.$$

Let $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ be the standard basis for $\mathbb{R}^3$ which is an orthonormal basis for $\mathbb{R}^3$. Then

$$[T]_E = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad ([T]_E)^* = ([T]_E)^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

As $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y+z \\ y+z \\ z \end{pmatrix}$, the adjoint T^* of T is the linear operator defined by

$$T^*((x,y,z)) = (x+y+z)\boldsymbol{e_1} + (y+z)\boldsymbol{e_2} + z\boldsymbol{e_3} = (x+y+z,\ y+z,\ z) \quad \text{for } (x,y,z) \in \mathbb{R}^3.$$

Note that for $(x_1,y_1,z_1),\ (x_2,y_2,z_2) \in \mathbb{R}^3$,

$$\begin{aligned} \langle T((x_1,y_1,z_1)),\ (x_2,y_2,z_2)\rangle &= \langle (x_1,\ x_1+y_1,\ x_1+y_1+z_1),\ (x_2,y_2,z_2)\rangle \\ &= x_1x_2 + (x_1+y_1)y_2 + (x_1+y_1+z_1)z_2 \\ &= x_1(x_2+y_2+z_2) + y_1(y_2+z_2) + z_1z_2 \\ &= \langle (x_1,y_1,z_1),\ (x_2+y_2+z_2,\ y_2+z_2,\ z_2)\rangle \\ &= \langle (x_1,y_1,z_1),\ T^*((x_2,y_2,z_2))\rangle. \end{aligned}$$

Remark 12.5.6 Theorem 12.5.4.2(a) is not always true when V is infinite dimensional.

Let V be the l_2-space defined in Example 12.1.6.6. As in Remark 12.4.5, consider the subspace $W = \operatorname{span}\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \dots\}$ of V. Observe that W consists of all real infinite sequences which has only finite number of nonzero terms. Define a linear operator T on W such that

$$T((a_n)_{n\in\mathbb{N}}) = \left(\sum_{i=n}^{\infty} a_i\right)_{n\in\mathbb{N}} \quad \text{for } (a_n)_{n\in\mathbb{N}} \in W.$$

(The sum $\displaystyle\sum_{i=n}^{\infty} a_i$ converges because there are only finite number of nonzero a_i.)

Note that $T(\boldsymbol{e_n}) = (1, 1, \dots, 1, 0, 0, \dots)$ where the first n entries are 1 and all other entries are 0.

Assume T^* exists. Let $T^*(\boldsymbol{e_1}) = (b_n)_{n\in\mathbb{N}}$. By (12.3), for all $n \in \mathbb{N}$,

$$b_n = \langle \boldsymbol{e_n},\, T^*(\boldsymbol{e_1})\rangle = \langle T(\boldsymbol{e_n}),\, \boldsymbol{e_1}\rangle = 1.$$

But then $T^*(\boldsymbol{e_1})$ is not a sequence in W which contradicts that T^* is an linear operator on W. (Actually, $T^*(\boldsymbol{e_1})$ is not even a sequence in V.)

Hence the adjoint of T does not exist.

Proposition 12.5.7 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and let V be an inner product space over $\mathbb{F}$. Suppose S and T are linear operators on V such that S^* and T^* exists. Then

1. $(S + T)^*$ exists and $(S + T)^* = S^* + T^*$;
2. for any $c \in \mathbb{F}$, $(cT)^*$ exists and $(cT)^* = \bar{c}T^*$;
3. $(S \circ T)^*$ exists and $(S \circ T)^* = T^* \circ S^*$;
4. $(T^*)^*$ exists and $(T^*)^* = T$; and
5. if W is a subspace of V which is both T-invariant and T^*-invariant, then $(T|_W)^*$ exists and $(T|_W)^* = T^*|_W$.

Proof We only show the proof of Part 3:

(Note that before we have shown that the adjoint of $S \circ T$ exists, we cannot use the term $(S \circ T)^*$.) For all $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\begin{aligned}
\langle (S \circ T)(\boldsymbol{u}),\, \boldsymbol{v}\rangle &= \langle S(T(\boldsymbol{u})),\, \boldsymbol{v}\rangle \\
&= \langle T(\boldsymbol{u}),\, S^*(\boldsymbol{v})\rangle \\
&= \langle \boldsymbol{u},\, T^*(S^*(\boldsymbol{v}))\rangle = \langle \boldsymbol{u},\, (T^* \circ S^*)(\boldsymbol{v})\rangle.
\end{aligned}$$

So $(S \circ T)^*$ exists and $(S \circ T)^* = T^* \circ S^*$.

(Proofs of the other parts are left as exercises. See Question 12.31).

Definition 12.5.8 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

1. Let V be an inner product space over $\mathbb{F}$ and T a linear operator on V such that T^* exists. Suppose T is invertible and $T^{-1} = T^*$, i.e. $T \circ T^* = T^* \circ T = I_V$.
   (a) If $\mathbb{F} = \mathbb{C}$, then T is called a *unitary operator*.
   (b) If $\mathbb{F} = \mathbb{R}$, then T is called an *orthogonal operator*.
2. Let $\boldsymbol{A}$ be an invertible matrix over $\mathbb{F}$ such that $\boldsymbol{A}^{-1} = \boldsymbol{A}^*$, i.e. $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}$.

(a) If $\mathbb{F} = \mathbb{C}$, then $\boldsymbol{A}$ is called a *unitary matrix*.

(b) If $\mathbb{F} = \mathbb{R}$, then $\boldsymbol{A}$ is called an *orthogonal matrix*. Note that only real square matrices satisfying $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$ are called orthogonal matrices. (See also Section 5.4.)

An orthogonal matrix is also a unitary matrix.

Proposition 12.5.9 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$, and T a linear operator on V. Take any ordered orthonormal basis B for V.

1. If $\mathbb{F} = \mathbb{C}$, then T is unitary if and only if $[T]_B$ is a unitary matrix.

2. If $\mathbb{F} = \mathbb{R}$, then T is orthogonal if and only if $[T]_B$ is an orthogonal matrix.

Proof Since the proof of Part 2 is the same as Part 1, we only prove Part 1:

Let $\boldsymbol{A} = [T]_B$. By Theorem 12.5.4.2, $[T^*]_B = \boldsymbol{A}^*$. Thus by Theorem 9.3.3, $[T \circ T^*]_B = \boldsymbol{A}\boldsymbol{A}^*$ and $[T^* \circ T]_B = \boldsymbol{A}^*\boldsymbol{A}$.

Hence

$$\begin{aligned} T \text{ is unitary} \quad &\Leftrightarrow \quad T \circ T^* = T^* \circ T = I_V \\ &\Leftrightarrow \quad \boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I} \qquad \text{(by Lemma 9.2.3)} \\ &\Leftrightarrow \quad \boldsymbol{A} \text{ is unitary.} \end{aligned}$$

Example 12.5.10

1. The following are some examples of unitary matrices where the first three matrices are also orthogonal matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

$$\begin{pmatrix} \frac{1}{\sqrt{5}}(1+2\mathrm{i}) & 0 & 0 \\ 0 & 0 & \mathrm{i} \\ 0 & -\mathrm{i} & 0 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{\sqrt{3}}(1+\mathrm{i}) & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}(1-\mathrm{i}) \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}}\mathrm{i} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}}\mathrm{i} \\ \frac{1}{\sqrt{2}}\mathrm{i} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{2}}\mathrm{i} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

2. Let $\mathbb{R}^2$ be equipped with the usual inner product. Consider the linear operator $L_{\boldsymbol{A}}$ on $\mathbb{R}^2$ where $\boldsymbol{A} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ for some $\theta \in [0, 2\pi]$.

   Using the orthonormal basis $E = \{(1,0), (0,1)\}$ for $\mathbb{R}^2$, $[L_{\boldsymbol{A}}]_E = \boldsymbol{A}$. Since $\boldsymbol{A}$ is an orthogonal matrix, by Proposition 12.5.9, $L_{\boldsymbol{A}}$ is an orthogonal operator.

3. Let $\mathbb{C}^3$ be equipped with the usual inner product. Let T be the linear operator on $\mathbb{C}^3$ defined by

$$T((x, y, z)) = \left(\tfrac{1}{\sqrt{2}}(x + \mathrm{i}y),\ \tfrac{1}{\sqrt{2}}(x - \mathrm{i}y),\ z\right) \quad \text{for } (x, y, z) \in \mathbb{C}^3.$$

Using the orthonormal basis $E = \{(1, 0, 0),\ (0, 1, 0)\ (0, 0, 1)\}$ for $\mathbb{C}^3$,

$$[T]_E = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\mathrm{i} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\mathrm{i} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It can be checked that $[T]_E$ is a unitary matrix and hence by Proposition 12.5.9, T is unitary.

Theorem 12.5.11 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$, and T a linear operator on V. Then the following are equivalent:

1. T is unitary (when $\mathbb{F} = \mathbb{C}$) or orthogonal (when $\mathbb{F} = \mathbb{R}$).
2. For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle T(\boldsymbol{u}),\, T(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$.
3. For all $\boldsymbol{u} \in V$, $||T(\boldsymbol{u})|| = ||\boldsymbol{u}||$.
4. There exists an orthonormal basis $\{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \dots,\, \boldsymbol{w_n}\}$ for V, where $n = \dim(V)$, such that $\{T(\boldsymbol{w_1}),\, T(\boldsymbol{w_2}),\, \dots,\, T(\boldsymbol{w_n})\}$ is also orthonormal.

Proof

$(1 \Rightarrow 2)$ For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle T(\boldsymbol{u}),\, T(\boldsymbol{v})\rangle = \langle \boldsymbol{u},\, T^*(T(\boldsymbol{v}))\rangle = \langle \boldsymbol{u},\, I_V(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$.

$(2 \Rightarrow 3)$ For all $\boldsymbol{u} \in V$, $||T(\boldsymbol{u})|| = \sqrt{\langle T(\boldsymbol{u}),\, T(\boldsymbol{u})\rangle} = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u}\rangle} = ||\boldsymbol{u}||$.

$(3 \Rightarrow 4)$ Take any orthonormal basis $\{\boldsymbol{w_1},\, \boldsymbol{w_2},\, \dots,\, \boldsymbol{w_n}\}$ for V. Since $||T(\boldsymbol{w_i})|| = ||\boldsymbol{w_i}|| = 1$ for all i, it suffices to show that $\langle T(\boldsymbol{w_i}),\, T(\boldsymbol{w_j})\rangle = 0$ for $i \neq j$. Let $\varepsilon = \langle T(\boldsymbol{w_i}),\, T(\boldsymbol{w_j})\rangle$. Then $\overline{\varepsilon} = \langle T(\boldsymbol{w_j}),\, T(\boldsymbol{w_i})\rangle$.

$$\begin{aligned} & ||T(\boldsymbol{w_i} + \varepsilon\boldsymbol{w_j})|| = ||\boldsymbol{w_i} + \varepsilon\boldsymbol{w_j}|| \\ \Rightarrow\ & \langle T(\boldsymbol{w_i}) + \varepsilon T(\boldsymbol{w_j}),\, T(\boldsymbol{w_i}) + \varepsilon T(\boldsymbol{w_j})\rangle = \langle \boldsymbol{w_i} + \varepsilon\boldsymbol{w_j},\, \boldsymbol{w_i} + \varepsilon\boldsymbol{w_j}\rangle \\ \Rightarrow\ & 1 + \overline{\varepsilon}\langle T(\boldsymbol{w_i}),\, T(\boldsymbol{w_j})\rangle + \varepsilon\langle T(\boldsymbol{w_j}),\, T(\boldsymbol{w_i})\rangle + \varepsilon\overline{\varepsilon} = 1 + \varepsilon\overline{\varepsilon} \\ \Rightarrow\ & 2\varepsilon\overline{\varepsilon} = 0 \\ \Rightarrow\ & \varepsilon = 0. \end{aligned}$$

Thus $\langle T(\boldsymbol{w_i}),\, T(\boldsymbol{w_j})\rangle = 0$.

$(4 \Rightarrow 1)$ Let $B = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_n}\}$ be an orthonormal basis for V such that $\{T(\boldsymbol{w_1}), T(\boldsymbol{w_2}), \dots, T(\boldsymbol{w_n})\}$ is also orthonormal. For $i = 1, 2, \dots, n$, by Theorem 12.3.6,

$$\begin{aligned}
&(T^* \circ T)(\boldsymbol{w_i}) \\
&= \langle (T^* \circ T)(\boldsymbol{w_i}), \boldsymbol{w_1} \rangle \boldsymbol{w_1} + \langle (T^* \circ T)(\boldsymbol{w_i}), \boldsymbol{w_2} \rangle \boldsymbol{w_2} + \cdots + \langle (T^* \circ T)(\boldsymbol{w_i}), \boldsymbol{w_n} \rangle \boldsymbol{w_n} \\
&= \langle T^*(T(\boldsymbol{w_i})), \boldsymbol{w_1} \rangle \boldsymbol{w_1} + \langle T^*(T(\boldsymbol{w_i})), \boldsymbol{w_2} \rangle \boldsymbol{w_2} + \cdots + \langle T^*(T(\boldsymbol{w_i})), \boldsymbol{w_n} \rangle \boldsymbol{w_n} \\
&= \langle T(\boldsymbol{w_i}), T(\boldsymbol{w_1}) \rangle \boldsymbol{w_1} + \langle T(\boldsymbol{w_i}), T(\boldsymbol{w_2}) \rangle \boldsymbol{w_2} + \cdots + \langle T(\boldsymbol{w_i}), T(\boldsymbol{w_n}) \rangle \boldsymbol{w_n} \\
&= \boldsymbol{w_i}.
\end{aligned}$$

This means that $T^* \circ T = I_V$. By Theorem 9.6.8, $T^{-1} = T^*$. So T is unitary if $\mathbb{F} = \mathbb{C}$ and T is orthogonal if $\mathbb{F} = \mathbb{R}$.

Example 12.5.12

1. In Example 12.5.10.2, $L_{\boldsymbol{A}}$ maps the standard basis $\{(1, 0)^{\mathrm{T}}, (0, 1)^{\mathrm{T}}\}$ to
$$\{(\cos(\theta), \sin(\theta))^{\mathrm{T}}, \ (-\sin(\theta), \cos(\theta))^{\mathrm{T}}\}$$
which is also an orthonormal basis for $\mathbb{R}^2$ using the usual inner product.

2. In Example 12.5.10.3, T maps the standard basis $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ to
$$\left\{ \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0 \right), \ \left(\tfrac{1}{\sqrt{2}}\mathrm{i}, -\tfrac{1}{\sqrt{2}}\mathrm{i}, 0 \right), \ (0, 0, 1) \right\}$$
which is also an orthonormal basis for $\mathbb{C}^3$ using the usual inner procuct.

Remark 12.5.13 Theorem 12.5.11 is not always true when V is infinite dimensional.

The linear operator S^* in Example 12.5.3.4 satisfies Part 2 and Part 3 of Theorem 12.5.11. But S^* is not surjective and hence not invertible. So S^* is not orthogonal.

Theorem 12.5.14 Let $\boldsymbol{A}$ be an $n \times n$ complex matrix. Suppose $\mathbb{C}^n$ is equipped with the usual inner product. The following statements are equivalent:

1. $\boldsymbol{A}$ is unitary.

2. The rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{C}^n$.

3. The columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{C}^n$.

Proof The proof follows the same argument as the proof for Theorem 5.4.6. (We can also prove the theorem by applying Theorem 12.5.11 to $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{A}^*}$.)

Theorem 12.5.15 Let V be a complex finite dimensional inner product space where $\dim(V) \geq 1$. If B and C are ordered orthonormal bases for V, then the transition matrix from B to C is a unitary matrix, i.e. $[I_V]_{B,C} = (\,[I_V]_{C,B})^{-1} = ([I_V]_{C,B})^*$. (See also Theorem 5.4.7.)

Proof Let $B = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ and $C = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_n}\}$. Define a linear operator T on V such that $T(\boldsymbol{w_i}) = \boldsymbol{v_i}$ for $i = 1, 2, \dots, n$. Note that

$$[T]_{B,C} = \Big([T(\boldsymbol{w_1})]_B \quad [T(\boldsymbol{w_2})]_B \quad \cdots \quad [T(\boldsymbol{w_n})]_B \Big) = \Big([\boldsymbol{v_1}]_B \quad [\boldsymbol{v_2}]_B \quad \cdots \quad [\boldsymbol{v_n}]_B \Big) = \boldsymbol{I_n}.$$

By Theorem 12.5.11, T is a unitary operator and hence by Proposition 12.5.9, $[T]_C$ is a unitary matrix. But then the transition matrix from B to C is

$$[I_V]_{C,B} = [I_V]_{C,B}\boldsymbol{I_n} = [I_V]_{C,B}[T]_{B,C} = [I_V \circ T]_{C,C} = [T]_C$$

which is a unitary matrix.

Section 12.6 Unitary and Orthogonal Diagonalization

Definition 12.6.1

1. Let V be an inner product space and let T be a linear operator on V such that T^* exists.

   (a) T is called a *self-adjoint operator* if $T = T^*$.

   (b) T is called a *normal operator* if $T \circ T^* = T^* \circ T$.

   All self-adjoint operators, orthogonal operators and unitary operators are normal.

2. Let $\boldsymbol{A}$ be a complex square matrix.

   (a) $\boldsymbol{A}$ is called a *Hermitian matrix* if $\boldsymbol{A} = \boldsymbol{A}^*$.

   Note that a real matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A} = \boldsymbol{A}^*$ is a symmetric matrix.

   (b) $\boldsymbol{A}$ is called a *normal matrix* if $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.

   All Hermitian matrices, real symmetric matrices, unitary matrices and orthogonal matrices are normal.

Proposition 12.6.2 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$, and T a linear operator on V. Take an ordered orthonormal basis B for V and let $\boldsymbol{A} = [T]_B$.

1. If $\mathbb{F} = \mathbb{C}$, T is self-adjoint if and only if $\boldsymbol{A}$ is a Hermitian matrix.

   If $\mathbb{F} = \mathbb{R}$, then T is self-adjoint if and only if $\boldsymbol{A}$ is a symmetric matrix.

2. T is normal if and only if $\boldsymbol{A}$ is a normal matrix

Proof The proof follows the same argument as the proof for Proposition 12.5.9.

Example 12.6.3

1. Let $\mathbb{C}^2$ be equipped with the usual inner product. Let T be the linear operator on $\mathbb{C}^2$ defined by
$$T((x, y)) = (2x - \mathrm{i}(x + y),\ 2y - \mathrm{i}(x + y)) \quad \text{for } (x, y) \in \mathbb{C}^2.$$
Using the orthonormal basis $E = \{(1, 0),\ (0, 1)\}$ for $\mathbb{C}^2$, $[T]_E = \begin{pmatrix} 2 - \mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & 2 - \mathrm{i} \end{pmatrix}$. Since
$$([T]_E)^* [T]_E = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} = [T]_E ([T]_E)^*$$
$[T]_E$ is a normal matrix and hence by Proposition 12.6.2.2, T is a normal operator.

2. Let $\mathbb{R}^n$ be equipped with the usual inner product and $\boldsymbol{A}$ a real $n \times n$ matrix.

   If $\boldsymbol{A}$ is symmetric, then the linear operator $L_{\boldsymbol{A}}$ is self-adjoint (and normal).

   If $\boldsymbol{A}$ is nonzero and skew symmetric, then $L_{\boldsymbol{A}}$ is normal but not self-adjoint.

Lemma 12.6.4 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V an inner product space over $\mathbb{F}$ and T a normal operator on V.

1. For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle T(\boldsymbol{u}),\, T(\boldsymbol{v})\rangle = \langle T^*(\boldsymbol{u}),\, T^*(\boldsymbol{v})\rangle$.

2. For any $c \in \mathbb{F}$, the linear operator $T - cI_V$ is normal.

3. If $\boldsymbol{u}$ is an eigenvector of T associated with λ, then $\boldsymbol{u}$ is an eigenvector of T^* associated with $\overline{\lambda}$.

4. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvector of T associated with λ and μ, respectively, where $\lambda \neq \mu$, then $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal.

Proof

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

2. Note that $I_V^* = I_V$. By Proposition 12.5.7, $(T - cI_V)^*$ exists and $(T - cI_V)^* = T^* - \bar{c}I_V$. Since
$$\begin{aligned} (T - cI_V) \circ (T - cI_V)^* &= (T - cI_V) \circ (T^* - \bar{c}I_V) \\ &= T \circ T^* - \bar{c}\,T - c\,T^* + c\bar{c}I_V \\ &= T^* \circ T - \bar{c}\,T - c\,T^* + c\bar{c}I_V \quad \text{(because } T \text{ is normal)} \\ &= (T^* - \bar{c}I_V) \circ (T - cI_V) \\ &= (T - cI_V)^* \circ (T - cI_V), \end{aligned}$$
$T - cI_V$ is normal.

3. Since $\boldsymbol{u}$ is an eigenvector of T associated with λ, $T(\boldsymbol{u}) = \lambda\boldsymbol{u}$. So $(T - \lambda I_V)(\boldsymbol{u}) = T(\boldsymbol{u}) - \lambda\boldsymbol{u} = \boldsymbol{0}$ and by Part 1 and Part 2,
$$\langle (T - \lambda I_V)^*(\boldsymbol{u}),\, (T - \lambda I_V)^*(\boldsymbol{u})\rangle = \langle (T - \lambda I_V)(\boldsymbol{u}),\, (T - \lambda I_V)(\boldsymbol{u})\rangle = 0.$$
By (IP4), $(T - \lambda I_V)^*(\boldsymbol{u}) = \boldsymbol{0}$ and hence
$$T^*(\boldsymbol{u}) - \overline{\lambda}\boldsymbol{u} = (T^* - \overline{\lambda} I_V)(\boldsymbol{u}) = (T - \lambda I_V)^*(\boldsymbol{u}) = \boldsymbol{0},$$
i.e. $T^*(\boldsymbol{u}) = \overline{\lambda}\boldsymbol{u}$. So $\boldsymbol{u}$ is an eigenvector of T^* associated with $\overline{\lambda}$.

4. Note that $T(\boldsymbol{u}) = \lambda\boldsymbol{u}$ and $T(\boldsymbol{v}) = \mu\boldsymbol{v}$. By Part 3, $T^*(\boldsymbol{v}) = \overline{\mu}\boldsymbol{v}$. Then
$$\lambda\langle\boldsymbol{u}, \boldsymbol{v}\rangle = \langle\lambda\boldsymbol{u},\, \boldsymbol{v}\rangle = \langle T(\boldsymbol{u}),\, \boldsymbol{v}\rangle = \langle\boldsymbol{u},\, T^*(\boldsymbol{v})\rangle = \langle\boldsymbol{u},\, \overline{\mu}\boldsymbol{v}\rangle = \mu\langle\boldsymbol{u}, \boldsymbol{v}\rangle.$$
Since $\lambda \neq \mu$, $\langle\boldsymbol{u}, \boldsymbol{v}\rangle = 0$.

Remark 12.6.5 The results in Lemma 12.6.4 is still true if we replace V by $\mathbb{C}^n$ (equipped with the usual inner product) and T by an $n \times n$ normal matrix $\boldsymbol{A}$. In particular, if $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with λ, then $\boldsymbol{u}$ is also an eigenvector of $\boldsymbol{A}^*$ associated with $\overline{\lambda}$.

Example 12.6.6 Let $\boldsymbol{A} = \begin{pmatrix} 2-\mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & 2-\mathrm{i} \end{pmatrix}$. By Example 12.6.3.1, $\boldsymbol{A}$ is a normal matrix. Let $\boldsymbol{u} = (1, -1)^{\mathrm{T}}$ and $\boldsymbol{v} = (1, 1)^{\mathrm{T}}$. Since
$$\begin{pmatrix} 2-\mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & 2-\mathrm{i} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2-\mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & 2-\mathrm{i} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = (2-2\mathrm{i})\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors of $\boldsymbol{A}$ associated with the eigenvalues 2 and $2 - 2\mathrm{i}$ respectively. Thus $\boldsymbol{u}$ and $\boldsymbol{v}$ are also eigenvectors of $\boldsymbol{A}^*$ associated with the eigenvalues 2 and $2 + 2\mathrm{i}$ respectively. We can also check them directly:
$$\begin{pmatrix} 2+\mathrm{i} & \mathrm{i} \\ \mathrm{i} & 2+\mathrm{i} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2+\mathrm{i} & \mathrm{i} \\ \mathrm{i} & 2+\mathrm{i} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = (2+2\mathrm{i})\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Definition 12.6.7

1. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$, and T a linear operator on V. Suppose there exists an ordered orthonormal basis B for V such that $[T]_B$ is a diagonal matrix.

   If $\mathbb{F} = \mathbb{C}$, then T is called *unitarily diagonalizable.*

   If $\mathbb{F} = \mathbb{R}$, then T is called *orthogonally diagonalizable.*

2. A complex square matrix $\boldsymbol{A}$ is called *unitarily diagonalizable* if there exists a unitary matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

   A real square matrix $\boldsymbol{A}$ is called *orthogonally diagonalizable* if there exists an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix. (See Section 6.3.)

Example 12.6.8 Use the complex normal matrix $\boldsymbol{A}$ in Example 12.6.6. Let $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ which is an orthogonal matrix and hence a unitary matrix. Since

$$\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 2-\mathrm{i} & -\mathrm{i} \\ -\mathrm{i} & 2-\mathrm{i} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2-2\mathrm{i} \end{pmatrix},$$

$\boldsymbol{A}$ is unitarily diagonalizable.

Theorem 12.6.9

1. Let V be a complex finite dimensional inner product space where $\dim(V) \geq 1$. A linear operator T on V is unitarily diagonalizable if and only if T is normal.

2. A complex square matrix $\boldsymbol{A}$ is unitarily diagonalizable if and only if $\boldsymbol{A}$ is normal.

Proof We only prove Part 1 of the theorem.

($\Rightarrow$) Suppose $[T]_B$ is a diagonal matrix for some ordered orthonormal basis B for V. Since $([T]_B)^*$ is also a diagonal matrix, $[T]_B\,([T]_B)^* = ([T]_B)^*\,[T]_B$. By Proposition 12.6.2.2, T is normal.

($\Leftarrow$) We use induction on the dimension of V:

If $\dim(V) = 1$, all linear operators on V are normal and unitarily diagonalizable.

Assume that if $\dim(V) = n-1$, then all normal operators on V are unitarily diagonalizable.

Now, suppose $\dim(V) = n$. Since the characteristic polynomial $c_T(x)$ can always be factorized into linear factors over $\mathbb{C}$, T has at least one eigenvalue λ. Suppose $\boldsymbol{u}$ is an eigenvector of T associated with λ. Let $W = \mathrm{span}\{\boldsymbol{u}\}$ which is a T-invariant subspace of V (see Example 11.3.2.2). By Theorem 12.4.3.3,

$$V = W \oplus W^{\perp}.$$

Take any $\boldsymbol{w} \in W^{\perp}$, i.e. $\langle \boldsymbol{w}, \boldsymbol{u}\rangle = 0$. By Lemma 12.6.4.3, $\boldsymbol{u}$ is an eigenvector of T^* associated with $\overline{\lambda}$. Thus

$$\langle T(\boldsymbol{w}),\, \boldsymbol{u}\rangle = \langle \boldsymbol{w},\, T^*(\boldsymbol{u})\rangle = \langle \boldsymbol{w},\, \overline{\lambda}\boldsymbol{u}\rangle = \lambda\langle \boldsymbol{w}, \boldsymbol{u}\rangle = 0$$

and hence $T(\boldsymbol{w}) \in W^\perp$. So $W^\perp$ is T-invariant. Similarly, $W^\perp$ is also T^*-invariant (see Question 12.30).

By Proposition 12.5.7.5, $(T|_{W^\perp})^* = T^*|_{W^\perp}$. Then

$$\begin{aligned}
(T|_{W^\perp}) \circ (T|_{W^\perp})^* &= (T|_{W^\perp}) \circ (T^*|_{W^\perp}) \\
&= (T \circ T^*)|_{W^\perp} && \text{(by Proposition 11.3.3.1)} \\
&= (T^* \circ T)|_{W^\perp} && \text{(because } T \text{ is normal)} \\
&= (T^*|_{W^\perp}) \circ (T|_{W^\perp}) && \text{(again by Proposition 11.3.3.1)} \\
&= (T|_{W^\perp})^* \circ (T|_{W^\perp}).
\end{aligned}$$

So $T|_{W^\perp}$ is a normal operator on $W^\perp$. Since $\dim(W^\perp) = n-1$, by the inductive assumption, there exists an ordered orthonormal basis $C = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_{n-1}}\}$ for $W^\perp$ such that $[T|_{W^\perp}]_C$ is a diagonal matrix.

Let $B = \left\{ \frac{1}{||\boldsymbol{u}||}\boldsymbol{u}, \boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_{n-1}} \right\}$. Then B is an orthonormal basis for V. By Discussion 11.3.12, using B as an ordered basis,

$$[T]_B = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & [T|_{W^\perp}]_C & \\ 0 & & & \end{pmatrix}$$

which is a diagonal matrix. So T is unitarily diagonalizable.

(The proof of Part 2 is left as an exercise. See Question 12.41.)

Algorithm 12.6.10 Let T be a normal operator on a complex finite dimensional vector space V where $\dim(V) \geq 1$. We want to find an ordered orthonormal basis so that the matrix of T relative to this basis is a diagonal matrix.

Step 1: Find an orthonormal basis C for V and compute the matrix $\boldsymbol{A} = [T]_C$.

Step 2: Factorize the characteristic equation $c_{\boldsymbol{A}}(x)$ into linear factors, i.e. to express it in the form

$$c_{\boldsymbol{A}}(x) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of $\boldsymbol{A}$ and $r_1 + r_2 + \cdots + r_k = \dim(V)$.

Step 3: For each eigenvalue λ_i, find a basis for the eigenspace $E_{\lambda_i}(T) = \mathrm{Ker}(T - \lambda_i I_V)$ and then use the Gram-Schmidt Process (Theorem 12.3.7) to transform it to an orthonormal basis B_{λ_i} for $E_{\lambda_i}(T)$.

Step 4: Let $B = B_{\lambda_1} \cup B_{\lambda_2} \cup \cdots \cup B_{\lambda_k}$. Then B is an orthonormal basis for V. Using B as an ordered basis, $\boldsymbol{D} = [T]_B$ is a diagonal matrix.

Note that $\boldsymbol{D} = \boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P}$ where $\boldsymbol{P} = [I_V]_{C,B}$ is the transition matrix from B to C.

(See also Algorithm 6.3.5.)

Example 12.6.11 Suppose $\mathcal{M}_{2\times 2}(\mathbb{C})$ is equipped with the inner product defined in Example 12.1.6.3. Let $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathcal{M}_{2\times 2}(\mathbb{C})$ be the linear operator defined by

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} -b - \mathrm{i}c + \mathrm{i}d & -a + \mathrm{i}c - \mathrm{i}d \\ \mathrm{i}a - \mathrm{i}b - d & -\mathrm{i}a + \mathrm{i}b - c \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{C}).$$

Step 1: Take the standard basis $C = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ for $\mathcal{M}_{2\times 2}(\mathbb{C})$.

Then

$$\boldsymbol{A} = [T]_C = \begin{pmatrix} 0 & -1 & -\mathrm{i} & \mathrm{i} \\ -1 & 0 & \mathrm{i} & -\mathrm{i} \\ \mathrm{i} & -\mathrm{i} & 0 & -1 \\ -\mathrm{i} & \mathrm{i} & -1 & 0 \end{pmatrix}$$

which is a Hermitian matrix and hence is a normal matrix. So T is a normal operator.

Step 2: The characteristic polynomial is

$$c_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} x & 1 & \mathrm{i} & -\mathrm{i} \\ 1 & x & -\mathrm{i} & \mathrm{i} \\ -\mathrm{i} & \mathrm{i} & x & 1 \\ \mathrm{i} & -\mathrm{i} & 1 & x \end{vmatrix} = (x+1)^3(x+3).$$

Thus -1 and -3 are the eigenvalues.

Step 3: To find a basis for $E_{-1}(T)$:

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in E_{-1}(T) \quad &\Leftrightarrow \quad (T + I_{\mathcal{M}_{2\times 2}(\mathbb{C})})\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\
&\Leftrightarrow \quad (\boldsymbol{A} + \boldsymbol{I_4}) \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right]_C = \left[\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\right]_C \\
&\Leftrightarrow \quad \begin{pmatrix} 1 & -1 & -\mathrm{i} & \mathrm{i} \\ -1 & 1 & \mathrm{i} & -\mathrm{i} \\ \mathrm{i} & -\mathrm{i} & 1 & -1 \\ -\mathrm{i} & \mathrm{i} & -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&\Leftrightarrow \quad \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = r\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} \mathrm{i} \\ 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -\mathrm{i} \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{for } r, s, t \in \mathbb{C}.
\end{aligned}$$

Thus $\left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} \mathrm{i} & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -\mathrm{i} & 0 \\ 0 & 1 \end{pmatrix} \right\}$ is a basis for $E_{-1}(T)$ and by the Gram-Schmidt Process, we obtain an orthonormal basis $B_{-1} = \left\{ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ 2 & 0 \end{pmatrix}, \frac{1}{\sqrt{12}}\begin{pmatrix} -\mathrm{i} & \mathrm{i} \\ 1 & 3 \end{pmatrix} \right\}$ for $E_{-1}(T)$.

Similarly, $B_{-3} = \left\{ \frac{1}{2}\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ -1 & 1 \end{pmatrix} \right\}$ is an orthonormal basis for $E_{-3}(T)$.

Step 4: Let $B = \left\{ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ 2 & 0 \end{pmatrix}, \frac{1}{\sqrt{12}}\begin{pmatrix} -\mathrm{i} & \mathrm{i} \\ 1 & 3 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} \mathrm{i} & -\mathrm{i} \\ -1 & 1 \end{pmatrix} \right\}$. We have

$$\boldsymbol{D} = [T]_B = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

Let $\boldsymbol{P} = [I_{\mathcal{M}_{2\times 2}(\mathbb{R})}]_{C,B} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\mathrm{i} & -\frac{1}{\sqrt{12}}\mathrm{i} & \frac{1}{2}\mathrm{i} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\mathrm{i} & \frac{1}{\sqrt{12}}\mathrm{i} & -\frac{1}{2}\mathrm{i} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} & -\frac{1}{2} \\ 0 & 0 & \frac{3}{\sqrt{12}} & \frac{1}{2} \end{pmatrix}$. Then $\boldsymbol{D} = \boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P}$.

Theorem 12.6.12

1. Let V be a real finite dimensional inner product space where $\dim(V) \geq 1$. A linear operator T on V is orthogonally diagonalizable if and only if T is self-adjoint.

2. A real square matrix $\boldsymbol{A}$ is orthogonally diagonalizable if and only if $\boldsymbol{A}$ is symmetric. (See Theorem 6.3.4.)

Proof We only prove Part 1:

($\Rightarrow$) Suppose $[T]_B$ is a diagonal matrix for some ordered orthonormal basis B for V. Since $([T]_B)^{\mathrm{T}} = [T]_B$, by Proposition 12.6.2.1, T is self-adjoint.

($\Leftarrow$) Suppose T is self-adjoint.

Let $\boldsymbol{A} = [T]_C$ where C is an ordered orthonormal basis for V. By Proposition 12.6.2.1, $\boldsymbol{A}$ is an $n \times n$ real symmetric matrix where $n = \dim(V)$. We factorize $c_{\boldsymbol{A}}(x)$ into linear factors over $\mathbb{C}$, say,

$$c_{\boldsymbol{A}}(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_n)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$. Since each λ_i is an eigenvalue of $\boldsymbol{A}$, there exists a nonzero column vector $\boldsymbol{u} \in \mathbb{C}^n$ such that $\boldsymbol{A}\boldsymbol{u} = \lambda_i\boldsymbol{u}$. But then

$$\begin{aligned} \lambda_i\boldsymbol{u} &= \boldsymbol{A}\boldsymbol{u} \\ &= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} && \text{(because } \boldsymbol{A} \text{ is symmetric)} \\ &= \boldsymbol{A}^*\boldsymbol{u} && \text{(because } \boldsymbol{A} \text{ is real)} \\ &= \overline{\lambda}_i\boldsymbol{u} && \text{(by Proposition 12.6.2.2).} \end{aligned}$$

As $\boldsymbol{u}$ is nonzero, $\lambda_i = \overline{\lambda}_i$ and hence $\lambda_i \in \mathbb{R}$.

Since $c_T(x) = c_{\boldsymbol{A}}(x)$, we conclude that $c_T(x)$ can be factorized into linear factors over $\mathbb{R}$. As a self-adjoint operator is normal, by following the same argument as in the proof for Theorem 12.6.9, we can show that T is orthogonally diagonalizable. (In the proof of Theorem 12.6.9, complex numbers are used only because we want to factorize $c_T(x)$ into linear factors.)

(The proof of Part 2 is left as exercise. See Question 12.41.)

Exercise 12

Question 12.1 to Question 12.7 are exercises for Section 12.1.

1. Let $\boldsymbol{A}$ be a complex $m \times n$ matrix.

   (a) Prove that $\operatorname{rank}(\overline{\boldsymbol{A}}) = \operatorname{rank}(\boldsymbol{A})$.

   (b) Hence, or otherwise, prove that $\operatorname{rank}(\boldsymbol{A}^*) = \operatorname{rank}(\boldsymbol{A})$. (Hint: You can use the result from Remark 4.2.5.3.)

2. Let V be a vector space over $\mathbb{F}_2$ such that V has at least two nonzero vectors. Suppose there exists a mapping $\langle\,,\rangle : V \times V \to \mathbb{F}_2$ such that

   (I) for all $\boldsymbol{u}, \boldsymbol{v} \in V$, $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{u}\rangle$; and

   (II) for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $\langle \boldsymbol{u} + \boldsymbol{v},\, \boldsymbol{w}\rangle = \langle \boldsymbol{u}, \boldsymbol{w}\rangle + \langle \boldsymbol{v}, \boldsymbol{w}\rangle$.

   Show that there exists a nonzero vector $\boldsymbol{u} \in V$ such that $\langle \boldsymbol{u},\, \boldsymbol{u}\rangle = 0$.

3. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a real matrix.

   Define $\langle \boldsymbol{u},\, \boldsymbol{v}\rangle = \boldsymbol{u}\boldsymbol{A}\boldsymbol{v}^{\mathrm{T}}$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ where $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as row vectors. Find a necessary and sufficient condition on a, b, c, d so that $\langle\,,\rangle$ is an inner product on $\mathbb{R}^2$.

4. Determine which of the following mappings $\langle\,,\rangle$ are inner products on V.

   (a) $V = \mathbb{C}^2$ and $\langle \boldsymbol{u},\, \boldsymbol{v}\rangle = u_1 v_1 + u_2 v_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in V$.

   (b) $V = \mathbb{C}^2$ and $\langle \boldsymbol{u},\, \boldsymbol{v}\rangle = u_1 \overline{v}_1 + 4u_2 \overline{v}_2$ for $\boldsymbol{u} = (u_1, u_2)$, $\boldsymbol{v} = (v_1, v_2) \in V$.

   (c) $V = \mathcal{P}(\mathbb{R})$ and $\langle p(x),\, q(x)\rangle = \displaystyle\sum_{i=0}^{\min\{m,n\}} a_i b_i$ for $p(x) = \displaystyle\sum_{i=0}^{m} a_i x^i$, $q(x) = \displaystyle\sum_{i=0}^{n} b_i x^i \in V$.

   (d) $V = \mathcal{M}_{n\times n}(\mathbb{R})$ and $\langle \boldsymbol{A},\, \boldsymbol{B}\rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in V$.

5. Verify Example 12.1.6.5:

   Let $[a, b]$, with $a < b$, be a closed interval on the real line. Consider the vector space $C([a, b])$ defined in Example 8.3.6.5. Define

$$\langle f, g \rangle = \tfrac{1}{b-a} \int_a^b f(t)g(t)dt \quad \text{for } f, g \in C([a, b]).$$

   Prove that $\langle\ ,\ \rangle$ is an inner product on $C([a, b])$.

6. Verify Example 12.1.6.6:

   Let V be the set of all real infinite sequences $(a_n)_{n\in\mathbb{N}}$ such that $\sum_{n=1}^{\infty} a_n^2$ converges.

   (a) For any $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}} \in V$, prove that $\sum_{n=1}^{\infty} a_n b_n$ converges.

   (b) Prove that V is a subspace of $\mathbb{R}^{\mathbb{N}}$.

   (c) Define $\langle (a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \rangle = \sum_{n=1}^{\infty} a_n b_n$ for $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}} \in V$.

   Prove that $\langle\ ,\ \rangle$ is an inner product on V.

7. Let V be an inner product space over $\mathbb{C}$ and let

$$\boldsymbol{A} = \begin{pmatrix} \langle \boldsymbol{u_1}, \boldsymbol{u_1} \rangle & \langle \boldsymbol{u_1}, \boldsymbol{u_2} \rangle & \cdots & \langle \boldsymbol{u_1}, \boldsymbol{u_n} \rangle \\ \langle \boldsymbol{u_2}, \boldsymbol{u_1} \rangle & \langle \boldsymbol{u_2}, \boldsymbol{u_2} \rangle & \cdots & \langle \boldsymbol{u_2}, \boldsymbol{u_n} \rangle \\ \vdots & \vdots & & \vdots \\ \langle \boldsymbol{u_n}, \boldsymbol{u_1} \rangle & \langle \boldsymbol{u_n}, \boldsymbol{u_2} \rangle & \cdots & \langle \boldsymbol{u_n}, \boldsymbol{u_n} \rangle \end{pmatrix}$$

   where $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n} \in V$.

   (a) Prove that all eigenvalues of $\boldsymbol{A}$ are nonnegative real numbers. Hence show that $\det(\boldsymbol{A})$ is a nonnegative real number.

   (b) Prove that $\det(\boldsymbol{A}) = 0$ if and only if $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ are linearly dependent.

Question 12.8 to Question 12.15 are exercises for Section 12.2.

8. Find all unit vectors in $\mathcal{P}_1(\mathbb{R})$ if

   (a) $\mathcal{P}_1(\mathbb{R})$ is equipped with the inner product such that

$$\langle p(x), q(x) \rangle = \tfrac{1}{2} \int_{-1}^{1} p(t)q(t)dt \quad \text{for } p(x), q(x) \in \mathcal{P}_1(\mathbb{R});$$

(b) $\mathcal{P}_1(\mathbb{R})$ is equipped with the inner product such that

$$\langle p(x),\, q(x)\rangle = \int_0^1 p(t)q(t)dt \quad \text{for } p(x), q(x) \in \mathcal{P}_1(\mathbb{R});$$

(c) $\mathcal{P}_1(\mathbb{R})$ is equipped with the inner product defined in Question 12.4(c).

9. If $a_1, a_2, ..., a_n$ are positive real numbers, prove that

$$(a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right) \geq n^2.$$

10. For $\boldsymbol{u} = (x_1, x_2), \boldsymbol{v} = (y_1, y_2) \in \mathbb{R}^2$, let $\langle \boldsymbol{u},\, \boldsymbol{v}\rangle = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2$.

(a) Prove that $\langle\, ,\rangle$ is an inner product on $\mathbb{R}^2$.

(b) Prove that for any real numbers x_1, x_2, y_1, y_2,

$$(x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2)^2 \leq \left[(x_1 - x_2)^2 + 3x_2^2\right]\left[(y_1 - y_2)^2 + 3y_2^2\right].$$

11. **(Parallelogram Law)** Let V be an inner product space. Prove that for all $\boldsymbol{u}, \boldsymbol{v} \in V$, $||\boldsymbol{u} + \boldsymbol{v}||^2 + ||\boldsymbol{u} - \boldsymbol{v}||^2 = 2||\boldsymbol{u}||^2 + 2||\boldsymbol{v}||^2$.

12. Complete the proof of Theorem 12.2.4:

Let V be an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. For any $c \in \mathbb{F}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$, prove that

(a) $||\boldsymbol{0}|| = 0$ and if $\boldsymbol{u} \neq \boldsymbol{0}$, $||\boldsymbol{u}|| > 0$;

(b) $||c\boldsymbol{u}|| = |c|\, ||\boldsymbol{u}||$;

(c) $|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| = ||\boldsymbol{u}||\, ||\boldsymbol{v}||$ if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent; and

(d) $||\boldsymbol{u} + \boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$.

13. Let V and W be inner product spaces over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A mapping $T : V \to W$ is said to be *continuous at* $\boldsymbol{v} \in V$ if for any real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that for $\boldsymbol{u} \in V$, $||T(\boldsymbol{v}) - T(\boldsymbol{u})|| < \varepsilon$ whenever $||\boldsymbol{v} - \boldsymbol{u}|| < \delta$. The mapping T is called *continuous* if it is continuous at every $\boldsymbol{v} \in V$.

(a) Suppose V is finite dimensional. Prove that every linear transformation $T : V \to W$ is continuous.

(b) Let $\mathbb{F} = \mathbb{R}$. Prove that if a mapping $T : V \to W$ is continuous and satisfies (T1) of Definition 9.1.2, T is a linear transformation.

14. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *norm* on a vector space V over $\mathbb{F}$ is a mapping $||\,||: V \to \mathbb{R}$ that satisfies the following axioms:

    (N1) $||\mathbf{0}|| = 0$ and, for all nonzero $\boldsymbol{u} \in V$, $||\boldsymbol{u}|| > 0$.

    (N2) For all $c \in \mathbb{F}$ and $\boldsymbol{u} \in V$, $||c\boldsymbol{u}|| = |c|\,||\boldsymbol{u}||$.

    (N3) For all $\boldsymbol{u}, \boldsymbol{v} \in V$, $||\boldsymbol{u} + \boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$.

    A vector space equipped with a norm is called a *normed vector space.*

    For each of the following, determine whether V is a normed vector space.

    (a) V is an inner product space and $||\boldsymbol{u}|| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$ for $\boldsymbol{u} \in V$.

    (b) $V = \mathbb{F}^n$ and $||(a_1, a_2, \ldots, a_n)|| = \min\limits_{1 \leq i \leq n} |a_i|$ for $(a_1, a_2, \ldots, a_n) \in \mathbb{F}^n$.

    (c) $V = \mathbb{F}^n$ and $||(a_1, a_2, \ldots, a_n)|| = \max\limits_{1 \leq i \leq n} |a_i|$ for $(a_1, a_2, \ldots, a_n) \in \mathbb{F}^n$.

    (d) $V = \mathbb{F}^n$ and $||(a_1, a_2, \ldots, a_n)|| = |a_1| + |a_2| + \cdots + |a_n|$ for $(a_1, a_2, \ldots, a_n) \in \mathbb{F}^n$.

    (e) $V = \mathbb{F}^n$ and $||(a_1, a_2, \ldots, a_n)|| = |a_1|^2 + |a_2|^2 + \cdots + |a_n|^2$ for $(a_1, a_2, \ldots, a_n) \in \mathbb{F}^n$.

15. Let V be a normed vector space, as defined in Question 12.14, over $\mathbb{R}$. Define

    $$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \tfrac{1}{4}\left(||\boldsymbol{u} + \boldsymbol{v}||^2 - ||\boldsymbol{u} - \boldsymbol{v}||^2\right) \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in V.$$

    (a) Prove that $\langle\,,\,\rangle$ satisfies (IP1') and (IP4).

    (b) In addition, suppose the norm $||\,||$ satisfies the Parallelogram Law (see Question 12.11), i.e. for $\boldsymbol{u}, \boldsymbol{v} \in V$,

    $$||\boldsymbol{u} + \boldsymbol{v}||^2 + ||\boldsymbol{u} - \boldsymbol{v}||^2 = 2||\boldsymbol{u}||^2 + 2||\boldsymbol{v}||^2.$$

    (i) For all $\boldsymbol{u}, \boldsymbol{v} \in V$, show that $\langle \boldsymbol{u}, 2\boldsymbol{v} \rangle = 2\langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

    (ii) For all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, show that $\langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle = \frac{1}{2}\langle \boldsymbol{u} + \boldsymbol{v}, 2\boldsymbol{w} \rangle$.

    (iii) Prove that $\langle\,,\,\rangle$ satisfies (IP2).

    (iv) For all $r \in \mathbb{Q}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$, show that $\langle r\boldsymbol{u}, \boldsymbol{v} \rangle = r\langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

    (v) For all $\boldsymbol{u}, \boldsymbol{v} \in V$, show that $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq ||\boldsymbol{u}||\,||\boldsymbol{v}||$.

    (vi) For all $c \in \mathbb{R}$, $r \in \mathbb{Q}$ and $\boldsymbol{u}, \boldsymbol{v} \in V$, show that $|\langle c\boldsymbol{u}, \boldsymbol{v} \rangle - r\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq |c - r|\,||\boldsymbol{u}||\,||\boldsymbol{v}||$.

    (vii) Prove that $\langle\,,\,\rangle$ satisfies (IP3) and hence $\langle\,,\,\rangle$ is an inner product on V.

Question 12.16 to Question 12.19 are exercises for Section 12.3.

16. For each part of Question 12.4, write down an orthonormal basis for V if $\langle\,,\,\rangle$ is an inner product.

17. Suppose $C([0,2\pi])$ is equipped with an inner product such that

$$\langle f,\, g\rangle = \tfrac{1}{2\pi}\int_0^{2\pi} f(t)g(t)dt \quad \text{for } f, g \in C([0,2\pi]).$$

Let $\phi_0,\, \phi_1,\, \ldots,\, \phi_{2n}$ be functions in $C([0,2\pi])$ defined by

$$\phi_0(x) = 1, \quad \phi_{2m-1}(x) = \cos(mx), \quad \phi_{2m}(x) = \sin(mx) \quad (m = 1, 2, \ldots, n) \quad \text{for } x \in [0, 2\pi].$$

(a) Prove that $\{\phi_0,\, \phi_1,\, \ldots,\, \phi_{2n}\}$ is orthogonal.

(b) Find an orthonormal basis for $W_n = \operatorname{span}\{\phi_0,\, \phi_1,\, \ldots,\, \phi_{2n}\}$.

18. Let $\mathcal{M}_{2\times 2}(\mathbb{C})$ be equipped with the inner product $\langle \boldsymbol{A},\, \boldsymbol{B}\rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}^*)$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{M}_{2\times 2}(\mathbb{C})$. Let $W = \operatorname{span}\{\boldsymbol{A_1}, \boldsymbol{A_2}, \boldsymbol{A_3}, \boldsymbol{A_4}\}$ where

$$\boldsymbol{A_1} = \begin{pmatrix} 1 & \mathrm{i} \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{A_2} = \begin{pmatrix} 2 & 0 \\ \mathrm{i} & 0 \end{pmatrix}, \quad \boldsymbol{A_3} = \begin{pmatrix} 0 & 2\mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}, \quad \boldsymbol{A_4} = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & \mathrm{i} \end{pmatrix}.$$

(a) Find a subset B of $\{\boldsymbol{A_1}, \boldsymbol{A_2}, \boldsymbol{A_3}, \boldsymbol{A_4}\}$ such that B is a basis for W.

(b) Starting with the basis in (a), use the Gram-Schmidt Process to find an orthonormal basis for W.

19. Suppose $\mathcal{P}_n(\mathbb{R})$ is equipped with an inner product such that

$$\langle p(x),\, q(x)\rangle = \int_0^1 p(t)q(t)dt \quad \text{for } p(x), q(x) \in \mathcal{P}_n(\mathbb{R}).$$

Starting with the standard bases, use the Gram-Schmidt Process to find an orthonormal basis for each of $\mathcal{P}_1(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$.

Question 12.20 to Question 12.28 are exercises for Section 12.4.

20. Suppose $\mathbb{C}^4$ is equipped with the usual inner product. Let W be a subspace of $\mathbb{C}^4$ with a basis $B = \{(1, \mathrm{i}, 0, 0),\ (0, 1, \mathrm{i}, 0),\ (0, 0, 1, \mathrm{i})\}$.

(a) Find an orthonormal basis for W.

(b) Find an orthonormal basis for $W^\perp$.

21. Let $\mathbb{C}^4$ be equipped with the usual inner product and let

$$W = \{(a,\, a - \mathrm{i}b,\, a + 2\mathrm{i}b,\, a + 3\mathrm{i}b) \mid a, b \in \mathbb{C}\}.$$

(a) Find an orthonormal basis for W.

(b) Find a formula for the orthogonal projection $\operatorname{Proj}_W : \mathbb{C}^4 \to \mathbb{C}^4$.

22. Let V be an inner product space and W_1, W_2 subspaces of V.

    (a) If $W_1 \subseteq W_2$, prove that $W_2^\perp \subseteq W_1^\perp$.

    (b) Prove that $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$.

    (c) If V is finite dimensional, prove that $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$.

23. Let T be a linear operator on an inner product space V such that $T^2 = T$.

    (a) Prove that $\mathrm{R}(T) = \{\boldsymbol{v} \in V \mid T(\boldsymbol{v}) = \boldsymbol{v}\}$.

    (b) Prove that $V = \mathrm{R}(T) \oplus \mathrm{Ker}(T)$.

    (c) Suppose $||T(\boldsymbol{u})|| \le ||\boldsymbol{u}||$ for all $\boldsymbol{u} \in V$. Prove that T is the orthogonal projection of V onto $\mathrm{R}(T)$.

24. Let $V = \mathcal{M}_{2\times 2}(\mathbb{R})$ be equipped with the inner product $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^T)$ for $\boldsymbol{A}, \boldsymbol{B} \in V$. Let $W = \mathrm{span}\{\boldsymbol{A_1}, \boldsymbol{A_2}, \boldsymbol{A_3}, \boldsymbol{A_4}\}$ where

$$\boldsymbol{A_1} = \begin{pmatrix} 0 & 0 \\ 1 & -3 \end{pmatrix}, \quad \boldsymbol{A_2} = \begin{pmatrix} 2 & 1 \\ -2 & 2 \end{pmatrix}, \quad \boldsymbol{A_3} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{A_4} = \begin{pmatrix} 0 & 1 \\ 0 & -2 \end{pmatrix}.$$

    (a) Find an orthonormal basis for $W^\perp$.

    (b) Let $\boldsymbol{C} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$. Express $\boldsymbol{C}$ in the form $\boldsymbol{C} = \boldsymbol{P} + \boldsymbol{Q}$ where $\boldsymbol{P} \in W$ and $\boldsymbol{Q} \in W^\perp$.

    (c) Find the smallest value in the set $\{\, ||\boldsymbol{C} - \boldsymbol{X}|| \mid \boldsymbol{X} \in W\}$.

25. Suppose $C([0,1])$ is equipped with an inner product such that

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt \quad \text{for } f, g \in C([0,1]).$$

    Let f be a function in $C([0,1])$ defined by $f(x) = \sqrt{x}$ for $x \in [0,1]$. Find the best approximation of f in each of $\mathcal{P}_1(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$, where $\mathcal{P}_1(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$ are regarded as subspaces of $C([0,1])$. (Hint: Use the orthonormal bases found in Question 12.19.)

26. Suppose $C([0,2\pi])$ is equipped with the inner product defined in Question 12.17. Let f be a function in $C([0,2\pi])$ defined by $f(x) = e^x$ for $x \in [0,2\pi]$. Follow the notation in Question 12.17. Find the best approximation of f in $W_2 = \mathrm{span}\{\phi_0, \phi_1, \phi_2\}$.

27. Let V be an inner product space and W a subspace of V. The distance from a vector $\boldsymbol{u}$ to W is defined to be the value

$$d(\boldsymbol{u}, W) = \min_{\boldsymbol{w} \in W} d(\boldsymbol{u}, \boldsymbol{w}).$$

    For each of the following, find the distance from the given vector to W.

(a) $V = \mathbb{R}^3$ with the usual inner product, $W = \{(x, y, z) \in V \mid x + y + z = 0\}$ and the vector is $\boldsymbol{u} = (1, -1, 2)$.

(b) $V = \mathbb{C}^3$ with the usual inner product, $W = \{(x, y, z) \in V \mid x + y + z = 0\}$ and the vector is $\boldsymbol{u} = (1 + \mathrm{i},\ 0,\ -1)$.

(c) $V = C([0, 1])$ with the inner product as in Question 12.25, $W = \mathcal{P}_1(\mathbb{R})$ and the vector is the function f defined in Question 12.25.

28. Let V be an inner product space and W a subspace of V. The distance from a vector $\boldsymbol{u}$ to a coset $W + \boldsymbol{v}$ is defined to be the value

$$d(\boldsymbol{u},\ W + \boldsymbol{v}) = \min_{\boldsymbol{w} \in W} d(\boldsymbol{u},\ \boldsymbol{w} + \boldsymbol{v}).$$

For each Part of Question 12.27, find the distance from the given vector to the coset of W stated below.

(a) $\{(x, y, z) \in V \mid x + y + z = 1\}$.

(b) $\{(x, y, z) \in V \mid x + y + z = -1 + \mathrm{i}\}$.

(c) $W + g$ where $g \in V$ is the function defined by $g(x) = 1 + \sqrt{x} - 3x$ for $x \in [0, 1]$.

Question 12.29 to Question 12.37 are exercises for Section 12.5.

29. For each of the following linear operator T on the inner product space V, find the adjoint of T.

(a) $V = \mathbb{C}^3$ is equipped with the usual inner product and $T : V \to V$ is defined by $T((x, y, z)) = (x,\ x + yi,\ x + yi - z)$ for $(x, y, z) \in V$.

(b) $V = \mathbb{R}^2$ is equipped with the inner product

$$\langle (u_1, u_2),\ (v_1, v_2) \rangle = u_1 v_1 + 2 u_2 v_2 \quad \text{for } (u_1, u_2),\ (v_1, v_2) \in V$$

and $T : V \to V$ is defined by $T((u_1, u_2)) = (u_2, u_1)$ for $(u_1, u_2) \in V$.

(c) $V = \mathcal{M}_{n \times n}(\mathbb{C})$ is equipped with the inner product

$$\langle \boldsymbol{X},\ \boldsymbol{Y} \rangle = \mathrm{tr}(\boldsymbol{X}\boldsymbol{Y}^*) \quad \text{for } \boldsymbol{X}, \boldsymbol{Y} \in V$$

and $T : V \to V$ is defined by $T(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X}$ for $\boldsymbol{X} \in V$, where $\boldsymbol{A}$ is an $n \times n$ complex matrix.

30. (a) Let T be a linear operator on an inner product space V such that T^* exists and let W be a T-invariant subspace of T. Prove that $W^\perp$ is T^*-invariant.

(b) Let $\mathbb{C}^3$ be equipped with the usual inner product and let $T : \mathbb{C}^3 \to \mathbb{C}^3$ be the linear operator defined by

$$T((x, y, z)) = (x + \mathrm{i}y,\ z - \mathrm{i}y,\ x + y + (1 - \mathrm{i})z) \quad \text{for } (x, y, z) \in \mathbb{C}^3.$$

Let $W = \operatorname{span}\{(1, -1, -1)\}$.

(i) Write down a formula for T^*.

(ii) Find an orthonormal basis for W and an orthonormal basis for $W^\perp$.

(iii) Verify that W is T-invariant and $W^\perp$ is T^*-invariant.

31. Complete the proof of Proposition 12.5.7:

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and let V be an inner product space over $\mathbb{F}$. Suppose S and T are linear operators on V such that S^* and T^* exists. Prove that

(a) $(S + T)^*$ exists and $(S + T)^* = S^* + T^*$;

(b) for any $c \in \mathbb{F}$, $(cT)^*$ exists and $(cT)^* = \bar{c}T^*$;

(c) $(T^*)^*$ exists and $(T^*)^* = T$; and

(d) if W is a subspace of V which is both T-invariant and T^*-invariant, then $(T|_W)^*$ exists and $(T|_W)^* = T^*|_W$.

32. Let T be a linear operator on an inner product space such that T^* exists.

(a) If T is surjective, prove that T^* is injective.

(b) If T is injective, must T^* be surjective?

33. Let T be a linear operator on an inner product space V such that T is invertible and T^* exists.

(a) If V is finite dimensional, prove that T^* is invertible.

(b) If T^* is invertible, prove that $(T^{-1})^*$ exists and $(T^{-1})^* = (T^*)^{-1}$.

(c) Let V be the the l_2-space. Consider the subspace $W = \operatorname{span}\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}, \dots\}$ of V defined in Remark 12.4.5. Define a linear operator T on W such that

$$T((a_n)_{n\in\mathbb{N}}) = (a_n - a_{n+1})_{n\in\mathbb{N}} \ \text{ for } (a_n)_{n\in\mathbb{N}} \in W.$$

(i) Find T^{-1} and T^*.

(ii) Is T^* invertible? Does $(T^{-1})^*$ exist?

34. Let T be a linear operator on an inner product space V such that T^* exists.

(a) Prove that $\mathrm{Ker}(T^* \circ T) = \mathrm{Ker}(T)$.

(b) Is it true that $\mathrm{Ker}(T \circ T^*) = \mathrm{Ker}(T)$? Justify your answer.

35. Let T be a linear operator on an inner product space V such that T^* exists.

(a) For $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k} \in V$, show that if $T(\boldsymbol{v_1}), T(\boldsymbol{v_2}), \dots, T(\boldsymbol{v_k})$ are linearly independent, then $T^*(T(\boldsymbol{v_1})), T^*(T(\boldsymbol{v_2})), \dots, T^*(T(\boldsymbol{v_k}))$ are linearly independent.

(b) Suppose $\mathrm{R}(T)$ is finite dimensional. (Note that V may not be finite dimensional.) Prove that $\mathrm{R}(T^*)$ is finite dimensional and $\mathrm{rank}(T^*) = \mathrm{rank}(T)$.

(Hint: By substituting T by T^* in (a), if $T^*(\boldsymbol{v_1}), T^*(\boldsymbol{v_2}), \dots, T^*(\boldsymbol{v_k})$ are linearly independent, then $T(T^*(\boldsymbol{v_1})), T(T^*(\boldsymbol{v_2})), \dots, T(T^*(\boldsymbol{v_k}))$ are linearly independent.)

36. Let T be a linear operator on an inner product space V such that T^* exists.

(a) Given $\boldsymbol{b} \in V$, show that $\boldsymbol{x} = \boldsymbol{u}$ is a solution to $(T^* \circ T)(\boldsymbol{x}) = T^*(\boldsymbol{b})$ if and only if $T(\boldsymbol{u})$ is the orthogonal projection of $\boldsymbol{b}$ onto $\mathrm{R}(T)$.

(b) Given $\boldsymbol{b} \in \mathrm{R}(T)$, show that $\{\boldsymbol{u} \in V \mid T(\boldsymbol{u}) = \boldsymbol{b}\} = \{\boldsymbol{u} \in V \mid (T^* \circ T)(\boldsymbol{u}) = T^*(\boldsymbol{b})\}$.

37. Determine which of the following complex square matrices are unitary.

$$\text{(i)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}, \text{(ii)} \begin{pmatrix} 0 & \mathrm{i} & 0 \\ 0 & 0 & \mathrm{i} \\ \mathrm{i} & 0 & 0 \end{pmatrix}, \text{(iii)} \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 1 & 0 \\ \mathrm{i} & 0 & 1 \end{pmatrix}, \text{(iv)} \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 1 & 1 \\ -\mathrm{i} & 1 & 1 \end{pmatrix}, \text{(v)} \begin{pmatrix} 1 & 0 & \mathrm{i} \\ 0 & 1 & 0 \\ \mathrm{i} & 0 & 0 \end{pmatrix}.$$

Question 12.38 to Question 12.45 are exercises for Section 12.6.

38. Determine which of the complex square matrices in Question 12.37 are Hermitian and/or normal.

39. Let $\boldsymbol{A} = \begin{pmatrix} 0 & -1 & -\mathrm{i} \\ 1 & 0 & 1 \\ -\mathrm{i} & -1 & 0 \end{pmatrix}$.

(a) Verify that $\boldsymbol{A}$ is a normal matrix.

(b) Find an unitary matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

40. (a) For each of the following matrices $\boldsymbol{A}$, determine whether $\boldsymbol{A}$ is orthogonally diagonalizable. If so, find an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

$$\text{(i) } \boldsymbol{A} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}; \quad \text{(ii) } \boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}; \quad \text{(iii) } \boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(b) For each of the matrices $\boldsymbol{A}$ in (a), determine whether $\boldsymbol{A}$ is unitarily diagonalizable. If so, find a unitary matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^*\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

41. Prove Theorem 12.6.9.2 and Theorem 12.6.12.2:

(a) Prove that a complex square matrix $\boldsymbol{A}$ is unitarily diagonalizable if and only if $\boldsymbol{A}$ is normal.

(b) Prove that a real square matrix $\boldsymbol{A}$ is orthogonally diagonalizable if and only if $\boldsymbol{A}$ is symmetric.

42. (a) Let V be a finite dimensional complex inner product space and T a normal operator on V. Prove that

(i) T is self-adjoint if and only if all eigenvalues of T are real; and

(ii) T is unitary if and only if all eigenvalues of T have modulus 1.

(b) Restate the results in Part (a) using square matrices.

43. Let V be a finite dimensional complex inner product space and $T : V \to V$ a self-adjoint linear operator on V.

(a) Prove that the linear operator $T - \mathrm{i}I_V$ is invertible.

(b) Prove that the linear operator $S = (T + \mathrm{i}I_V) \circ (T - \mathrm{i}I_V)^{-1}$ is unitary.

44. Let $\boldsymbol{A}$ be an $n \times n$ Hermitian matrix, i.e. $\boldsymbol{A}^* = \boldsymbol{A}$. Define

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}\boldsymbol{A}\boldsymbol{v}^* \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^n$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as row vectors.

(a) Prove that $\langle\, , \rangle$ satisfies (IP1), (IP2) and (IP3).

(b) Give a necessary and sufficient condition on the eigenvalues of $\boldsymbol{A}$ so that $\langle\, , \rangle$ is an inner product. (See also Question 12.3.)

45. Let V be a finite dimensional inner product space over $\mathbb{C}$.

(a) Let T be a self-adjoint operator on V. Prove that $\langle T(\boldsymbol{u}), \boldsymbol{u} \rangle$ is a real number for all $\boldsymbol{u} \in V$.

(b) A linear operator P on V is called *positive semi-definite* if P is self-adjoint and $\langle P(\boldsymbol{u}), \boldsymbol{u} \rangle \geq 0$ for all $\boldsymbol{u} \in V$.

(i) Let S be a linear operator on V and let $P = S^* \circ S$. Prove that P is positive semi-definite.

(ii) Suppose P is a positive semi-definite operator on V. Show that there exists a linear operator S on V such that $P = S^* \circ S$.

Index