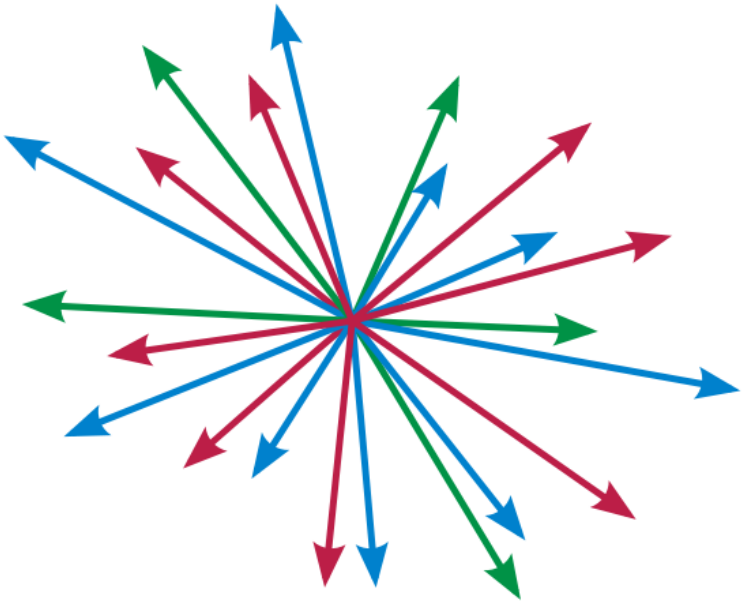
Linear Algebra I

MAT 2141

Fall 2018



Alistair Savage

Department of Mathematics and Statistics

University of Ottawa

Contents

Preface

These are lecture notes for the course *Linear Algebra I* (MAT 2141) at the University of Ottawa. In this course, we will take a more abstract approach to linear algebra than the one taken in MAT 1341 (the prerequisite for this course). Instead of working only with real or complex numbers, we will generalize to the setting where our coefficients lie in a field. The real numbers and complex numbers are both examples of fields, but there are others as well. We will revisit familiar topics such as matrices, linear maps, determinants, and diagonalization, but now in this more general setting and at a deeper level. We will also discuss more advanced topics such as dual spaces, multilinear maps, and inner product spaces. Compared to MAT 1341, this course will concentrate more on justification of results and mathematical rigour as opposed to computation. Almost all results will be accompanied by a proof and students will be expected to do proofs on assignments and exams.

The appendices contains some introductory abstract algebra and discuss the topic of quotient vector spaces, including the important First Isomorphism Theorem. This material will not be covered in the course, and is included here only for the interested student who would like to explore delve further into the subject matter.

Notation: In this course $\mathbb{N} = \mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \dots\}$ denotes the set of nonnegative integers. If Y is a subset of a set X, then

$$X \setminus Y = \{x \in X \mid x \notin Y\}.$$

Acknowledgement: Portions of these notes are based on lecture notes by Barry Jessup, Daniel Daigle, and Kirill Zainoulline.

Alistair Savage

Course website: https://alistairsavage.ca/mat2141

Chapter 1

Vector spaces

In this chapter we introduce vector spaces, which form the central subject of the course. You have seen vector spaces in MAT 1341, and so much of the material in this chapter should be familiar. The topics in this chapter roughly correspond to [Tre, §1.1, §1.2, §1.7].

1.1 Fields

In MAT 1341, you did linear algebra over the real numbers and complex numbers. In fact, linear algebra can be done in a much more general context. For much of linear algebra, one can consider "scalars" from any mathematical object known as a *field*. The topic of fields is an interesting subject in its own right. We will not undertake a detailed study of fields in this course. The interested student is referred to Appendix A.6 for a discussion of this topic.

For the purposes of this course, we will use the word *field* to mean a subfield of the complex numbers or a finite field, which we now explain.

Definition 1.1.1 (Subfield of $\mathbb{C}$)**.** A *subfield* of the complex numbers $\mathbb{C}$ is a subset $F \subseteq \mathbb{C}$ that

(a) contains 1,

(b) is closed under addition (that is, $x + y \in F$ for all $x, y \in F$),

(c) is closed under subtraction (that is, $x - y \in F$ for all $x, y \in F$),

(d) is closed under multiplication (that is, $xy \in F$ for all $x, y \in F$),

(e) is closed under taking the multiplicative inverse of a nonzero element (that is, $x^{-1} \in F$ for all $x \in F$, $x \neq 0$).

Definition 1.1.2 (Finite fields)**.** Let p be a prime number. The *finite field* with p elements is the set

$$\mathbb{F}_p = \{\bar{0}, \bar{1}, \ldots, \overline{n-1}\}$$

of integers modulo p, together with its usual addition and multiplication:

$$\bar{x} + \bar{y} = \text{remainder after dividing } x + y \text{ by } p,$$
$$\bar{x} \cdot \bar{y} = \text{remainder after dividing } x \cdot y \text{ by } p.$$

As long as the context is clear, we sometimes omit the bars and, for instance, write 2 instead of $\bar{2}$.

Example 1.1.3. (The field $\mathbb{F}_2$) We have $\mathbb{F}_2 = \{0, 1\}$, where $0 \neq 1$. The operations $+$ and $\cdot$ on F are defined by:

$+$	0	1
0	0	1
1	1	0

and

$\cdot$	0	1
0	0	0
1	0	1

.

Example 1.1.4. In $\mathbb{F}_5$, we have

$$\bar{3} \cdot \bar{4} = \bar{2}, \quad \bar{2} + \bar{4} = \bar{1}, \quad -(\bar{2}) = \bar{3}, \quad \bar{2}^{-1} = \bar{3}.$$

Remark 1.1.5. We require p to be prime in Definition 1.1.2 since we want all nonzero elements of a field to have multiplicative inverses. See Exercise 1.1.3.

For the purposes of this course, the term *field* will mean either a subfield of $\mathbb{C}$ or a finite field. (See Definition A.6.1 for the more general definition of a field.)

Examples 1.1.6. (a) $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{Q}$ are fields.

(b) The integers $\mathbb{Z}$ are *not* a field since they are not closed under taking the inverse of a nonzero element.

(c) The set $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$ is *not* a field because it is not closed under subtraction.

Example 1.1.7 (The field $\mathbb{Q}(\sqrt{2})$). Let

$$\mathbb{Q}(\sqrt{2}) := \{x + y\sqrt{2} \mid x, y \in \mathbb{Q}\}.$$

Then

$$\mathbb{Q} \subsetneq \mathbb{Q}(\sqrt{2}) \subsetneq \mathbb{R}.$$

The fact that $\mathbb{Q} \neq \mathbb{Q}(\sqrt{2})$ follows from the fact that $\sqrt{2}$ is not a rational number. The fact that $\mathbb{Q}(\sqrt{2}) \neq \mathbb{R}$ follows, for example, from the fact that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$ (see Exercise 1.1.2). We leave it as an exercise (Exercise 1.1.1) to show that $\mathbb{Q}(\sqrt{2})$ is a field.

Examples 1.1.8. (a) $\mathbb{Q}(\sqrt{3})$ is also a subfield of $\mathbb{C}$.

(b) $\mathbb{Q}(i) = \mathbb{Q}(\sqrt{-1})$ is a subfield of $\mathbb{C}$. We have $\mathbb{Q} \subsetneq \mathbb{Q}(i) \subsetneq \mathbb{C}$ and $\mathbb{Q}(i) \not\subseteq \mathbb{R}$.

We write $F^\times$ for the set of nonzero elements of F, that is

$$F^\times = F \setminus \{0\}.$$

We will see that most of the linear algebra that you saw in MAT 1341 can be done over any field. That is, fields can form the "scalars" in systems of equations and vector spaces. For example, we can solve the system

$$\begin{array}{rcrcl} x_1 & - & 2x_2 & = & 2 \\ x_1 & + & x_2 & = & 0 \end{array} \tag{1.1}$$

in $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, or $\mathbb{F}_3$ (Exercise 1.1.4).

Exercises.

1.1.1. Prove that $\mathbb{Q}(\sqrt{2})$, as defined in Example 1.1.7 is a subfield of $\mathbb{C}$.

1.1.2. Show that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$ and so $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ are different fields. *Hint*: Prove the result by contradiction. Assume that $\sqrt{3} = a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$. Show that this leads to a contradiction, using the fact that $\sqrt{2}$ and $\sqrt{3}$ are irrational.

1.1.3. Consider the set $\mathbb{F}_6$ of integers modulo 6, together with the multiplication and addition of Definition 1.1.2. Show that $\bar{2}$ has no multiplicative inverse.

1.1.4. Find all the solutions to the system (1.1) over the fields $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, and $\mathbb{F}_3$.

1.2 Vector spaces

For the remainder of the chapter, F is a field.

Definition 1.2.1 (Vector space)**.** A *vector space over* F is

- a set V (whose objects are called *vectors*),
- a binary operation $+$ on V called *vector addition*, and
- *scalar multiplication:* for each $c \in F$ and $v \in V$, an element $cv \in V$, called the *scalar product* of c and v,

such that the following axioms are satisfied:

(V1) For all $u, v \in V$, we have $u + v = v + u$. (*commutativity of vector addition*)

(V2) For all $u, v, w \in V$, we have $(u + v) + w = u + (v + w)$. (*associativity of vector addition*)

(V3) There is an element $\mathbf{0} \in V$ such that, for all $v \in V$, $v + \mathbf{0} = \mathbf{0} + v = v$. The element $\mathbf{0}$ is unique and is called the *zero vector*.

(V4) For any $v \in V$, there exists an element $-v \in V$ such that $v + (-v) = \mathbf{0}$. The element $-v$ is uniquely determined by v and is called the *additive inverse* of v.

(V5) For all $a \in F$ and $u, v \in V$, we have $a(u + v) = au + av$. (*distributivity of scalar multiplication over vector addition*)

(V6) For all $a, b \in F$ and $v \in V$, we have $(a + b)v = av + bv$. (*distributivity of scalar multiplication over field addition*)

(V7) For all $a, b \in F$ and $v \in V$, we have $a(bv) = (ab)v$. (*compatibility of scalar multiplication with field multiplication*)

(V8) For all $v \in V$, we have $1v = v$, where 1 denotes the multiplicative identity in F. (*unity law*)

Remark 1.2.2 (Notation). In the setting of vector spaces, elements of F will be called *scalars*. Some references used boldface for vectors (e.g. $\mathbf{v}$), while other use arrows over vectors (e.g. $\vec{v}$). We will only use boldface for the zero vector, to distinguish it from the zero element 0 of the field F. In class, we will write $\vec{0}$ for the zero vector (since boldface is hard to write on a blackboard).

Definition 1.2.3 (Real and complex vector spaces)**.** When $F = \mathbb{R}$ in Definition 1.2.1, V is called a *real vector space*. When $F = \mathbb{C}$ in Definition 1.2.1, V is called a *complex vector space*.

Examples 1.2.4. (a) For each positive integer n, $\mathbb{R}^n$ is a real vector space and $\mathbb{C}^n$ is a complex vector space. Here the vector addition and scalar multiplication are the operations you learned in MAT 1341.

(b) Suppose F is a field. For each positive integer n,

$$F^n = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in F\}$$

is a vector space over F with operations defined by

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$
$$c(x_1, \ldots, x_n) = (cx_1, \ldots, cx_n),$$

for $c, x_1, \ldots, x_n, y_1, \ldots, y_n \in F$. The previous examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ are special cases of this example (where $F = \mathbb{R}$ and $F = \mathbb{C}$).

(c) Suppose F is a field. Then for positive integers m and n,

$$M_{m,n}(F) = \{A \mid A \text{ is an } m \times n \text{ matrix with entries in } F\}$$

is a vector space over F with the usual operations of matrix addition and scalar multiplication. Note that matrix multiplication does not play a role when we consider $M_{m,n}(F)$ as a vector space.

(d) In addition to being a complex vector space, $\mathbb{C}$ is *also* a real vector space. See Exercise 1.2.1. In addition, $\mathbb{R}$ and $\mathbb{C}$ are both vector spaces over $\mathbb{Q}$.

For the next example, recall that two functions $f, g\colon X \to Y$ are *equal*, and we write $f = g$, if $f(x) = g(x)$ for all $x \in X$ (see Definition A.5.4).

Example 1.2.5 (Function spaces). Suppose X is a nonempty set and F is a field. Let $\mathcal{F}(X, F)$ or F^X denote the set of functions from X to F (i.e. F valued functions on X). When $X = F$, we sometimes write $\mathcal{F}(F)$ in instead of $\mathcal{F}(F, F)$. If $f, g \in \mathcal{F}(X, F)$ and $c \in F$, we define functions $f + g, cf \in \mathcal{F}(X, F)$ by

$$(f + g)(x) = f(x) + g(x), \qquad (cf)(x) = cf(x), \qquad \forall\, x \in X.$$

So we define addition and scalar multiplication pointwise. Let $\mathbf{0} \in \mathcal{F}(X, F)$ be the function defined by $\mathbf{0}(x) = 0$ for all $x \in X$ (note the important difference between the zero *function*

and the *number* zero). For $f \in \mathcal{F}(X, F)$, define $-f \in \mathcal{F}(X, F)$ by $(-f)(x) = -f(x)$ for all $x \in X$. With these definitions, $\mathcal{F}(X, F)$ is a vector space over F.

For example, we can verify the distributivity axiom as follows: For all $f, g \in \mathcal{F}(X, F)$ and $c \in F$, we need to show that $c(f + g) = cf + cg$. These two functions are equal if they are equal at all points of X, that is, if

$$(c(f + g))(x) = (cf + cg)(x) \quad \forall\, x \in X.$$

Now, since we know that distributivity holds in the field F, for all $x \in X$ we have

$$(c(f+g))(x) = c(f+g)(x) = c(f(x)+g(x)) = cf(x)+cg(x) = (cf)(x)+(cg)(x) = (cf+cg)(x).$$

Thus $c(f + g) = cf + cg$ and so the distributivity axiom in Definition 1.2.1 holds. We leave it as an exercise (Exercise 1.2.2) to verify the remaining axioms of Definition 1.2.1.

Example 1.2.6 ($C^\infty(\mathbb{R})$). Consider the field $F = \mathbb{R}$. Let

$$C^\infty(\mathbb{R}) = \{f \mid f\colon \mathbb{R} \to \mathbb{R} \text{ has derivatives of all orders}\} \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

As in Example 1.2.5, we define addition and scalar multiplication pointwise. For example,

$$\sin x,\ \cos x,\ e^x, x^2 + x^3 \in C^\infty(\mathbb{R}),$$

since these functions have derivatives of all orders (they are *infinitely differentiable*). We claim that $C^\infty(\mathbb{R})$ is a vector space over $\mathbb{R}$.

First of all, we must ask ourselves if the operations of addition and scalar multiplication are well-defined on the set $C^\infty(\mathbb{R})$. But we know from calculus that if the nth derivatives of f and g exist, then so do the nth derivatives of $f + g$ and cf, for all $c \in \mathbb{R}$ (we have $(f + g)^{(n)} = f^{(n)} + g^{(n)}$ and $(cf)^{(n)} = cf^{(n)}$). Thus, $C^\infty(\mathbb{R})$ is *closed* under addition and scalar multiplication and so the operations are well-defined.

Next, we must check the axioms of Definition 1.2.1 are satisfied. Since the zero function $\mathbf{0}$ is infinitely differentiable ($\mathbf{0}' = \mathbf{0}$, $\mathbf{0}'' = \mathbf{0}$, etc.), we have $\mathbf{0} \in C^\infty(\mathbb{R})$. The axioms of Definition 1.2.1 then hold since they hold in $\mathcal{F}(\mathbb{R}, \mathbb{R})$. Thus $C^\infty(\mathbb{R})$ is a vector space over $\mathbb{R}$.

Example 1.2.7. Let $F = \mathbb{R}$ and

$$V = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = \mathbf{0}\} \subseteq C^\infty(\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

For example, if $f(x) = \sin x$, then $f'(x) = \cos x$, $f''(x) = -\sin x$, and so

$$f''(x) + f(x) = -\sin x + \sin x = \mathbf{0},$$

and so $\sin x \in V$. To show that V is a vector space over $\mathbb{R}$, we need to show that it is closed under addition and scalar multiplication (then the rest of the axioms of Definition 1.2.1 will follow from the fact that $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is a vector space). If $f, g \in V$, then

$$(f + g)'' + (f + g) = f'' + g'' + f + g = (f'' + f) + (g'' + g) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

and so $f + g \in V$. Thus V is closed under addition. Now, if $f \in V$ and $c \in \mathbb{R}$, we have

$$(cf)'' + cf = cf'' + cf = c(f'' + f) = c\mathbf{0} = \mathbf{0},$$

and so $cf \in V$. Thus V is a vector space over $\mathbb{R}$.

Example 1.2.8 (Polynomials). Let

$$\begin{aligned}\mathcal{P}(F) &= \{p \mid p \text{ is a polynomial with coefficients in } F\} \\ &= \{p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid n \in \mathbb{N},\ a_i \in F\ \forall\ i\}.\end{aligned}$$

Here t is an "indeterminate" (a formal symbol that is manipulated as if it were a number). For $n \in \mathbb{N}$, let

$$\begin{aligned}\mathcal{P}_n(F) &= \{p \in \mathcal{P}(F) \mid p = 0 \text{ or } \deg p \leq n\} \\ &= \{a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid a_i \in F \text{ for all } i\} \subseteq \mathcal{P}(F).\end{aligned}$$

Note the difference between $\mathcal{P}(F)$ and $\mathcal{P}_n(F)$. The elements of $\mathcal{P}(F)$ are polynomials of *any* degree and the elements of $\mathcal{P}_n(F)$ are polynomials of degree at most n. Both $\mathcal{P}(F)$ and $\mathcal{P}_n(F)$ are vector spaces over F.

Remark 1.2.9. A polynomial $p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0$ defines a function $p : F \to F$. Specifically, $p(t)$ defines the function which maps $c \in F$ to $p(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_0$. There is subtle difference between the *polynomial* $p(t)$ and the *polynomial function* p. For example, if $F = \mathbb{F}_2$, then the polynomials

$$t^2 + t \quad \text{and} \quad 0$$

are different, but they both define the zero function. If the field F is infinite, then two polynomials are equal if and only if they define the same function.

Example 1.2.10 (Infinite sequences). Suppose F is a field. Let

$$V = \{(a_1, a_2, \dots) \mid a_i \in F\ \forall\ i\}$$

be the set of all infinite sequence of elements of F. We define addition and scalar multiplication componentwise. That is,

$$(a_1, a_2, \dots) + (b_1, b_2, \dots) = (a_1 + b_1, a_2 + b_2, \dots)$$

and

$$c(a_1, a_2, \dots) = (ca_1, ca_2, \dots)$$

for $(a_1, a_2, \dots), (b_1, b_2, \dots) \in V$, $c \in F$. We leave it as an exercise (Exercise 1.2.5) to show that V is a vector space over F.

Definition 1.2.11 (Product vector space)**.** Suppose $V_1, V_2, \dots, V_n$ are vector spaces over a field F. Define

$$V_1 \times V_2 \times \cdots \times V_n = \{(v_1, v_2, \dots, v_n) \mid v_i \in V_i,\ 1 \leq i \leq n\}.$$

We write $(v_1, v_2, \dots, v_n) = (w_1, w_2, \dots, w_n)$ if $v_i = w_i$ for all $1 \leq i \leq n$. We define addition and scalar multiplication componentwise:

$$\begin{aligned}(v_1, v_2, \dots, v_n) + (w_1, w_2, \dots, w_n) &= (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n), \\ c(v_1, v_2, \dots, v_n) &= (cv_1, cv_2, \dots, cv_n),\end{aligned}$$

for $(v_1, v_2, \dots, v_n), (w_1, w_2, \dots, w_n) \in V_1 \times V_2 \times \cdots \times V_n$ and $c \in F$. We call $V_1 \times V_2 \times \cdots \times V_n$ the *product vector space* of $V_1, V_2, \dots, V_n$. See Exercise 1.2.6.

Exercises.

1.2.1. Check that the conditions of Definition 1.2.1 are satisfied with $V = \mathbb{C}$ and $F = \mathbb{R}$.

1.2.2. Verify the remaining axioms of Definition 1.2.1 in Example 1.2.5.

1.2.3. Show that $\mathcal{P}(F)$ and $\mathcal{P}_n(F)$ are vector spaces over F.

1.2.4. Fix a positive integer n. Why is the set

$$\{a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid a_i \in F \text{ for all } i,\ a_n \neq 0\}$$

of polynomials of degree exactly n *not* a vector space over F?

1.2.5. Show that V, as defined in Example 1.2.10 is a vector space.

1.2.6. Show that the product vector space (Definition 1.2.11) is actually a vector space over F. That is, show that it satisfies the axioms of Definition 1.2.1. You will need to use the fact that each V_i, $1 \leq i \leq n$, is a vector space over F.

1.2.7 ([Ber14, Ex. 1.3.4]). Suppose V is a real vector space. Let $W = V \times V$ be the (real) product vector space and define multiplication by complex scalars by the formula

$$(a + bi)(u, v) = (au - bv, bu + av), \quad a, b \in \mathbb{R}, \quad (u, v) \in W.$$

Show that W satisfies the axioms for a complex vector space.

1.3 Some properties of vector spaces

In this section we deduce some basic properties of vector spaces that will be used throughout the course.

Theorem 1.3.1. *Suppose V is a vector space over the field F.*

(a) If $\mathbf{0}' \in V$ is a vector such that $\mathbf{0}' + v = v$ for all $v \in V$, then $\mathbf{0}' = \mathbf{0}$. In other words, the zero vector is unique.

(b) If $v + w = \mathbf{0}$ for some $v, w \in V$, then $w = -v$. That is, the additive inverse of a vector is unique (or, negatives of vectors are unique).

(c) For all $v \in V$, we have $-(-v) = v$.

(d) For all $v \in V$, we have $0v = \mathbf{0}$.

(e) For all $c \in F$, $c\mathbf{0} = \mathbf{0}$.

Proof. To prove (d), note that

$$0v + 0v = (0 + 0)v = 0v = 0v.$$

Adding $-0v$ to both sides then gives

$$0v + 0v + (-0v) = 0v + (-0v) \implies 0v = \mathbf{0}.$$

We leave the proof of the remaining statements as Exercise 1.3.1. □

Corollary 1.3.2. *For every vector v and scalar c, we have*

$$c(-v) = -(cv) = (-c)v.$$

Proof. To prove the first equality, note that

$$c(-v) + cv = c(-v + v) = c\mathbf{0} = \mathbf{0}.$$

Thus $c(-v) = -cv$ by the uniqueness of negatives (Theorem 1.3.1(b)). Similarly,

$$cv + (-c)v = (c - c)v = 0v = \mathbf{0},$$

and so $(-c)v = -(cv)$ by the uniqueness of negatives. □

Corollary 1.3.3. *For every vector v, we have $(-1)v = -v$.*

Proof. We have $(-1)v = -(1v) = v$. □

Theorem 1.3.4 (The zero vector has no divisors)**.** *Let V be a vector space, $v \in V$, and c a scalar. Then $cv = \mathbf{0}$ if and only if $c = 0$ or $v = \mathbf{0}$.*

Proof. We already know from Theorem 1.3.1 that if $c = 0$ or $v = \mathbf{0}$, then $cv = \mathbf{0}$. So it remains to prove that if $cv = \mathbf{0}$, then either $c = 0$ or $v = \mathbf{0}$. Suppose $cv = \mathbf{0}$. Either $c = 0$ or $c \neq 0$. If $c = 0$, then we're done. So let's consider the remaining case of $c \neq 0$. Then, since c is a nonzero element of a field, it has a multiplicative inverse. Multiplying both sides of the equation $cv = \mathbf{0}$ by c^{-1} gives

$$c^{-1}cv = c^{-1}\mathbf{0} \implies 1v = \mathbf{0} \implies v = \mathbf{0}.$$ □

Corollary 1.3.5 (Cancellation laws)**.** *Suppose u, v are vectors and c, d are scalars.*

(a) If $cu = cv$ and $c \neq 0$, then $u = v$.

(b) If $cv = dv$ and $v \neq \mathbf{0}$, then $c = d$.

Proof. (a) Since $c \neq 0$ and all nonzero elements of a field have multiplicative inverses, we can multiply both sides of the equation $cu = cv$ by c^{-1} to get

$$c^{-1}(cu) = c^{-1}cv \implies (c^{-1}c)u = (c^{-1}c)v \implies 1u = 1v \implies u = v.$$

(b) $cv = dv \implies cv + (-dv) = dv + (-dv) \implies (c-d)v = \mathbf{0}$. Then, since $v \neq \mathbf{0}$, we have $c - d = 0$ by Theorem 1.3.4. Adding d to both sides gives $c - d + d = 0 + d \implies c + 0 = d \implies c = d$.

□

Definition 1.3.6 (Subtraction of vectors)**.** For vectors u, v, we define

$$u - v = u + (-v).$$

Exercises.

1.3.1. Complete the proof of Theorem 1.3.1.

1.3.2 ([Ber14, Ex. 1.4.1]). Prove that, in a vector space:

(a) $u - v = \mathbf{0}$ if and only if $u = v$;

(b) $u + v = z$ if and only if $u = z - v$.

1.3.3 ([Ber14, Ex. 1.4.2]). Let V be a vector space over a field F.

(a) Show that if $v \in V$ is a fixed nonzero vector, then the mapping $f\colon F \to V$ defined by $f(c) = cv$ is injective.

(b) Show that if c is a fixed nonzero scalar, then the mapping $g\colon V \to V$ defined by $g(v) = cv$ is bijective.

You should directly use the definition of injective and bijective mapppings, and *not* use any properties of linear maps (a topic we have yet to discuss).

1.3.4 ([Ber14, Ex. 1.4.3]). Suppose that v is a fixed vector in a vector space V. Prove that the mapping

$$\tau\colon V \to V, \quad \tau(u) = u + v,$$

is bijective. (This map is called *translation* by the vector v.)

1.3.5 ([Ber14, Ex. 1.4.4]). Prove that if a is a nonzero scalar and v is a fixed vector in a vector space V, then the equation $ax + v = \mathbf{0}$ has a unique solution $x \in V$.

1.3.6. Suppose that, in some vector space V, a vector $u \in V$ has the property that $u + v = v$ for *some* $v \in V$. Prove that $u = \mathbf{0}$.

1.4 Linear combinations

Definition 1.4.1 (Linear combination)**.** Suppose V is a vector space over a field F. If $v_1, v_2, \ldots, v_n \in V$ and $c_1, c_2, \ldots, c_n \in F$, then the vector

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

is called a *linear combination* of v_1, v_2, v_n. The scalars $c_1, c_2, \ldots, c_n$ are called *coefficients*.

Example 1.4.2. In the vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$,

$$2\sin x - 3e^x + 5x^3 - 8|x|$$

is a linear combination of the vectors $\sin x$, e^x, x^3 and $|x|$.

Example 1.4.3 (Wave functions in Physics). The theory of vector spaces plays an important role in many areas of physics. For example, in quantum mechanics, the space of wave functions that describe the state of a system of particles is a vector space. The so-called *superposition principle*, that a system can be in a linear combination of states, corresponds to the fact that in vector spaces, one can form linear combinations of vectors. This is the theory underlying the thought experiment known as Schrödinger's cat.

Definition 1.4.4 (Span)**.** Suppose V is a vector space over F and $v_1, v_2, \ldots, v_n \in V$. Then

$$\text{Span}\{v_1, v_2, \ldots, v_n\} = \{c_1v_1 + c_2v_2 + \cdots + c_nv_n \mid c_1, \ldots, c_n \in F\}$$

is the set of all linear combinations of $v_1, v_2, \ldots, v_n$ and is call the *span* of this set of vectors. When we wish to emphasize the field we're working with (for instance, when we are working with multiple fields), we write $\text{Span}_F\{v_1, \ldots, v_n\}$. Note that [Tre] uses the notation $\mathcal{L}\{v_1, \ldots, v_n\}$.

Example 1.4.5. Recall $\mathcal{P}_n(F) = \{a_nt^n + \cdots + a_0 \mid a_i \in F\}$ from Example 1.2.8. We have

$$\mathcal{P}_n(F) = \text{Span}\{1, t, t^2, \ldots, t^n\},$$

where here 1 denotes the constant polynomial function $1(c) = 1$ for all $c \in F$.

Example 1.4.6. Consider the real vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$. Using the trigonometric sum formulas, we have

$$\cos\left(x + \frac{\pi}{4}\right) = \cos x \cos\frac{\pi}{4} - \sin x \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}\cos x - \frac{1}{\sqrt{2}}\sin x.$$

Thus $\cos\left(x + \frac{\pi}{4}\right)$ is a linear combination of $\cos x$ and $\sin x$ and so $\cos\left(x + \frac{\pi}{4}\right) \in \text{Span}\{\cos x, \sin x\}$.

Examples 1.4.7. (a) Consider $\mathbb{C}$ as a vector space over $\mathbb{C}$. We have $\mathbb{C} = \text{Span}_{\mathbb{C}}\{1\}$ since we can write $z = z1$ for any $z \in \mathbb{C}$.

(b) Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$. Then $\mathbb{C} = \text{Span}_{\mathbb{R}}\{1, i\}$. Since we can write *any* complex number as $a + bi$ for some *real* numbers a, b.

(c) Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Then $\text{Span}_{\mathbb{Q}}\{1, \sqrt{2}\} = \mathbb{Q}(\sqrt{2})$, the field of Example 1.1.7.

Example 1.4.8. Consider the vector space F^n over F, for some field F. Let

$$e_1 = (1, 0, 0, \ldots, 0), \quad e_2 = (0, 1, 0, \ldots, 0), \quad \ldots, \quad e_n = (0, \ldots, 0, 0, 1).$$

So for $1 \leq i \leq n$, e_i has a 1 in the ith position and a zero everywhere else. Then

$$F^n = \text{Span}_F\{e_1, e_2, \ldots, e_n\}$$

since every vector $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in F^n$ can be written as

$$\mathbf{a} = a_1e_1 + a_2e_2 + \cdots + a_ne_n = \sum_{k=1}^{n} a_ke_k.$$

Therefore every vector in F^n is a linear combination of the vectors $e_1, e_2, \ldots, e_n$.

Exercises.

1.4.1. In $\mathbb{C}^4$, express the vector $(2 + i, 3 - 7i, 0, -6)$ as a linear combination of e_1, e_2, e_3, e_4 (see Example 1.4.8).

1.4.2. Show that, in any vector space, $u - v$ is a linear combination of u and v.

1.4.3. For each of the following statements, deduce whether the statement is true or false. Justify your answers.

(a) In $\mathbb{R}^3$, the vector $(3, 1, -7)$ is a linear combination of $(1, 0, 1)$ and $(2, 0, 3)$.

(b) In $\mathbb{R}^3$, the vector $(3, 0, -7)$ is a linear combination of $(1, 0, 1)$ and $(2, 0.3)$.

(c) For any vector space V and $u, v \in V$, the vector u is a linear combination of $u - v$ and $u + v$.

1.4.4 ([Ber14, Ex. 1.5.8]). In the vector space $\mathcal{P}(F)$ (Example 1.2.8), let

$$p(t) = 2t^3 - 5t^2 + 6t - 4, \quad q(t) = t^3 + 6t^2 + 3t + 5, \quad r(t) = 4t^2 - 3t + 7.$$

Is r a linear combination of p and q? Justify your answer.

1.5 Subspaces

Definition 1.5.1 (Subspace)**.** Suppose V is a vector space over a field F. A subset $U \subseteq V$ is called a *subspace* (or *linear subspace*) of V if

(a) $\mathbf{0} \in U$,

(b) U is closed under vector addition: $u + v \in U$ for all $u, v \in U$.

(c) U is closed under scalar multiplication: $cu \in U$ for all $u \in U$ and $c \in F$.

Theorem 1.5.2 (Subspaces are vector spaces)**.** *If U is a subspace of a vector space V, then U is a vector space.*

Proof. By Definition 1.5.1, vector addition and scalar multiplication are well-defined on U. Since $\mathbf{0} \in U$ and $-v = (-1)v \in U$ for all $v \in U$, we see that axioms (V3) and (V4) of a vector space are satisfied.. The remaining axioms hold since they hold in V (because V is a vector space). □

Examples 1.5.3. (a) If V is a vector space over a field F, then $\{\mathbf{0}\}$ and V are both subspaces of V. These are called the *trivial* subspaces. They are different if $V \neq \{\mathbf{0}\}$.

(b) Suppose $A \in M_{m,n}(F)$. Then

$$\operatorname{Ker} A = \{v \in F^n \mid Av = \mathbf{0}\}$$

is a subspace of F^n. We check the axioms of Definition 1.5.1. Since $A\mathbf{0} = \mathbf{0}$, we have $\mathbf{0} \in \operatorname{Ker} A$. If $u, v \in \operatorname{Ker} A$, then $A(u + v) = Au + Av = \mathbf{0} + \mathbf{0} = \mathbf{0}$ and so $u + v \in A$. If $u \in \operatorname{Ker} A$ and $c \in F$, then $A(cv) = c(Av) = c\mathbf{0} = \mathbf{0}$ and so $cv \in \operatorname{Ker} A$.

(c) Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$. Then $\operatorname{Span}_{\mathbb{R}}\{1\} = \mathbb{R}$ is a subspace of $\mathbb{C}$ over $\mathbb{R}$.

(d) Consider $\mathbb{C}$ as a vector space over $\mathbb{C}$. Then the only subspaces are $\{\mathbf{0}\}$ and $\mathbb{C}$.

(e) More generally, for any field F, we can consider F as a vector space over itself. We leave it as an exercise (Exercise 1.5.1) to show that the only subspaces are $\{\mathbf{0}\}$ and F.

(f) The set

$$V = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f(x) \geq 0 \ \forall \ x \in \mathbb{R}\},$$

is *not* a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$ because it does not contain the additive inverse of every element in V. (So it is not a vector space, hence not a subspace by Theorem 1.5.2.) For instance, $f(x) = x^2$ is a function in V, but the additive inverse $-f$ is defined by $(-f)(x) = -x^2$ and so $-f$ is not in V (since, for example $(-f)(1) = -1 < 0$). The set V is also not closed under scalar multiplication (see Exercise 1.5.2).

(g) The set $C^\infty(\mathbb{R})$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$. The set

$$\{f \in C^\infty(\mathbb{R}) \mid f'' + f = \mathbf{0}\}$$

is a subspace of $C^\infty(\mathbb{R})$ *and* a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

(h) For any $n \in \mathbb{N}$, $\mathcal{P}_n(F)$ is a subspace of $\mathcal{P}(F)$. When $F = \mathbb{R}$, if we view polynomials as polynomial *functions*, then $\mathcal{P}_n(\mathbb{R})$ is even a subspace of $C^\infty(\mathbb{R})$ since polynomials are infinitely differentiable.

(i) For a field F, let
$$W = \{p \in \mathcal{P}(F) \mid p(1) = 0\}.$$
We check that W is a subspace of $\mathcal{P}(F)$. Since $\mathbf{0}(1) = 0$, we have $\mathbf{0} \in W$. Suppose $p, q \in W$, then
$$(p+q)(1) = p(1) + q(1) = 0 + 0 = 0,$$
and so $p + q \in W$. Finally, if $p \in W$ and $c \in F$, then
$$(cp)(1) = cp(1) = c \cdot 0 = 0,$$
and so $cp \in W$. Thus, W is indeed a subspace of $\mathcal{P}$.

(j) For a field F, let
$$V = \{p \in \mathcal{P}(F) \mid p(1) = 1\}.$$
Since $\mathbf{0}(1) = 0 \neq 1$, $\mathbf{0} \notin V$ and so V is *not* a subspace of $\mathcal{P}(F)$. The set V also violates the other axioms of a subspace but we can stop here because as soon as a set violates *one* of the axioms, we know it is not a subspace.

Theorem 1.5.4. *If $v_1, v_2, \ldots, v_n$ are vectors in a vector space V, then* $\text{Span}\{v_1, v_2, \ldots, v_n\}$ *is a subspace of V.*

Proof. The proof of this theorem is exactly like the proof of the special case where the field is $\mathbb{R}$ (which you saw in MAT 1341). □

Definition 1.5.5 (Sum and intersection of subsets)**.** Suppose M and N are subsets of a vector space V (note that they do not need to be subspaces). We define
$$M \cap N = \{v \in V \mid v \in M \text{ and } v \in N\}, \text{ and}$$
$$M + N = \{u + v \mid u \in M,\ v \in N\}.$$

These are called the *intersection* and *sum* (respectively) of M and N.

Theorem 1.5.6. *If U and W are* subspaces *of a vector space V, then $U \cap W$ and $U + W$ are subspaces as well.*

Proof. Let's prove that $U + W$ is a subspace. Since $\mathbf{0} \in U$ and $\mathbf{0} \in W$, we have that
$$\mathbf{0} = \mathbf{0} + \mathbf{0} \in U + W.$$

Next we show that $U + W$ is closed under vector addition. Suppose that $v = u + w$ and $v' = u' + w'$ are two vectors in $U + W$ (i.e. $u, u' \in U$ and $w, w' \in W$). Then
$$v + v' = (u + w) + (u' + w') = (u + u') + (w + w').$$

Since U is a subspace, it is closed under vector addition and so $u + u' \in U$. Similarly, since W is a subspace, $w + w' \in W$. Thus $v + v' \in U + W$. Finally, we show that $U + W$ is closed under scalar multiplication. Suppose that $v = u + w \in U + W$ (with $u \in U$ and $w \in W$). Then
$$cv = c(u + w) = cu + cw.$$

Since U is a subspace, it is closed under scalar multiplication. Thus $cu \in U$. Similarly, $cw \in W$. Therefore, $cv \in U + W$. So we have shown that $U + W$ is a subspace.

We leave it as an exercise (Exercise 1.5.3) to show that $U \cap W$ is a subspace of V. In fact, it is also a subspace of both U and W (since $U \cap W \subseteq U$ and $U \cap W \subseteq W$). □

Example 1.5.7. Suppose $V = F^3$, $U = \{(x, 0, 0) \mid x \in F\}$, $W = \{(0, y, 0) \mid y \in F\}$. We leave it as an exercise (Exercise 1.5.4) to show that U and W are subspaces of V, and that $U + W = \{(x, y, 0) \mid x, y \in F\}$ (which is also a subspace). Note that

$$U = \text{Span}\{(1, 0, 0)\}, \qquad W = \text{Span}\{(0, 1, 0)\},$$
$$U + W = \text{Span}\{(1, 0, 0), (0, 1, 0)\}, \qquad U \cap W = \{0\}.$$

Corollary 1.5.8. *Suppose U, W are subspaces of a vector space V over F. Then,*

(a) $U \cap W$ is the largest subspace of V contained in both U and W (that is, if X is any subspace of V such that $X \subseteq U$ and $X \subseteq W$, then $X \subseteq U \cap W$), and

(b) $U + W$ is the smallest subspace of V containing both U and W (that is, if Y is any subspace of V such that $U \subseteq Y$ and $V \subseteq Y$, then $U + W \subseteq Y$).

Proof. (a) Suppose X is a subspace of V such that $X \subseteq U$ and $X \subseteq W$. Then $X \subseteq U \cap W$ by the definition of the intersection $U \cap W$.

(b) Suppose Y is a subspace of V such that $U \subseteq Y$ and $W \subseteq Y$. Let $v \in U + W$. then $v = u + w$ for some $u \in U$ and $w \in W$. Since $u \in U \subseteq Y$, we have $u \in Y$. Similarly, $w \in W \subseteq Y$ implies $w \in Y$. Since Y is a subspace, it is closed under vector addition and so $v = u + w \in Y$. So we have shown that every element of $U + W$ is an element of Y. Therefore $U + W \subseteq Y$.

□

Definition 1.5.9 (Direct sum)**.** Suppose V is a vector space and U, W are subspaces of V such that

(a) $U \cap W = \{\mathbf{0}\}$, and

(b) $U + W = V$.

We say that V is the *direct sum* of U and W and write $V = U \oplus W$.

Example 1.5.10. Suppose F is a field, $V = F^2$,

$$U = \{(x, 0) \mid x \in F\}, \quad \text{and} \quad W = \{(0, x) \mid x \in F\}.$$

Then $U \cap W = \{\mathbf{0}\}$ and $U + W = V$ since any vector $(x, y) \in V$ can be written as $(x, y) = (x, 0) + (0, y) \in U + W$. Thus $V = U \oplus W$. Sometimes this as written as $F^2 = F \oplus F$, where we identity U and W with F (by considering only the nonzero component).

Theorem 1.5.11. *Suppose U and W are subspace of a vector space V. The following statements are equivalent:*

(a) $V = U \oplus W$,

(b) For each $v \in V$, there are unique *elements $u \in U$ and $w \in W$ such that $v = u + w$.*

Proof. We first proof that (a) implies (b). So we assume (a) is true. Suppose $v \in V$. Since $V = U + W$, there exist $u \in U$ and $w \in W$ such that $v = u + w$. Now suppose $v = u' + w'$ for some $u' \in U$ and $w' \in W$. Then

$$\mathbf{0} = v - v = (u + w) - (u' - w') = (u - u') + (w - w') \implies u - u' = w' - w.$$

Now $u - u' \in U$ since U is a subspace (hence closed under vector addition and scalar multiplication). Similarly, $w' - w \in W$. So $u - u' \in U$ and $u - u' = w - w' \in W$. Thus $u - u' \in U \cap W$. But $U \cap W = \{\mathbf{0}\}$. Therefore $u - u' = \mathbf{0}$ and $w - w' = u - u' = \mathbf{0}$. So $u = u'$ and $w = w'$. Hence the representation of v in the form $v = u + w$ is *unique*.

We leave it as an exercise (Exercise 1.5.7) to prove that (b) implies (a). □

If U, W, and V satisfy the equivalent conditions of Theorem 1.5.11, we say that W is a *complement* to U in V. Of course, it follows that U is also a complement to W.

Remark 1.5.12. To show that S is a subspace of a vector space V over a field F, it is enough to that $\mathbf{0} \in S$ and

$$u, v \in S,\ c, d \in F \implies cu + dv \in S.$$

Why? Take $c = d = 1$ to see that S is closed under vector addition and then take $d = 0$ to see that S is closed under scalar multiplication.

Exercises.

1.5.1. Suppose F is a field and consider F as a vector space over itself. Show that the only subspaces are $\{0\}$ and F.

1.5.2. Come up with an example to show that V, as defined in Example 1.5.3(f) is not closed under scalar multiplication.

1.5.3. Complete the proof of Theorem 1.5.6 by showing that $U \cap W$ is a subspace of V.

1.5.4. Consider Example 1.5.7.

(a) Show that U and W are subspaces of V.

(b) Show that $U + W = \{(x, y, 0) \mid x, y \in F\}$.

1.5.5. Suppose $V = F^3$, $U = \{(x, 0, z) \mid x, z \in F\}$, $W = \{(y, y + z, z) \mid y, z \in F\}$.

(a) Show that U and W are subspaces of V and that

$$U = \text{Span}\{(1, 0, 0), (0, 0, 1)\}, \quad W = \text{Span}\{(1, 1, 0), (0, 1, 1)\} = \text{Span}\{(1, 0, -1), (0, 1, 1)\}.$$

(b) Show that $U + W = F^3$.

(c) Show that $U \cap W = \text{Span}\{(1, 0, -1)\}$.

1.5.6 ([Ber14, Ex. 1.6.5]). Let M, N, and P be subspaces of a vector space V.

(a) Show that, if $M \subseteq P$, then $P \cap (M + N) = M + (P \cap N)$. (This is called the *modular law* for subspaces.)

(b) Give an example to show that, in general, $P \cap (M + N) \neq (P \cap M) + (P \cap N)$. *Hint*: Let $V = \mathbb{R}^2$ and let P, M, and N be three distinct lines through the origin.

1.5.7. Complete the proof of Theorem 1.5.11 by showing that (b) implies (a).

1.5.8. Let $U = \text{Span}_{\mathbb{R}}\{(1, 1)\}$ and $V = \text{Span}_{\mathbb{R}}\{1, -1)\}$. Show that $\mathbb{R}^2 = U \oplus V$.

1.5.9. Let X be a nonempty set and let F be a field. Recall that $\mathcal{F}(X, F)$ is the set of functions from X to F and is a vector space over F(see Example 1.2.5). Let Y be a nonempty subset of X. Show that

$$V = \{f \in \mathcal{F}(X, F) \mid f(x) = 0 \ \forall \ x \in Y\}$$

is a subspace of $\mathcal{F}(X, F)$.

1.5.10 ([Ber14, Ex. 1.6.13]). Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ be the real vector space of all functions $x\colon \mathbb{R} \to \mathbb{R}$ (see Example 1.2.5). We say that $y \in V$ is *even* if $y(-t) = y(t)$ for all $t \in \mathbb{R}$ and we say that $z \in V$ is *odd* if $z(-t) = -z(t)$ for all $t \in \mathbb{R}$. Let

$$M = \{y \in V \mid y \text{ is even}\} \quad \text{and} \quad N = \{z \in V \mid z \text{ is odd}\}.$$

(a) Prove that M and N are subspace of V and that $V = M \oplus N$. *Hint*: If $x \in V$, consider the functions y and z defined by

$$y(t) = \frac{1}{2}\left(x(t) + x(-t)\right) \quad \text{and} \quad z(t) = \frac{1}{2}\left(x(t) - x(-t)\right).$$

(b) What does (a) say for $x(t) = e^t$? *Hint*: Think about hyperbolic trigonometric functions.

(c) What does (a) say for a polynomial function x?

1.5.11. Suppose

$$V = V_1 \times V_2, \quad M_1 = \{(x_1, \mathbf{0}) \mid x_1 \in V_1\}, \quad M_2 = \{(\mathbf{0}, x_2) \mid x_2 \in V_2\}.$$

Prove that $V = M_1 \oplus M_2$.

1.5.12 ([Ber14, Ex. 1.6.18]). Let M and N be subspaces of a vector space V whose union $M \cup N$ is also a subspace of V. Prove that either $M \subseteq N$ or $N \subseteq M$. *Hint:* Try proof by contradiction. Assume that neither of M, N is contained in the other. Then we can choose vectors $y \in M$, $z \in N$ such that $y \notin N$, $z \notin M$. Think about the sum $y + z$.

1.5.13. Give an example of a vector space V with two subspaces U and W such that $U \cup W$ is *not* a subspace of V.

Chapter 2

Linear maps

As a general rule, whenever one introduces a new type of mathematical object (such as vector spaces), it is important to look the natural maps between them. This helps us understand the relationships between these objects. In the case of vector spaces, the natural maps between them are linear maps. The topics in this chapter roughly correspond to [Tre, §§1.3–1.6].

2.1 Definition and examples

Definition 2.1.1. Suppose V and W are vector spaces over the same field F. A function $T\colon V \to W$ is said to be a *linear map* if

(a) $T(v+w) = T(v) + T(w)$ for all $v, w \in V$, and

(b) $T(cv) = cT(v)$ for all $c \in F$ and $v \in V$.

Such a function is also called a *linear transformation*.

Remark 2.1.2. If $T\colon V \to W$ is a linear map and $v \in V$, we will sometimes write Tv instead of $T(v)$. (This should remind you of matrix multiplication.)

Remark 2.1.3. Suppose V and W are vector spaces over a field F and $T\colon V \to W$ is a linear map.

(a) It follows from the definition of a linear map that

$$T(cv + dw) = cT(v) + dT(w), \quad \forall\ c, d \in F,\ v, w \in V.$$

(b) In fact, if $c_1, \ldots, c_n \in F$, and $v_1, \ldots, v_n \in V$, then

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

(c) We have

$$T(\mathbf{0}_V) = T(0 \cdot \mathbf{0}_V) = 0 \cdot T(\mathbf{0}_V) = \mathbf{0}_W.$$

Here we use the notation $\mathbf{0}_V$ and $\mathbf{0}_W$ to distinguish between the zero vectors of V and W.

(d) For $v \in V$, we have

$$T(-v) = T((-1)v) = (-1)T(v) = -T(v).$$

Now that we have the definition of a linear map, we turn our attention to constructing some.

Theorem 2.1.4. *Suppose V is a vector space over a field F and $v_1, v_2, \ldots, v_n \in V$. (Note that we do not require that the v_i be distinct. In other words, some of the v_i may be equal.) Then the map $T\colon F^n \to V$ defined by*

$$T(a_1, a_2, \ldots, a_n) = a_1v_1 + a_2v_2 + \cdots + a_nv_n = \sum_{i=1}^{n} a_iv_i.$$

is linear. Furthermore, it is the unique linear map such that $Te_i = v_i$ for all $1 \leq i \leq n$, where the e_i are the vectors described in Example 1.4.8.

Proof. If $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ are vectors in F^n and $c \in F$, then

$$\begin{aligned} T(a+b) &= T(a_1 + b_1, \ldots, a_n + b_n) \\ &= (a_1 + b_1)v_1 + \cdots + (a_n + b_n)v_n \\ &= (a_1v_1 + \cdots + a_nv_n) + (b_1v_1 + \cdots + b_nv_n) \\ &= Ta + Tb. \end{aligned}$$

Also,

$$\begin{aligned} T(ca) &= T(ca_1, \ldots, ca_n) \\ &= (ca_1)v_1 + \cdots + (ca_n)v_n \\ &= c(a_1v_1) + \cdots + c(a_nv_n) \\ &= c(a_1v_1 + \cdots + a_nv_n) \\ &= c(Ta). \end{aligned}$$

Therefore T is linear. Now

$$Te_i = T(0, \ldots, 0, 1, 0, \ldots, 0) = 0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n = v_i.$$

If S is another linear map such that $Se_i = v_i$ for all i, then for all $a = (a_1, \ldots, a_n) \in F^n$, we have

$$Sa = S(a_1, \ldots, a_n) = S(a_1e_1 + \cdots + a_ne_n) = \sum_{i=1}^{n} a_iS(e_i) = \sum_{i=1}^{n} a_iv_i = Ta.$$

Thus $Sa = Ta$ for all $a \in F^n$ and so $S = T$. □

Remark 2.1.5. Suppose V and W are vector spaces over a field F. To show that a map $T\colon V \to W$ is linear, it is enough to show that

$$T(cu + dv) = cT(u) + dT(v) \ \forall\ c, d \in F,\ u, v \in V.$$

Why? Consider the case $c = d = 1$ and then the case $d = 0$.

Examples 2.1.6. (a) Let $V = C^\infty(\mathbb{R})$, $F = \mathbb{R}$, and define $D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by $D(f) = f'$ (i.e. the map D takes the derivative). Then we know from calculus that

$$D(cf + dg) = (cf + dg)' = cf' + dg' = cD(f) + dD(g),$$

for all $c, d \in \mathbb{R}$ and $f, g \in \mathbb{C}^\infty(\mathbb{R})$.

(b) If

$$C^n(\mathbb{R}) = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f \text{ is } n \text{ times differentiable}\},$$

then $D\colon C^n(\mathbb{R}) \to C^{n-1}(\mathbb{R})$ is linear ($n \geq 1$).

(c) Let $V = C^\infty(\mathbb{R})$, $F = \mathbb{R}$, and define $S\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by

$$(Sf)(x) = \int_0^x f(t)\,dt, \ \forall\ x \in \mathbb{R}.$$

Then we know from calculus that $S(f) \in C^\infty(\mathbb{R})$ and

$$\begin{aligned} S(cf + dg) &= \int_0^x (cf + dg)(t)\,dt = \int_0^x (cf(t) + dg(t))\,dt \\ &= c\int_0^x f(t)\,dt + d\int_0^x g(t)\,dt = cS(f) + dS(g), \end{aligned}$$

for all $c, d \in \mathbb{R}$ and $f, g \in C^\infty(\mathbb{R})$. Thus S is a linear map.

(d) We leave it as an exercise (Exercise 2.1.1) to show that the map $T\colon \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ defined by $T(at^2 + bt + c) = (a, b, c)$ is linear.

Examples 2.1.7. The following are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^3$ (Exericse 2.1.2):

$$\begin{aligned} S(x_1, x_2, x_3) &= (x_3, x_1, x_2) \\ T(x_1, x_2, x_3) &= (2x_1 - 5x_3, 0, 2x_2) \end{aligned}$$

Example 2.1.8. The maps

$$\begin{aligned} &T\colon \mathbb{R}^3 \to \mathbb{R}^2, \quad T(x_1, x_2, x_3) = (x_2 - x_1, 2x_3), \\ &S\colon \mathbb{R}^2 \to \mathbb{R}^4, \quad S(x_1, x_2) = (2x_2 - x_1, 0, x_1, -4x_2) \end{aligned}$$

are linear (Exercise 2.1.3).

Example 2.1.9. If V is any vector space and a is any scalar, then the map $T\colon V \to V$ defined by $Tv = av$ is linear since for any $u, v \in V$ and scalars c, d, we have

$$\begin{aligned} T(cu + dv) &= a(cu + dv) = a(cu) + a(dv) = (ac)u + (ad)v \\ &= (ca)u + (da)v = c(au) + d(av) = cTu + dTv. \end{aligned}$$

Note that we used the commutativity of the field of scalars here.

Definition 2.1.10 (Linear form and dual space)**.** Suppose V is a vector space over a field F. Then a *linear form* on V is a linear map $V \to F$ and

$$V^* := \{f \mid f\colon V \to F \text{ is linear}\}$$

is called the *dual space* (or simply *dual*) of V.

We will soon see that V^* is itself a vector space.

Exercises.

2.1.1. Prove that the map T defined in 2.1.6(d) is linear.

2.1.2. Verify that the maps of Exercise 2.1.7 are linear.

2.1.3. Verify that the maps of Exercise 2.1.8 are linear.

2.1.4. Suppose V is a vector space. Prove that the map

$$T\colon V \times V \to V, \quad T(u,v) = u - v,$$

is linear.

2.1.5. Fix a vector $v \in \mathbb{R}^3$ and define

$$T\colon \mathbb{R}^3 \to \mathbb{R}^3, \quad T(u) = u \times v,$$

where $u \times v$ denotes the cross product of u and v. Prove that T is linear.

2.1.6. Suppose F is a field, X is a set, and $x \in X$. Prove that the map

$$T\colon \mathcal{F}(X,F) \to F, \quad T(f) = f(x),$$

is a linear form on $\mathcal{F}(X,F)$.

2.1.7. Prove that if $T\colon V \to W$ is a linear map, then $T(u-v) = T(u) - T(v)$ for all $u, v \in V$.

2.2 Kernel and image

Definition 2.2.1. If $f\colon A \to B$ is any map of sets (in particular, f could be a linear map between vector spaces) and $A' \subseteq A$, then

$$f(A') = \{f(a) \mid a \in A'\} \subseteq B$$

is called the *image* of A' under f. If $B' \subseteq B$, then

$$f^{-1}(B') = \{a \in A \mid f(a) \in B'\} \subseteq A$$

is called the *inverse image* (or *preimage*) of B' under f. Note that we use the notation $f^{-1}(B')$ here even though we do *not* assume that f is invertible. If $B' = \{b\}$ consists of a single element, we will sometimes write $f^{-1}(b)$ for $f^{-1}(\{b\})$ (here one must be extra careful to not confuse this with the inverse function, which may or may not exist).

If $f\colon \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2$, then f is not invertible since it is not injective. However, we can still compute inverse images. For example

$$f^{-1}(\{0,1\}) = \{-1,0,1\}, \quad f^{-1}(4) = \{-2,2\}, \quad f^{-1}(-1) = \varnothing.$$

We also have

$$f(\{-1,1,4\}) = \{1,16\}, \quad f(\{x \in \mathbb{R} \mid -3 \le x \le 5\}) = \{y \in \mathbb{R} \mid 0 \le y \le 25\}.$$

Theorem 2.2.2. *Suppose V and W are vector spaces and $T\colon V \to W$ is a linear map.*

(a) If M is a subspace of V, then $T(M)$ is a linear subspace of W.

(b) If N is a subspace of W, then $T^{-1}(N)$ is a subspace of V.

Proof. We will use Remark 1.5.12 in this proof.

(a) Since M is a subspace, we have $\mathbf{0} \in M$ and so $\mathbf{0} = T\mathbf{0} \in T(M)$. Now suppose that $y, y' \in T(M)$ and c, c' are scalars. Then there are $x, x' \in M$ such that $y = Tx$, $y' = Tx'$ and so

$$cy + c'y' = cTx + c'Tx' = T(cx) + T(c'x') = T(cx + c'x') \in T(M)$$

since $cx + c'x' \in M$ (because M is a subspace).

(b) Since $T\mathbf{0} = \mathbf{0} \in N$, we have $\mathbf{0} \in T^{-1}(N)$. Now suppose that $x, x' \in T^{-1}(N)$ and c, c' are scalars. Then

$$T(cx + c'x') = cTx + c'Tx' \in N,$$

since $Tx, Tx' \in N$. Thus $cx + c'x' \in T^{-1}(N)$.

□

Definition 2.2.3 (Kernel and image)**.** If $T\colon V \to W$ is a linear map, then

$$\operatorname{Ker} T = T^{-1}(\mathbf{0}) = \{x \mid Tx = \mathbf{0}\}$$

is called the *kernel* (or *null space*) of T and

$$\operatorname{Im} T = T(V) = \{Tx \mid x \in V\}$$

is called the *image* (or *range*) of T.

Corollary 2.2.4. *If $T\colon V \to W$ is a linear map, then* $\operatorname{Ker} T$ *is a subspace of V and* $\operatorname{Im} T$ *is a subspace of W.*

Proof. We take $M = V$ and $N = \{\mathbf{0}\}$ in Theorem 2.2.2. □

Example 2.2.5. Let $D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the linear map given by differentiation. Then

$$\operatorname{Ker} D = \{f \in C^\infty(\mathbb{R}) \mid f \text{ is constant}\},$$
$$\operatorname{Im} D = C^\infty(\mathbb{R}).$$

Why is $\operatorname{Im} T$ all of $C^\infty(\mathbb{R})$? Suppose $f \in C^\infty(\mathbb{R})$. Define F by

$$F(x) = \int_0^x f(t)\,dt.$$

Then we know from calculus that F is differentiable and $DF = F' = f$ (this shows that $F \in C^\infty(\mathbb{R})$). Hence every $f \in C^\infty(\mathbb{R})$ is in the image of D.

Example 2.2.6. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the linear form defined by $f(x_1, x_2) = x_2 - 3x_1$. Then $\operatorname{Im} f = \mathbb{R}$ and the kernel of f is a line through the origin (the line $x_2 = 3x_1$).

Example 2.2.7. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x_1, x_2) = (x_2, 0, x_1)$. Then $\operatorname{Ker} T = \{\mathbf{0}\}$ and the image of T is the 'x, z-plane' in $\mathbb{R}^3$.

Theorem 2.2.8. *Suppose $T\colon V \to W$ is a linear map. Then T is injective if and only if* $\operatorname{Ker} T = \{\mathbf{0}\}$.

Proof. Suppose $\operatorname{Ker} T = \{\mathbf{0}\}$. Then, for $v, v' \in V$,

$$Tv = Tv' \implies Tv - Tv' = \mathbf{0} \implies T(v - v') = \mathbf{0} \implies v - v' = \mathbf{0} \implies v = v'.$$

Thus T is injective. Now suppose T is injective. Then for $v \in V$,

$$Tv = \mathbf{0} = T\mathbf{0} \implies v = \mathbf{0}.$$

Hence $\operatorname{Ker} T = \{\mathbf{0}\}$. □

Remark 2.2.9. Note that Theorem 2.2.8 only applies to linear maps. For instance, let $f\colon \mathbb{R} \to \mathbb{R}$ by the map defined by $f(x) = x^2$. Then $f^{-1}(\{0\}) = \{0\}$ but f is *not* injective since, for instance, $f(1) = f(-1)$.

Exercises.

2.2.1 ([Ber14, Ex. 2.1.7]). Let $\mathcal{P}(\mathbb{R})$ be the vector space of all real polynomial functions (see Example 1.2.8) and define

$$f\colon \mathcal{P} \to \mathbb{R}, \quad f(p) = p'(1).$$

Prove that f is a linear form on $\mathcal{P}$. What is the geometric meaning of the kernel of f?

2.2.2. Suppose $S\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is the map of Example 2.1.6(c). What is $\operatorname{Im} S$? *Hint:* S is *not* surjective.

2.2.3 ([Ber14, Ex. 2.2.2]). Let V be a vector space over F, and let $f\colon V \to F$ be a linear form on V. Assume that f is not identically zero and choose a vector v such that $f(v) \neq 0$. Let $N = \operatorname{Ker} f$. Prove that, for every vector $u \in V$, there exist unique $z \in N$ and $c \in F$ such that $u = z + cv$. *Hint*: If $u \in V$, compute the value of f on the vector $u - (f(u)/f(v))v$.

2.2.4 ([Ber14, Ex. 2.2.3]). Let $S\colon V \to W$ and $T\colon V \to W$ be linear maps, and let

$$M = \{v \in V \mid S(v) \in T(V)\}.$$

Prove that M is a subspace of V.

2.2.5 ([Ber14, Ex. 2.2.6]). Let $\mathcal{P}$ be the space of real polynomial functions and let $T\colon \mathcal{P} \to \mathcal{P}$ be the mapping defined by $Tp = p - p'$, where p' is the derivative of p. Prove that T is linear and injective.

2.2.6 ([Ber14, Ex. 2.3.15]). Let $T\colon U \to V$ and $S\colon V \to W$ be linear mappings. Prove:

(a) If S is injective, then $\operatorname{Ker}(ST) = \operatorname{Ker}(T)$.

(b) If T is surjective, then $\operatorname{Im}(ST) = \operatorname{Im}(S)$.

2.2.7. Let V be a vector space and $S, T \in \mathcal{L}(V)$. Show that if $ST = TS$, then $S(\operatorname{Ker} T) \subseteq \operatorname{Ker} T$.

2.2.8. Suppose V is a vector space and consider the linear mapping (see Exercise 2.1.4)

$$T\colon V \times V \to V, \quad T(u, v) = u - v.$$

Determine the kernel and image of T.

2.3 Vector spaces of linear maps

Definition 2.3.1. Suppose V and W are vector spaces over a field F. We define

$$\mathcal{L}(V, W) = \{T\colon V \to W \mid T \text{ is linear}\}.$$

In the case where $V = W$, we write $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$.

Example 2.3.2 (Zero linear map). For any vector spaces V and W over a field F, we have the *zero linear map* which maps every element of V to $\mathbf{0} \in W$. We denote the map by 0 (so $0(v) = \mathbf{0}$ for all $v \in V$). Thus $0 \in \mathcal{L}(V, W)$.

Example 2.3.3 (Identity map). The map $I\colon V \to V$ defined by $Iv = v$ for all $v \in V$ is linear and is called the *identity linear map*. We sometimes write I_V when we want to keep track of the vector space on which it acts.

Example 2.3.4 (Scalar linear map). If a is any scalar, then the map $T\colon V \to V$ defined by $Tv = av$ is linear (see Example 2.1.9).

Example 2.3.5 (Linear forms). The elements of $\mathcal{L}(V, F)$ are precisely the linear forms on V.

Recall that in MAT 1341, you learned that every linear map f from $\mathbb{R}^n$ to $\mathbb{R}^m$ (whose elements are written as column vectors) is given by multiplication by the $m \times n$ matrix

$$A = \begin{bmatrix} f(e_1) & f(e_2) & \cdots & f(e_n) \end{bmatrix},$$

where $\{e_1, \ldots, e_n\}$ is the *standard basis* of $\mathbb{R}^n$ (see Example 1.4.8) and $f(e_j)$ is the jth column of A. The matrix A is called the *standard matrix* of the linear map.

Example 2.3.6. The linear map $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T(x, y, z) = (y, z)$$

has standard matrix

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

since

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix} = T \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Later in the course, we'll return to the topic of matrices. We'll see matrices with entries in an arbitrary field F and their relation to linear maps between vectors spaces over F.

Our goal now is to turn $\mathcal{L}(V, W)$ into a vector space itself. So we need to define vector addition and scalar multiplication and then check the axioms of a vector space.

Definition 2.3.7. Suppose V and W are vector spaces over a field F and $c \in F$. For $S, T \in \mathcal{L}(V, W)$, define $S + T$ and cT by

$$\begin{aligned} (S + T)(v) &= S(v) + T(v), \quad \forall\ v \in V, \\ (cT)(v) &= cT(v), \quad \forall\ v \in V. \end{aligned}$$

Lemma 2.3.8. *If $S, T \in \mathcal{L}(V, W)$ and a is a scalar, then $S + T$, $aT \in \mathcal{L}(V, W)$.*

Proof. We first show that $S + T$ is linear. For $u, v \in V$ and $c, d \in F$, we have

$$\begin{aligned} (S + T)(cu + dv) &= S(cu + dv) + T(cu + dv) \\ &= cS(u) + dS(v) + cT(u) + dT(v) \\ &= cS(u) + cT(u) + dS(v) + dT(v) \\ &= c(S(u) + T(u)) + d(S(v) + T(v)) \\ &= c(S + T)(u) + d(S + T)(v). \end{aligned}$$

Therefore, $S + T$ is linear. We leave it as an exercise (Exercise 2.3.1) to show that aT is linear. □

Theorem 2.3.9. *If V and W are vector spaces over a field F, then $\mathcal{L}(V, W)$ is also a vector space over F (with the operations defined in Definition 2.3.7).*

Proof. We must show that the axioms in Definition 1.2.1 are satisfied.

Vector addition is commutative and associative since for all $S, T, U \in \mathcal{L}(V, W)$ and $v \in V$, we have

$$(S + T)(v) = S(v) + T(v) = T(v) + S(v) = (T + S)(v),$$

and

$$\begin{aligned} ((S + T) + U)(v) &= (S + T)(v) + U(v) = S(v) + T(v) + U(v) \\ &= S(v) + (T + U)(v) = (S + (T + U))(v). \end{aligned}$$

The zero linear map is an identity for the operation of vector addition since $T + \mathbf{0} = T$ for all $T \in \mathcal{L}(V, W)$.

For all $T \in \mathcal{L}(V, W)$, the map $-T = (-1)T$ is linear and is the inverse of T under the operation of vector addition since

$$(T + (-T))(v) = T(v) + (-1)T(v) = \mathbf{0} \quad \forall\, v \in V.$$

We leave it as an exercise (Exercise 2.3.1) to verify the remaining axioms of Definition 1.2.1. □

Example 2.3.10. Fix $c \in \mathbb{R}$ and define a map $\varphi_c \colon \mathcal{F}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ by

$$\varphi_c(f) = f(c) \quad \forall\, f \in \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

We show that $\varphi_c \in \mathcal{F}(\mathbb{R}, \mathbb{R})^*$ (see Definition 2.1.10). To do this, we need to show that φ_c is linear. For $k_1, k_2 \in \mathbb{R}$ and $f_1, f_2 \in \mathcal{F}(\mathbb{R}, \mathbb{R})$, we have

$$\varphi_c(k_1 f_1 + k_2 f_2) = (k_1 f_1 + k_2 f_2)(c) = k_1 f_1(c) + k_2 f_2(c) = k_1 \varphi_c(f_1) + k_2 \varphi_c(f_2).$$

Thus φ_c is linear. This example is important in quantum physics. The so-called 'delta function' which describes the state of a particle located at the point is this type of linear form (i.e. evaluation at the point).

Definition 2.3.11 (Composition of maps)**.** If A, B and C are sets (so, for instance, they could be vector spaces), and $T \colon A \to B$ and $S \colon B \to C$ are maps, we write $ST \colon A \to C$ for the *composite map* (or simply the *composition*) defined by

$$(ST)a = S(Ta) \quad \forall a \in A.$$

Composition of maps is associative, so if $R \colon C \to D$ is a third map, we have $(RS)T = R(ST)$. We simply write RST for this triple composition.

Remark 2.3.12. Sometimes composition maps are written as $S \circ T$. We use the shorter notation ST to remind us of matrix multiplication. As in MAT 1341, we'll see that composition and matrix multiplication are closely related.

Theorem 2.3.13. *The composition of linear maps is a linear map. In other words, if $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $ST \in \mathcal{L}(U, W)$.*

Proof. For $u, u' \in U$ and c a scalar, we have

$$\begin{aligned}
(ST)(u + u') &= S(T(u + u')) \\
&= S(Tu + Tu') \\
&= S(Tu) + S(Tu') \\
&= (ST)(u) + (ST)(u'), \\
(ST)(cu) = S(T(cu)) &= S(c(Tu)) = c(S(Tu)) = c((ST)u).
\end{aligned}$$

□

Note that if $U = V = W$, then the above theorem tells us that when $S, T \in \mathcal{L}(V)$, we have $ST \in \mathcal{L}(V)$. Therefore we have *three* operations on $\mathcal{L}(V)$: addition, multiplication by scalars, and composition.

Definition 2.3.14 (Powers of a map)**.** If T is a map from some set to itself (for instance, if $T \in \mathcal{L}(V)$), then the *powers* of T are defined by

$$T^1 = T, \quad T^2 = TT, \quad T^3 = TT^2, \quad \ldots, \quad T^n = TT^{n-1}, \quad n \geq 2.$$

We also define T^0 to be the identity map.

Exercises.

2.3.1. Complete the proof of Lemma 2.3.8 by showing that aT is linear.

2.3.2. Complete the proof of Theorem 2.3.9 by verifying the remaining axioms of Definition 1.2.1.

2.3.3 ([Ber14, Ex. 2.3.2]). Let V be a vector space. If $R, S, T \in \mathcal{L}(V)$ and c is a scalar, prove the following statements:

(a) $(RS)T = R(ST)$;

(b) $(R + S)T = RT + ST$;

(c) $R(S + T) = RS + RT$;

(d) $(cS)T = c(ST) = S(cT)$;

(e) $TI = T = IT$, where I is the identity mapping.

2.3.4. If $S, T \in \mathcal{L}(\mathbb{R}^3)$ are defined by

$$S(x, y, z) = (x - 2y + 3z, y - 2z, 4y),$$
$$T(x, y, z) = (y - z, 2x + y, x + 2x),$$

find explicit expressions for $S + T$, $2T$, $S - T$, and ST.

2.3.5. Suppose $T\colon V \to V$ is a linear map such that $\operatorname{Im} T \subseteq \operatorname{Ker}(T - I)$. Prove that $T^2 = T$.

2.3.6 ([Ber14, Ex. 2.3.9]). Let U, V, W be vector spaces over F. Prove the following:

(a) For fixed $T \in \mathcal{L}(U, V)$, the map

$$\mathcal{L}(V, W) \to \mathcal{L}(U, W), \quad S \mapsto ST,$$

is linear.

(b) For fixed $S \in \mathcal{L}(V, W)$, the map

$$\mathcal{L}(U, V) \to \mathcal{L}(U, W), \quad T \mapsto ST,$$

is linear.

2.3.7 ([Ber14, Ex. 2.3.12]). Suppose $V = U \oplus W$. For each $v \in V$, let $v = u + w$ be its unique decomposition with $u \in U$ and $w \in W$, and define $Pv = u$, $Qv = w$. Prove that $P, Q \in \mathcal{L}(V)$, $P^2 = P$, $Q^2 = Q$, $P + Q = I$, and $PQ = QP = \mathbf{0}$.

2.3.8 ([Ber14, Ex. 2.3.13]). Let $T\colon V \to V$ be a linear map such that $T^2 = T$, and let

$$U = \operatorname{Im} T, \quad W = \operatorname{Ker} T.$$

Prove that $U = \{v \in V \mid Tv = v\} = \operatorname{Ker}(T - I)$ and that $V = U \oplus W$.

2.3.9 ([Ber14, Ex. 2.3.14]). Give an example of $S, T \in \mathcal{L}(\mathbb{R}^2)$ such that $ST \neq TS$.

2.3.10 ([Ber14, Ex. 2.3.16]). Let V be a real or complex vector space and let $T \in \mathcal{L}(V)$ be such that $T^2 = I$. Define

$$M = \{v \in V \mid Tv = v\}, \quad N = \{v \in V \mid Tv = -v\}.$$

Prove that M and N are subspaces of V and that $V = M \oplus N$. *Hint:* For every vector v, we have

$$v = \frac{1}{2}(v + Tv) + \frac{1}{2}(v - Tv).$$

2.3.11 ([Ber14, Ex. 2.3.19]). Let $S, T \in \mathcal{L}(V, W)$ and let $U = \{v \in V \mid Sv = Tv\}$. Prove that U is a subspace of V. *Hint*: Consider $\operatorname{Ker}(S - T)$.

2.3.12 ([Ber14, Ex. 2.3.20]). If $T\colon V \to W$ and $S\colon W \to V$ are linear maps such that $ST = I$, prove that $\operatorname{Ker} T = \{\mathbf{0}\}$ and $\operatorname{Im} S = V$.

2.3.13 ([Ber14, Ex. 2.3.24]). Let $V = \mathcal{P}$ be the vector space of real polynomial functions, and let $D\colon V \to V$ be the differentiation mapping $Dp = p'$. Let $u \in \mathcal{P}$ be the monomial $u(t) = t$, and define another linear map which is multiplication by t:

$$M\colon \mathcal{P} \to \mathcal{P}, \quad Mp = up.$$

Prove that $DM - MD = I$. *Hint*: You need to show that $(up)' - up' = p$ for all $p \in \mathcal{P}$. Remember the product rule.

2.4 Isomorphisms

We will now discuss a precise way in which certain vector spaces are the 'same' (but not necessarily equal).

Definition 2.4.1 (Isomorphism)**.** An *isomorphism* is a bijective linear map $T\colon V \to W$, where V and W are vector spaces. We say a vector space V is *isomorphic* to another vector space W (over the same field) if there is an isomorphism $T\colon V \to W$. We write $V \cong W$ to indicate that V is isomorphic to W.

Remark 2.4.2. One should think of isomorphic vector spaces as being 'the same' as far as their vector space properties go. Of course, they make look quite different (and are, in general, not equal). But the isomorphism identifies the two in a way that preserves the operations of a vector space (vector addition and scalar multiplication).

Example 2.4.3. Let

$$\begin{aligned} V &= \mathbb{R}^2, \\ W &= \{(x, y, 0) \mid x, y \in \mathbb{R}\}. \end{aligned}$$

Then $V \cong W$. In order to prove this, we need to find a specific isomorphism. One isomorphism is the map

$$T\colon V \to W, \quad T(x, y) = (x, y, 0).$$

We leave it as an exercise (Exercise 2.4.1) to check that the map T is an isomorphism.

Note that there is more than one isomorphism from V to W. For instance,

$$T_1(x, y) = (y, x, 0) \quad \text{and} \quad T_2(x, y) = (x + y, x - y, 0)$$

are two other isomorphisms from V to W.

Example 2.4.4. Consider the map

$$T\colon \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3, \quad T(at^2 + bt + c) = (a, b, c).$$

You were asked to show in Exercise 2.1.1 that the map T is linear. It is easy to see that T is bijective. Thus, T is an isomorphism. Hence $\mathcal{P}_2(\mathbb{R}) \cong \mathbb{R}^3$ (as real vector spaces).

Example 2.4.5. One can actually show that $\mathcal{P}_n(F) \cong F^{n+1}$ for any *infinite* field F. However, things can go wrong if the field is finite. For instance if we work over the field $\mathbb{F}_2$, then the polynomial functions t and t^2 are equal! To check this, we just need to check that they take the same value on all of the elements of $\mathbb{F}_2$. Since $0 = 0^2$ and $1 = 1^2$, this verifies that the polynomial functions t and t^2 are indeed equal.

Remark 2.4.6. There is a difference between the symbols $\to$ and $\mapsto$. We use $\to$ when we write $T\colon V \to W$ to indicate that T is a map with domain V and codomain W. We use $\mapsto$ to describe a map. So $x \mapsto Tx$ is the map that sends x to Tx.

Example 2.4.7. Let $U = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$. We know from techniques learned in MAT 1341 that U is a subspace of $\mathbb{R}^3$ since it is the solution set to a system of homogeneous equations. In fact, you can use the techniques of MAT 1341 to show that

$$U = \operatorname{Span}\{(-1, 1, 0), (-1, 0, 1)\}$$

Define

$$T\colon \mathbb{R}^2 \to U, \quad T(x,y) = x(-1,1,0) + y(-1,0,1) = (-x-y, x, y).$$

Then T is linear since it corresponds to multiplication by the matrix

$$\begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We leave it as an exercise (Exercise 2.4.4) to show that T is bijective. Note that it is *not* bijective as a map $\mathbb{R}^2 \to \mathbb{R}^3$, but it *is* bijective as a map $\mathbb{R}^2 \to U$. So T is an isomorphism and $\mathbb{R}^2 \cong U$.

Example 2.4.8. Remember that $\mathbb{C}$ can be thought of as a real vector space. Then

$$(a,b) \mapsto a + bi, \quad a, b \in \mathbb{R}$$

is an isomorphism from $\mathbb{R}^2$ to $\mathbb{C}$. So $\mathbb{R}^2 \cong \mathbb{C}$ *as real vector spaces*. Not that it is especially important here to state what *type* of isomorphism we have (that is, an isomorphism of real vector spaces). We are not using the complex vector space structure on $\mathbb{C}$.

Recall that an *inverse* of a map $f\colon A \to B$ is a map $g\colon B \to A$ such that gf is the identity map on A and fg is the identity map on B. In other words,

$$(gf)(a) = a \;\forall\; a \in A, \quad \text{and} \quad (fg)(b) = b \;\forall\; b \in B.$$

We write f^{-1} for such a map g.

Remember that bijective maps have inverses. If $f\colon A \to B$ is a bijective map from a set A to a set B, then we can define the inverse map $f^{-1}\colon B \to A$, by

$$f^{-1}(b) = a \iff f(a) = b.$$

In other words, for any $b \in B$, $f^{-1}(b)$ is defined to be the unique element a of A such that $f(a) = b$. Such an a exists since f is surjective and it is unique since f is injective.

Theorem 2.4.9. *Suppose V and W are vector spaces over a field F. If $T\colon V \to W$ is a bijective linear map (i.e. an isomorphism), then the inverse map $T^{-1}\colon W \to V$ is also linear (hence is an isomorphism, since it is bijective by Exercise 2.4.6).*

Proof. Let $w, w' \in W$ and $c, c' \in F$. We want to show that

$$T^{-1}(cw + c'w') = cT^{-1}w + c'T^{-1}w'. \tag{2.1}$$

Recall that since T is injective, for any $u, v \in V$, we have that $Tu = Tv$ implies $u = v$. So let's apply T to each side of 2.1 and try to show that the results are equal. Applying T to the left-hand side, we get

$$TT^{-1}(cw + c'w') = cw + c'w'.$$

Now, applying T to the right-hand side, we get

$$T(cT^{-1}w + c'T^{-1}w') = cTT^{-1}w + c'TT^{-1}w' = cw + cw',$$

where we have used the fact that T is linear. We thus see that

$$T(T^{-1}(cw + c'w')) = T(cT^{-1}w + c'T^{-1}w')$$

and so $T^{-1}(cw + c'w') = cT^{-1}w + c'T^{-1}w'$ by the injectivity of T. □

Remark 2.4.10. By Exercise 2.4.6, if a linear map $T\colon V \to W$ has an inverse, then it is an isomorphism. This can be a useful way to show that a give linear map is an isomorphism.

Example 2.4.11. If $A \in M_n(\mathbb{R})$ is any invertible matrix, then the map $\mathbb{R}^n \to \mathbb{R}^n$ defined by $v \mapsto Av$ for $v \in \mathbb{R}^n$ is an isomorphism. We know this because it is linear (from MAT 1341) and has an inverse, namely the map $\mathbb{R}^n \to \mathbb{R}^n$ given by $v \mapsto A^{-1}v$ for $v \in \mathbb{R}^n$. This is indeed an inverse map since

$$AA^{-1}v = Iv = v \quad \text{and} \quad A^{-1}Av = Iv = v \quad \forall v \in V.$$

We now collect some nice properties of isomorphism. The next theorem states that isomorphism is an *equivalence relation* (see Definition B.1.1).

Theorem 2.4.12. *Suppose U, V, W are vector spaces over the same field. Then*

(a) $V \cong V$ (isomorphism is reflexive*),*

(b) if $V \cong W$, then $W \cong V$ (isomorphism is symmetric*),*

(c) if $U \cong V$ and $V \cong W$, then $U \cong W$ (isomorphism is transitive*).*

Proof. (a) The identity map $I\colon V \to V$ is an isomorphism.

(b) We now know that the inverse of an isomorphism is itself an isomorphism. So suppose $V \cong W$. Then there is an isomorphism $T\colon V \to W$. Thus, the inverse $T^{-1}\colon W \to V$ is an isomorphism. Hence $W \cong V$.

(c) If $T\colon U \to V$ and $S\colon V \to W$ are isomorphisms, then the composition $ST\colon U \to W$ is also linear and bijective, hence is an isomorphism.

□

Exercises.

2.4.1. Prove that the maps T, T_1, and T_2 of Example 2.4.3 are linear. So you need to show they are linear and bijective.

2.4.2. Show that $\mathcal{P}_n(\mathbb{R})$ is isomorphic to $\mathbb{R}^{n+1}$.

2.4.3. Let $A = \{1, 2, \ldots, n\}$ and $V = \mathcal{F}(A, F)$ for some field F. Show $V \cong F^n$ via the bijection $x \mapsto (x(1), x(2), \ldots, x(n))$.

2.4.4. Show that the map T of Example 2.4.7 is bijective.

2.4.5. Show that $\mathbb{R}^{2n} \cong \mathbb{C}^n$ as real vector spaces.

2.4.6. Prove the following statements:

(a) The inverse of a bijective map is also bijective.

(b) If $f\colon A \to B$ has an inverse, then f is bijective. Combining this with the above remarks, this means that a map is bijective if and only if it has an inverse.

(c) The composition of two injective maps is itself an injective map.

(d) The composition of two surjective maps is itself a surjective map.

(e) The composition of two bijective maps is itself a bijective map.

Students who took MAT 1362 will have already seen these statements (see [Sav, §8.1]).

2.4.7 ([Ber14, Ex. 2.4.5]). Let V be a vector space, $T \in \mathcal{L}(V)$ bijective, and c a nonzero scalar. Prove that cT is bijective and that $(cT)^{-1} = c^{-1}T^{-1}$.

2.4.8 ([Ber14, Ex. 2.4.6]). Let $T\colon V \to W$ be an isomorphism. For each $S \in \mathcal{L}(V)$, define $\varphi(S) = TST^{-1}$ (note that the product is defined). Prove that $\varphi\colon \mathcal{L}(V) \to \mathcal{L}(W)$ is an isomorphism and that $\varphi(RS) = \varphi(R)\varphi(S)$ for all $R, S \in \mathcal{L}(V)$. Don't forget that you need to show φ is itself a linear map.

2.4.9 ([Ber14, Ex. 2.4.7]). Let V be a vector space, and let $T \in \mathcal{L}(V)$. Prove the following:

(a) If $T^2 = 0$, then $I - T$ is bijective.

(b) If $T^n = 0$ for some positive integer n, then $I - T$ is bijective.

Hint: In polynomial algebra, $(1 - t)(1 + t) = 1 - t^2$.

2.4.10. Let U, V, W be vector spaces over the same field. Prove the following:

(a) $(U \times V) \times W \cong U \times (V \times W)$.

(b) $V \times W \cong W \times V$.

2.4.11. For positive integers n and m, prove that $\mathbb{R}^n \times \mathbb{R}^m \cong \mathbb{R}^{n+m}$.

Chapter 3

Structure of vector spaces

In this chapter we will explore the structure of vector spaces is more detail. In particular, we analyze the concepts of generating sets, linear dependence/independence, bases, and dimension. We also discuss the important notation of the dual space. The material in this chapter roughly corresponds to [Tre, §1.2, §3.5, §8.1].

3.1 Spans and generating sets

Recall (Definition 1.4.4) that if A is a nonempty subset of a vector space V over a field F, then

$$\operatorname{Span} A = \{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n \mid n \in \mathbb{N},\ c_i \in F,\ v_i \in A \text{ for } 1 \leq i \leq n\}.$$

We sometimes write $\operatorname{Span}_F A$ when we want to emphasize the field.

Theorem 3.1.1. *Suppose A is a nonempty subset of a vector space V. Then $\operatorname{Span} A$ is a subspace of V and is the smallest subspace of V containing A. In other words*

(a) $A \subseteq \operatorname{Span} A$, and

(b) if W is a subspace of V such that $A \subseteq W$, then $\operatorname{Span} A \subseteq W$.

Proof. We already noted in Theorem 1.5.4 that $\operatorname{Span} A$ is a subspace of V, so it remains to proves that it is the smallest one containing A. It is clear that $A \subseteq \operatorname{Span} A$ since every element $a \in A$ is a linear combination of elements of A (with one term and coefficient one). Now suppose W is a subspace of V containing A. We wish to show that $\operatorname{Span} A \subseteq W$. Let $v \in \operatorname{Span} A$. Then, by definition,

$$v = \sum_{i=1}^{n} c_i v_i$$

for some $c_1, \ldots, c_n \in F$ and $v_1, \ldots, v_n \in A$. Since $A \subseteq W$, each $v_i \in W$, for $1 \leq i \leq n$. Then, since W is a subspace and therefore closed under scalar multiples and vector addition, $\sum_{i=1}^{n} c_i v_i \in W$. So $v \in W$. Thus $\operatorname{Span} A \subseteq W$ as desired. □

Remark 3.1.2. If we adopt the convention that $\operatorname{Span} \varnothing = \{\mathbf{0}\}$, then Theorem 3.1.1 remains true with the word 'nonempty' removed.

Definition 3.1.3 (Generating set)**.** The space $\operatorname{Span} A$ is called the subspace of V *generated* by A, we call A a *generating set* for $\operatorname{Span} A$, and say that A *generates* $\operatorname{Span} A$. So if $\operatorname{Span} A = V$, then we say A generates V.

Theorem 3.1.4. *Suppose $T\colon V \to W$ is a linear map and A is a subset of V. Then* $\operatorname{Span} T(A) = T(\operatorname{Span} A)$.

Proof. Since $A \subseteq \operatorname{Span} A$, we have $T(A) \subseteq T(\operatorname{Span} A)$. Also, since $\operatorname{Span} A$ is a subspace of A, we know $T(\operatorname{Span} A)$ is a subspace of W by Theorem 2.2.2. Hence, by Theorem 3.1.1, $\operatorname{Span} T(A) \subseteq T(\operatorname{Span} A)$.

It remains to prove the reverse in inclusion. Since $T(A) \subseteq \operatorname{Span} T(A)$, we have $A \subseteq T^{-1}(\operatorname{Span} T(A))$. Also, since $\operatorname{Span} T(A)$ is a subspace of W, we know $T^{-1}(\operatorname{Span} T(A))$ is a subspace of V by Theorem 2.2.2. Thus, by Theorem 3.1.1, we have $\operatorname{Span} A \subseteq T^{-1}(\operatorname{Span} T(A))$. Thus $T(\operatorname{Span} A) \subseteq \operatorname{Span} T(A)$. □

Corollary 3.1.5. *If $T\colon V \to W$ is a surjective linear map and A generates V, then $T(A)$ generates W.*

Proof. We have

$$\begin{aligned} W &= T(V) && \text{(since } T \text{ is surjective)} \\ &= T(\operatorname{Span} A) && \text{(since } A \text{ generates } V\text{)} \\ &= \operatorname{Span} T(A) && \text{(by Theorem 3.1.4).} \end{aligned}$$

Therefore $T(A)$ generates W. □

Exercises.

3.1.1 ([Ber14, Ex. 3.1.1]). If $T\colon V \to W$ is linear and A is a subset of V such that $T(A)$ generates W, then T is surjective.

3.1.2 ([Ber14, Ex. 3.1.3]). Suppose $x, x_1, \ldots, x_n$ are vectors such that x is a linear combination of $x_1, \ldots, x_n$, but not of $x_1, \ldots, x_{n-1}$. Prove that

$$\operatorname{Span}\{x_1, \ldots, x_n\} = \operatorname{Span}\{x_1, \ldots, x_{n-1}, x\}.$$

Hint: When x is represented as a linear combination of $x_1, \ldots, x_n$, the coefficient of x_n cannot be zero.

3.1.3 ([Ber14, Ex. 3.1.4]). Let $S\colon V \to W$ and $T\colon V \to W$ be linear mappings, and let A be a subset of V such that $\operatorname{Span} A = V$. Prove that, if $Sx = Tx$ for all $x \in A$, then $S = T$.

3.1.4. Let $T\colon V \to W$ be a linear map, and let $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ be two lists of vectors in V. Suppose that

(a) $x_1, \ldots, x_m$ generate $\operatorname{Ker} T$, and

(b) $T(y_1), \ldots, T(y_n)$ generate W.

Show that the list $x_1, \ldots, x_m, y_1, \ldots, y_n$ generates V.

3.2 Linear dependence/independence

Definition 3.2.1 (Linearly dependent/independent)**.** Suppose $v_1, v_2, \ldots, v_n$ is a finite list of vectors in a vector space V. We say that the list is (or the vectors $v_1, v_2, \ldots, v_n$ themselves are) *linearly dependent* (or simply *dependent*) if there exist scalars $c_1, c_2, \ldots, c_n$ such that

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0}.$$

and at least one of the c_i, $1 \leq i \leq n$, is nonzero. Such an equation (in which at least one of the c_i is nonzero), is called a *dependence relation* among the v_i.

If $v_1, v_2, \ldots, v_n$ are not linearly dependent, they are said to be *linearly independent*. In other words, the vectors $v_1, v_2, \ldots, v_n$ are linearly independent if

$$c_1v_1 + \cdots + c_nv_n = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0,$$

where $c_1, \ldots, c_n$ are scalars.

Remark 3.2.2. By the commutativity of vector addition, the order of the vectors in the list is not important in the definition of linear dependence/independence.

Example 3.2.3. Consider the case $n = 1$. The list v_1 is linearly dependent if and only if $v_1 = \mathbf{0}$. This is because $c_1v_1 = \mathbf{0}$ for some *nonzero* scalar c_1 if and only if $v_1 = 0$ (by Theorem 1.3.4).

Example 3.2.4. Consider the case $n = 2$. Then the list v_1, v_2 is linearly dependent if and only if one of the vectors is a multiple of the other. To see this, first suppose that $v_2 = cv_1$. Then $cv_1 + (-1)v_2 = \mathbf{0}$ is a dependence relation and so v_1, v_2 are linearly dependent. The same argument works for the case that v_1 is a multiple of v_2. Now suppose that v_1, v_2 are linearly dependent. Then we have a dependence relation $c_1v_1 + c_2v_2 = \mathbf{0}$ where c_1 or c_2 (or both) is nonzero. If $c_1 \neq 0$, then $v_1 = (-c_2/c_1)v_2$ and so v_1 is a multiple of v_2. Similarly, if $c_2 \neq 0$, then v_2 is a multiple of v_1.

Lemma 3.2.5. *(a) Any list of vectors containing the zero vector is linearly dependent.*

(b) If the vectors $v_1, v_2, \ldots, v_n$ are not distinct (i.e. some vector appears more than once in the list), then they are linearly dependent.

Proof. Consider a list of vectors $v_1, \ldots, v_n$.

(a) If $v_i = \mathbf{0}$ for some i, $1 \leq i \leq n$, then

$$0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n$$

is a dependence relation and so the vectors are linearly dependent.

(b) If $v_i = v_j$ for some $1 \leq i < j \leq n$, then

$$0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_{j-1} + (-1)v_j + 0v_{j+1} + \cdots + v_n$$

is a dependence relation and so the vectors are linearly dependent.

□

Corollary 3.2.6. *Consider a list of vectors $v_1, \dots, v_n$.*

(a) If $v_1, \dots, v_n$ are linearly independent, then none of them is the zero vector.

(b) If $v_1, \dots, v_n$ are linearly independent, then no two of them are equal.

Theorem 3.2.7. *Suppose V is a vector space over a field F and $v_1, v_2, \dots, v_n \in V$. Define a linear map $T\colon F^n \to V$ by*

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

(We know from Theorem 2.1.4 that this map is linear.) Then the vectors $v_1, \dots, v_n$ are linearly dependent if and only if T is not injective.

Proof. Since T is linear, we have

$$\begin{aligned} T \text{ is not injective} &\iff \operatorname{Ker} T \neq \{\mathbf{0}\} \\ &\iff \exists\ (a_1, \dots, a_n) \neq (0, \dots, 0) \text{ such that } T(a_1, \dots, a_n) = \mathbf{0} \\ &\iff \exists\ (a_1, \dots, a_n) \neq (0, \dots, 0) \text{ such that } a_1 v_1 + \cdots + a_n v_n = \mathbf{0} \\ &\iff v_1, \dots, v_n \text{are linearly dependent.} \end{aligned}$$

□

Corollary 3.2.8. *Under the hypotheses of Theorem 3.2.7, the map T is injective if and only if the vectors $v_1, \dots, v_n$ are linearly independent.*

Theorem 3.2.9. *Let V be a vector space and $v_1, \dots, v_n \in V$, $n \geq 2$. Then the following statements are equivalent:*

(a) $v_1, \dots, v_n$ are linearly dependent,

(b) some v_j is a linear combination of the v_i with $i \neq j$ (i.e. some vector is a linear combination of the others),

(c) either $v_1 = \mathbf{0}$ or there exists an index $j > 1$ such that v_j is a linear combination of $v_1, \dots, v_{j-1}$.

Proof. You saw this result for real vector spaces in MAT 1341. The proof for vector spaces over an arbitrary field is essentially the same and so we will omit it. See also [Tre, Prop. 2.6].

□

Corollary 3.2.10. *Let V be a vector space and $v_1, \dots, v_n \in V$, $n \geq 2$. Then the following statements are equivalent:*

(a) $v_1, \dots, v_n$ are linearly independent,

(b) no v_j is a linear combination of the v_i with $i \neq j$,

(c) $v_1 \neq \mathbf{0}$ and for every $j > 1$, v_j is not a linear combination of the $v_1, \dots, v_{j-1}$.

Example 3.2.11. The vectors $e_1, \dots, e_n \in F^n$ are linearly independent. This is because

$$c_1 e_1 + \cdots + c_n e_n = \mathbf{0} \implies (c_1, \dots, c_n) = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0.$$

Example 3.2.12. Consider the vectors $\sin x, \sin 2x$ in $\mathcal{F}(\mathbb{R})$. Suppose

$$a \sin x + b \sin 2x = 0 \quad \forall\, x \in \mathbb{R}.$$

Note that we insist that the equality hold for *all* $x \in \mathbb{R}$ since equality in the vector space $\mathcal{F}(\mathbb{R})$ is equality *of functions*. So it must hold, in particular, for $x = \pi/2$, and so

$$a \sin \frac{\pi}{2} + b \sin \frac{2\pi}{2} = 0 \implies a + 0 = 0 \implies a = 0.$$

Then, taking $x = \pi/4$, gives

$$a \sin \frac{\pi}{4} + b \sin \frac{2\pi}{4} = 0 \implies 0 \sin \frac{\pi}{4} + b = 0 \implies b = 0.$$

Thus, the vectors $\sin x, \sin 2x$ are linearly independent.

Theorem 3.2.13. *If $T\colon V \to W$ is an injective linear map and $v_1, \dots, v_n$ are linearly independent vectors in V, then $Tv_1, \dots, Tv_n$ are linearly independent in W.*

Proof. For scalars $c_1, \dots, c_n$, we have

$$\begin{aligned}
&c_1(Tv_1) + \cdots + c_n(Tv_n) = \mathbf{0} && \\
\implies\ &T(c_1v_1 + \cdots + c_nv_n) = \mathbf{0} && (T \text{ is linear}) \\
\implies\ &c_1v_1 + \cdots + c_nv_n = \mathbf{0} && (T \text{ is injective}) \\
\implies\ &c_1 = c_2 = \cdots = c_n = 0 && (\text{the vectors } v_1, \dots, v_n \text{ are linearly independent}).
\end{aligned}$$

□

Note that it is very important in the above theorem that T be injective. For instance, it would certainly fail to be true if T were the zero map since then the list $Tv_1, \dots, Tv_n$ would contain the zero vector and so could not be linearly independent.

Definition 3.2.14 (Linear dependence/independence of sets)**.** If A is any (possibly infinite) set of vectors in a vector space V, we say A is *linearly dependent* if some finite list $v_1, v_2, \dots, v_n$ of distinct vectors in A is linearly dependent (in other words, there is some dependence relation involves a finite number of the vectors of A). We say A is *linearly independent* if every finite list of distinct vectors in A is linearly independent.

Exercises.

3.2.1 ([Ber14, Ex. 3.2.1]). Let $V = \mathcal{F}(\mathbb{R})$ be the vector space of all real-valued functions of a real variable (see Example 1.2.5). Define

$$f\colon \mathbb{R} \to \mathbb{R}, \quad f(t) = \sin\left(t + \frac{\pi}{6}\right).$$

Prove that the functions f, sin, and cos are linearly dependent in V.

3.2.2. Show that in $\mathbb{R}^3$, the vectors e_1, e_2, e_3, and $(-2, 5, 4)$ are linearly dependent. (See Example 1.4.8.)

3.2.3. Show that the vectors $(1, 1, 1)$, $(1, 2, -1)$, and $(1, -1, 2)$ in $\mathbb{R}^3$ are linearly independent.

3.2.4. Suppose $T\colon V \to W$ is a linear map and $x_1, \dots, x_n$ are vectors in V.

(a) If $x_1, \dots, x_n$ are linearly dependent, then the vectors $Tx_1, \dots, Tx_n$ are linearly dependent in W.

(b) If the vectors $Tx_1, \dots, Tx_n$ are linearly independent, then $x_1, \dots, x_n$ are linearly independent.

3.2.5 ([Ber14, Ex. 3.2.7]). If $T\colon V \to W$ is an injective linear map and $x_1, \dots, x_n$ are vectors in V such that $Tx_1, \dots, Tx_n$ are linearly dependent in W, then $x_1, \dots, x_n$ are linearly dependent in V.

3.2.6. If E is a subset of $\mathbb{R}$, then we define a function $\chi_E\colon \mathbb{R} \to \mathbb{R}$ (called the *characteristic function* of E) by

$$\chi_E(x) = \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \in \mathbb{R} \setminus E. \end{cases}$$

Note that χ_E is a vector in the vector space $\mathcal{F}(\mathbb{R})$. For example, $\chi_\varnothing$ is the zero function (the zero vector in $\mathcal{F}(\mathbb{R})$) and $\chi_\mathbb{R}$ is the constant function with value 1.

Let A and B be subsets of $\mathbb{R}$ and consider the list

$$\ell : \ \chi_A, \ \chi_B, \ \chi_{A\cap B},$$

of vectors in $\mathcal{F}(\mathbb{R})$.

(a) Show that if the condition

$$A \cap B \neq \varnothing \text{ and } A \not\subseteq B \text{ and } B \not\subseteq A \tag{$*$}$$

is satisfied, then the list ℓ is linearly independent.

(b) Show that if the condition $(*)$ is not satisfied, then the list ℓ is linearly dependent.

3.2.7. For $y \in \mathbb{R}$, define $\delta_y \in \mathcal{F}(\mathbb{R})$ by

$$\delta_y(x) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

Show that $\{\delta_y \mid y \in \mathbb{R}\}$ is a linearly independent subset of $\mathcal{F}(\mathbb{R})$. *Note*: Since the subset is infinite, you will need to use Definition 3.2.14.

3.2.8. Suppose A is a linearly independent set of vectors in a vector space V. Prove that every nonempty subset of A is also linearly independent.

3.2.9 ([Ber14, Ex. 3.3.3]). If $x_1, \dots, x_n$ are linearly independent, then the following implication is true:

$$a_1x_1 + \cdots + a_nx_n = b_1x_1 + \cdots + b_nx_n \implies a_i = b_i \text{ for all } i = 1, \dots, n.$$

Conversely, if the above implication is true, then the vectors $x_1, \dots, x_n$ are linearly independent.

3.2.10 ([Ber14, Ex. 3.3.4]). Prove that if the vectors x_1, x_2, x_3 are linearly independent, then so are the vectors $x_1, x_1 + x_2, x_1 + x_2 + x_3$.

3.2.11 ([Ber14, Ex. 3.3.5]). Let a, b, c be distinct real numbers. Prove that the vectors

$$(1, 1, 1), \ (a, b, c), \ (a^2, b^2, c^2)$$

in $\mathbb{R}^3$ are linearly independent. Can you come up with a generalization of this result?

3.2.12 ([Ber14, Ex. 3.3.7]). (a) Let V be a real or complex vector space. Prove the following statement: If x, y, z are linearly independent vectors in V, then so are $x + y$, $y + z$, $z + x$.

(b) Does this statement hold if V is instead a vector space over the field $\mathbb{F}_2$ of two elements (see Definition 1.1.2)? *Hint*: $2x = \mathbf{0}$ for every vector x.

3.2.13 ([Ber14, Ex. 3.3.11]). Prove that the functions $\sin t$, $\sin 2t$, and $\sin 3t$ are linearly independent.

3.2.14 ([Ber14, Ex. 3.3.12]). Let $V = \mathbb{C}^2$, $u = (1, i)$, and $v = (i, -1)$.

(a) Show that u, v are linearly independent in V when V is considered as a *real* vector space.

(b) Show that u, v are linearly dependent in V when V is considered as a *complex* vector space.

3.3 Finitely generated vector spaces

Definition 3.3.1 (Finitely generated vector space)**.** A vector space V over a field F is *finitely generated* if there is a finite subset $\{v_1, \dots, v_n\} \subseteq V$ such that $\operatorname{Span}\{v_1, \dots, v_n\} = V$. In other words, V is finitely generated if it can be spanned by finitely many vectors. In the case where V can be considered as a vector space over multiple fields (for example, $\mathbb{C}$ is a vector space over both $\mathbb{R}$ and $\mathbb{C}$), we say V is *finitely generated over* F if there is a finite subset $\{v_1, v_2, \dots, v_n\} \subseteq V$ such that $\operatorname{Span}_F\{v_1, \dots, v_n\} = V$.

Remark 3.3.2. Recall (Definition 3.1.3) that if $\operatorname{Span}\{v_1, \dots, v_n\} = V$, then we say that the set $\{v_1, \dots, v_n\}$ generates V. This explains the term 'finitely generated'. Later, once we've defined dimension, we'll often use the term *finite dimensional* instead of *finitely generated*.

Examples 3.3.3. (a) The zero space $\{\mathbf{0}\}$ is finitely generated since $\operatorname{Span}\{\mathbf{0}\} = \{\mathbf{0}\}$.

(b) F^n is finitely generated over F since $\text{Span}_F\{e_1, e_2, \ldots, e_n\} = F$.

(c) $\mathcal{P}_n(F) = \text{Span}_F\{1, t, t^2, \ldots, t^n\}$ and so $\mathcal{P}_n(F)$ is finitely generated over F. However, if F is infinite, then $\mathcal{P}(F)$ is *not* finitely generated. See Exercise 3.3.3.

(d) $\mathbb{C}$ is finitely generated over $\mathbb{C}$ since $\mathbb{C} = \text{Span}_{\mathbb{C}}\{1\}$.

(e) $\mathbb{C}$ is finitely generated over $\mathbb{R}$ since $\mathbb{C} = \text{Span}_{\mathbb{R}}\{1, i\}$.

(f) $\mathbb{R}$ is finitely generated over $\mathbb{R}$ since $\mathbb{R} = \text{Span}_{\mathbb{R}}\{1\}$.

(g) $\mathbb{R}$ is *not* finitely generated over $\mathbb{Q}$ (but this is not so easy to see).

(h) $\mathcal{F}(\mathbb{R})$ is *not* finitely generated over $\mathbb{R}$.

Example 3.3.4. The vector space $V = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\}$ is finitely generated over $\mathbb{R}$. Indeed, we can show that $V = \text{Span}_{\mathbb{R}}\{\sin, \cos\}$. To do this, we need to prove that any $f \in V$ can be written as a linear combination of sin and cos. Suppose $f \in V$ and let

$$g(x) = f(x)\sin x + f'(x)\cos x.$$

Then

$$g'(x) = f'(x)\sin x + f(x)\cos x + f''(x)\cos x - f'(x)\sin x = (f''(x) + f(x))\cos x = 0.$$

Therefore, $g(x)$ is constant. That is $g(x) = a$ (for all $x \in \mathbb{R}$) for some $a \in \mathbb{R}$. Similarly, if

$$h(x) = f(x)\cos x - f'(x)\sin x,$$

then one can show that $h'(x) = 0$ (Exercise 3.3.1) and so $h(x) = b$ (for all $x \in \mathbb{R}$) for some $b \in \mathbb{R}$ (that is, h is a constant function). Therefore we have

$$\underbrace{\begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix}}_{A} \begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Since $\det A = -\sin^2 x - \cos^2 x = -1$, for all $x \in \mathbb{R}$, the matrix A is invertible with inverse

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} -\sin x & -\cos x \\ -\cos x & \sin x \end{bmatrix} = \begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix}.$$

Multiplying both sides of our matrix equation on the left by A^{-1} gives

$$\begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} = \begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\sin x + b\cos x \\ a\cos x - b\sin x \end{bmatrix}.$$

Thus $f(x) = a\sin x + b\cos x$ for all $x \in \mathbb{R}$.

Example 3.3.5. If $\{v_1, \ldots, v_n\}$ generates V, and $v_{n+1} \in V$, then $\{v_1, \ldots, v_n, v_{n+1}\}$ is dependent. This is because v_{n+1} must be a linear combination of $v_1, \ldots, v_n$ and so $\{v_1, \ldots, v_{n+1}\}$ is dependent by Theorem 3.2.9.

In contrapositive form, we have that if $x_1, \ldots, x_n$ are independent, then $x_1, \ldots, x_{n-1}$ cannot generate V.

Theorem 3.3.6. *Suppose V is a vector space over a field F and $v_1, \dots, v_n \in V$. Define a linear map $T\colon F^n \to V$ by*

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

Then T is surjective if and only if $v_1, \dots, v_n$ generate V.

Proof. By definition, T is surjective if and only if every vector of V can be written as $a_1 v_1 + \cdots + a_n v_n$. This is true if and only if $\operatorname{Span}\{v_1, \dots, v_n\} = V$. □

Theorem 3.3.7. *If $T\colon V \to W$ is a surjective linear map and V is finitely generated, then W is finitely generated.*

Proof. Since V is finitely generated, there is a finite subset $A = \{v_1, \dots, v_n\} \subseteq V$ such $\{v_1, \dots, v_n\}$ generates V. Then, since T is surjective, $T(A) = \{Tv_1, \dots, Tv_n\}$ generates W by Corollary 3.1.5. Hence W is finitely generated. □

Theorem 3.3.8. *A vector space V over F is finitely generated if and only if there exists a positive integer n and a surjective linear map $T\colon F^n \to V$.*

Proof. If V is finitely generated, then $\operatorname{Span}\{v_1, \dots, v_n\} = V$ for some finite subset $\{v_1, \dots, v_n\} \subseteq V$. Then, by Theorem 3.3.6, there is a surjective linear map $T\colon F^n \to V$.

Conversely, if there is a positive integer n and a linear surjective map $T\colon F^n \to V$, then, by Theorem 3.3.7, V is finitely generated since F^n is. □

Exercises.

3.3.1. In the notation of Exercise 3.3.4, show that $h'(x) = 0$.

3.3.2. Do the vectors $(2, 1, 1)$, $(3, -1, 1)$, $(10, 5, 5)$, and $(6, -2, 2)$ generate $\mathbb{R}^3$? Justify your answer.

3.3.3. Prove that $\mathcal{P}(\mathbb{R})$ is not finitely generated over $\mathbb{R}$. *Hint*: Suppose $p_1, \dots, p_n \in \mathcal{P}(\mathbb{R})$. Find $p \in \mathcal{P}(\mathbb{R})$ such that $p \notin \operatorname{Span}\{p_1, \dots, p_n\}$.

3.3.4 ([Ber14, Ex. 3.4.4]). Let $V = C^\infty(\mathbb{R})$ be the vector space of all functions $\mathbb{R} \to \mathbb{R}$ having derivatives of all orders (see Example 1.2.6) and let

$$U = \{f \in V \mid f'' = f\},$$

where f'' is the second derivative of f. Prove that U is the subspace of V generated by the two functions $t \mapsto e^t$ and $t \mapsto e^{-t}$. *Hint*: If $f'' = f$, consider $f = \frac{1}{2}(f + f') + \frac{1}{2}(f - f')$.

3.4 Basis and dimension

Definition 3.4.1 (Basis)**.** A finite list of vectors $v_1, \ldots, v_n$ in a vector space V is called a *basis* of V if it is both independent and generating. In other words, every vector $v \in V$ can be written as a linear combination

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

(since $\operatorname{Span}\{v_1, \ldots, v_n\} = V$) and the coefficients $a_1, \ldots, a_n$ are *unique* by Corollary 3.2.8. These coefficients are called the *coordinates* of v with respect to the basis $v_1, \ldots, v_n$.

We also speak of the set $\{v_1, \ldots, v_n\}$ being a basis. If we wish to emphasize the field, we say $v_1, \ldots, v_n$ is a basis of V *over* F. The plural of 'basis' is 'bases' (pronounced bay-sees).

Examples 3.4.2. (a) For any field F, the set $\{e_1, \ldots, e_n\}$ is a basis of F^n, called the *canonical basis*, the *standard basis*, or the *natural basis* of F^n.

(b) If $v_1, \ldots, v_n$ are independent vectors in a vector space V, then $\{v_1, \ldots, v_n\}$ is a basis of $\operatorname{Span}\{v_1, \ldots, v_n\}$.

(c) $\{1\}$ is a basis for $\mathbb{C}$ over $\mathbb{C}$.

(d) $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$.

(e) $\{1, t, \ldots, t^n\}$ is a basis for $\mathcal{P}_n(\mathbb{R})$ over $\mathbb{R}$.

Example 3.4.3. Suppose $v_1, \ldots, v_n$ is a basis of V. If v_{n+1} is any vector in V, then $v_1, \ldots, v_{n+1}$ is dependent (since v_{n+1} is a linear combination of the other vectors and so the set is dependent by Theorem 3.2.9) and so is not a basis. Additionally, the list $v_1, \ldots, v_{n-1}$ is not generating (since if it were, v_n would be a linear combination of the vectors in this shorter list and so $v_1, \ldots, v_n$ would not be independent).

Theorem 3.4.4. *Suppose V is a vector space over a field F and $v_1, \ldots, v_n$ is a finite list of vectors in V. Define a linear map $T\colon F^n \to V$ by*

$$T(a_1, \ldots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

Then the following statements are equivalent:

(a) $v_1, \ldots, v_n$ is a basis of V,

(b) T is bijective.

Proof. We have that $v_1, \ldots, v_n$ is a basis if and only if it is both independent and generating. This is true if and only if T is injective and surjective by Theorem 3.3.6 and Corollary 3.2.8. □

Corollary 3.4.5. *A vector space V over a field F has a (finite) basis if and only if $V \cong F^n$ for some positive integer n.*

Proof. If $V \cong F^n$, then there is a bijective linear map $T\colon F^n \to V$. Let $e_1, \ldots, e_n$ be the canonical basis of F^n. Then $Te_1, \ldots, Te_n$ is independent (by Theorem 3.2.13) and generating (by Corollary 3.1.5) and hence is a basis of V.

Conversely, if V has a basis $v_1, \ldots, v_n$, then $V \cong F^n$ by Theorem 3.4.4. □

Theorem 3.4.6. *Every nonzero finitely generated vector space has a basis.*

Proof. Let $V \neq \{\mathbf{0}\}$ be a finitely generated vector space. We know that there is some finite set of vectors that spans V. Among all such sets, choose one of minimal size, say $\{u_1, \ldots, u_k\}$. So $\{u_1, \ldots, u_k\}$ generates V. We will show that it is also linearly independent (by contradiction). Suppose that this set is dependent. Then, by Theorem 3.2.9, there exists a j, $1 \leq j \leq k$, such that $u_j \in \text{Span}\{u_1, u_2, \ldots, \hat{u}_j, \ldots, u_k\}$ (where the notation $\hat{u}_j$ means we omit the vector u_j). Then $V = \text{Span}\{u_1, \ldots, \hat{u}_j, \ldots, u_k\}$, which contradicts the minimality of k. Therefore, $\{u_1, \ldots, u_k\}$ independent and hence is a basis (since it is also generating). □

Remarks 3.4.7. (a) In fact, *every* vector space (not just finitely generated ones) has a basis. The proof of this fact uses Zorn's Lemma and is beyond the scope of this course.

(b) The above proof shows that we can get a basis by taking a sublist of some generating set. We just keep removing vectors until the vectors are independent.

(c) A vector space can have more than one basis.

Even though a vector space can have more than one basis, any two bases have the same number of vectors, as we will see (Theorem 3.4.9).

Theorem 3.4.8. *Suppose V is a vector space and*

- *R is a linearly independent subset of V with $|R| = m$ and*
- *S is a generating set for V with $|S| = n$.*

Then $m \leq n$ and there exists a subset $S' \subseteq S$ with $|S'| = n - m$ such that

$$V = \text{Span}(R \cup S').$$

Proof. We prove the result by induction on m. In the base case $m = 0$, we have $R = \varnothing$. Thus, taking $S' = S$ gives the desired statement.

For the induction step, we now assume that the theorem holds for some integer $m \geq 0$ (and arbitrary n). We wish to show that it holds for $m + 1$.

Let $R = \{v_1, \ldots, v_{m+1}\}$ be a linearly independent subset of V. Then the subset $\{v_1, \ldots, v_m\}$ is also linearly independent (Exercise 3.2.8). Therefore, by the induction hypothesis, we have $m \leq n$ and there is a subset $S' = \{u_1, \ldots, u_{n-m}\}$ of S such that

$$V = \text{Span}\left(\{v_1, \ldots, v_m\} \cup \{u_1, \ldots, u_{n-m}\}\right).$$

We cannot have $V = \text{Span}(\{v_1, \ldots, v_m\})$, since that would imply that R is linearly dependent (since we could write v_{m+1} as a linear combination of $v_1, \ldots, v_m$). Thus we have $n - m \geq 1$, which gives us $m + 1 \leq n$, proving part of our induction step.

Since $v_{m+1} \in V$, this means that there exist scalars $a_1, \ldots, a_m \in F$ and $b_1, \ldots, b_{n-m} \in F$ such that

$$v_{m+1} = a_1 v_1 + \cdots + a_m v_m + b_1 v_1 + \cdots + b_{n-m} u_{n-m}.$$

We claim that the $b_1, \ldots, b_{n-m}$ are not all zero. Indeed, if $b_1 = b_2 = \cdots = b_{n-m} = 0$, then we have

$$v_{m+1} = a_1 v_1 + \cdots + a_m v_m \in \text{Span}(\{v_1, \ldots, v_m\}).$$

Thus, by Theorem 3.2.9, the set $\{v_1, \ldots, v_{m+1}\}$ in linearly independent, contradicting our assumption.

By the above, we can choose $1 \leq i \leq n-m$ such that $b_i \neq 0$. Then we have

$$u_i = (-b_i^{-1}a_1)v_1 + \cdots + (-b_i^{-1}a_m)v_m + b_i^{-1}v_{m+1} + (-b_i^{-1}b_1)u_1 + \cdots + (-b_i^{-1}b_{i-1})u_{i-1} + (-b_i^{-1}b_{i+1})u_{i+1} + \cdots + (-b_i^{-1}b_{n-m})u_{n-m}. \quad (3.1)$$

Let

$$S' = \{u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{n-m}\}.$$

Then, by (3.1), we have $u_i \in \operatorname{Span}(R \cup S')$ and so

$$\begin{aligned} V &= \operatorname{Span}\left(\{v_1, \ldots, v_m\} \cup \{u_1, \ldots, u_{n-m}\}\right) \\ &= \operatorname{Span}\left(\{v_1, \ldots, v_m, v_{m+1}\} \cup \{u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{n-m}\}\right) \\ &= \operatorname{Span}(R \cup S'). \end{aligned}$$

Thus the result holds for $|R| = m+1$, completing the proof of the induction step. □

Theorem 3.4.8 tells us that the size of any spanning set is greater than the size of any linearly independent set (of the same vector space).

Theorem 3.4.9. *If V has a basis with n elements, then* every *basis of V has n elements.*

Proof. Suppose $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_m\}$ are both bases of V. Since B spans V and C is linearly independent, we have $m \leq n$ (by Theorem 3.4.8). Since C spans V and B is linearly independent, we also have $n \leq m$. Hence $m = n$. □

Definition 3.4.10 (Dimension)**.** Suppose V is a finitely generated vector space. If $V \neq \{\mathbf{0}\}$, then the number of vectors in any basis of V is called the *dimension* of V and is written $\dim V$. If $V = \{\mathbf{0}\}$, we say $\dim V = 0$. If we wish to emphasize the field F, we write $\dim_F V$ for the dimension of V over F. Finitely generated vector spaces are also called *finite dimensional*. If $\dim V = n$, we say V is *n-dimensional*. If V is not finite dimensional, we say it is *infinite dimensional*.

Examples 3.4.11. (a) $\dim_F F^n = n$. (See Exercise 3.4.2.)

(b) $\dim_{\mathbb{C}} \mathbb{C} = 1$.

(c) $\dim_{\mathbb{R}} \mathbb{C} = 2$.

(d) $\dim_{\mathbb{R}} \mathcal{P}_n(\mathbb{R}) = n+1$.

(e) $\dim_{\mathbb{R}} \mathcal{P}_n(\mathbb{C}) = 2(n+1)$.

Lemma 3.4.12. *Suppose V is a vector space. If $\{v_1, \ldots, v_n\}$ spans V, then $\dim V \leq n$. If $\{w_1, \ldots, w_m\}$ is linearly independent, then $m \leq \dim V$.*

Proof. This follows from Theorem 3.4.8. □

Lemma 3.4.13. *Suppose $\{v_1, \ldots, v_n\}$ is independent and v is a vector. Then $\{v, v_1, \ldots, v_n\}$ is independent if and only if $v \notin \operatorname{Span}\{v_1, \ldots, v_n\}$.*

Proof. We prove the contrapositive: That $\{v, v_1, \ldots, v_n\}$ is dependent if and only if $v \in \text{Span}\{v_1, \ldots, v_n\}$. We already know by Theorem 3.2.9 that if $v \in \text{Span}\{v_1, \ldots, v_n\}$, then $\{v, v_1, \ldots, v_n\}$ is dependent. Now suppose $\{v, v_1, \ldots, v_n\}$ is dependent. Then there are scalars $a, a_1, \ldots, a_n$, not all zero, such that

$$av + a_1v_1 + \cdots + a_nv_n = 0.$$

If $a = 0$, then $a_1v_1 + \cdots + a_nv_n = 0$. Since $\{v_1, \ldots, v_n\}$ is independent, we have $a_1 = \cdots = a_n = 0$. This contradicts the fact that $a, a_1, \ldots, a_n$ are not all zero. So $a \neq 0$. Then

$$v = -a^{-1}a_1v_1 - \cdots - a^{-1}a_nv_n \in \text{Span}\{v_1, \ldots, v_n\}. \qquad \square$$

Theorem 3.4.14. *Suppose V is an n-dimensional vector space. Then the following statements are equivalent.*

(a) $\{v_1, \ldots, v_n\}$ *spans* V.

(b) $\{v_1, \ldots, v_n\}$ *is linearly independent.*

(c) $\{v_1, \ldots, v_n\}$ *is a basis of* V.

Proof. (a) $\Rightarrow$ (b): Suppose $\{v_1, \ldots, v_n\}$ spans V but is dependent. Then there exists an i such that $V = \text{Span}\{v_1, \ldots, \hat{v}_i, \ldots, v_n\}$. But then V has a spanning set with $n-1$ elements. But this contradicts Lemma 3.4.12. Thus, (b) holds.

(b) $\Rightarrow$ (a): Suppose $\{v_1, \ldots, v_n\}$ is independent but does not span V. Then we can find a vector $v \in V$ such that $v \notin \text{Span}\{v_1, \ldots, v_n\}$. Then, by Lemma 3.4.13, $\{v, v_1, \ldots, v_n\}$ is an independent set with $n+1$ elements. But this contradicts Lemma 3.4.12. Thus, (a) holds.

It is now clear that (a) and (b) are equivalent to (c). $\square$

Remark 3.4.15. The point of the above theorem is that if we know the dimension of a vector space ahead of time, to show that some set is a basis we only need to check *one* of the two defining properties of a basis (independence or spanning).

Example 3.4.16. We know $\dim_{\mathbb{R}} \mathbb{C} = 2$. Since the set $\{1, 1+i\}$ is linearly independent (Exercise 3.4.1), it is a basis.

Example 3.4.17. We showed in Example 3.3.4 that

$$W = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\} = \text{Span}\{\sin, \cos\}.$$

We can also check that sin and cos are linearly independent (Exercise 3.4.3). Thus $\dim W = 2$.

Define

$$f(x) = \sin x + \cos x$$
$$g(x) = \sin x - \cos x.$$

Then f and g are linearly independent (Exercise 3.4.3). Therefore $\{f, g\}$ is a basis for W.

Example 3.4.18. Let

$$U = \{(x, y, z) \in \mathbb{C}^3 \mid x + y + z = 0\} = \text{Span}\{(-1, 1, 0), (-1, 0, 1)\}.$$

(You learned how to solve homogeneous systems like this in MAT 1341.) Moreover,

$$\{(-1, 1, 0), (-1, 0, 1)\}$$

is linearly independent (it is a two vector set, and neither vector is a multiple of the other). Thus $\{(-1, 1, 0), (-1, 0, 1)\}$ is a basis of U and so $\dim U = 2$.

Since $(1, 1, -2)$, $(0, 1, -1)$ both belong to U and are linearly independent,

$$\{(1, 1, -2), (0, 1, -2)\}$$

is another basis for U.

Lemma 3.4.19. *Every linearly independent set of vectors in a finitely-generated vector space V can be extended to a basis of V.*

Proof. Suppose V is a finitely-generated vector space, and let $n = \dim V$. Let R be a linearly independent set of vectors in V. Choose a basis S of V. By Theorem 3.4.8, there is a subset S' of S with $|S'| = n - m$ and $V = \text{Span}(R \cup S')$. By Lemma 3.4.12, $|R \cup S'| \geq n$. On the other hand, $n = |R| + |S'| \geq |R \cup S'|$. Thus $|R \cup S'| = n$. Thus, by Theorem 3.4.14, $R \cup S'$ is a basis of V. □

Theorem 3.4.20. *Suppose V and W are finite-dimensional vector spaces. Then $V \cong W$ if and only if $\dim V = \dim W$.*

Proof. Suppose $\dim V = n$. Then V has a basis $\{v_1, \ldots, v_n\}$ with n vectors. If $V \cong W$, then there exists an isomorphism $T \colon V \to W$. Then $Tv_1, \ldots, Tv_n$ generates W (Corollary 3.1.5) and is independent (Theorem 3.2.13). Then $\{Tv_1, \ldots, Tv_n\}$ is a basis of W and so $\dim W = n = \dim V$.

Suppose suppose $\dim W = \dim V = n$. Then W has a basis $\{w_1, \ldots, w_n\}$ with n elements. Then, by Theorem 3.4.4, $V \cong F^n$ and $W \cong F^n$. Hence $V \cong W$ by Theorem 2.4.12. □

Theorem 3.4.21. *Suppose W is a subspace of a finite-dimensional vector space V. Then*

(a) W is finite dimensional and $\dim W \leq \dim V$, and

(b) $\dim W = \dim V$ if and only if $W = V$.

Proof. Let $n = \dim V$. If $W = \{\mathbf{0}\}$, then the theorem is trivially true. Therefore, assume $W \neq \{\mathbf{0}\}$. Hence $V \neq \{\mathbf{0}\}$ and $n \geq 1$. Choose $w_1 \in W$, $w_1 \neq 0$. Then $\{w_1\}$ is independent. If $\text{Span}\{w_1\} = W$, then W is finitely-generated. Otherwise, choose $w_2 \in W$ such that $w_2 \notin \text{Span}\{w_1\}$. Then $\{w_1, w_2\}$ is independent. We continue in the manner, extending the list. By Lemma 3.4.12, this process must stop and so we obtain a list $\{w_1, \ldots, w_k\}$ which is both independent and spans W, with $k \leq n$. Therefore, $\dim W \leq \dim V$. If $\dim W = \dim V$ (i.e. $k = n$), then $\{w_1, \ldots, w_k\}$ must span V since it is independent (Theorem 3.4.14) and so $W = \text{Span}\{w_1, \ldots, w_k\} = V$. It is clear that if $W = V$, then $\dim W = \dim V$. □

Example 3.4.22. Take

$$V = \mathcal{P}_2(\mathbb{R}) = \{a_0 + a_1 t + a_2 t^2 \mid a_i \in \mathbb{R}\}.$$

The set $\{1 - t\}$ is independent in V (since $1 - t$ is not the zero polynomial, e.g. $1 - t \neq 0$ at $t = 0$). Since $\operatorname{Span}\{1 - t\} \neq \mathcal{P}_2(\mathbb{R})$, we continue. Choose a polynomial, for instance, $t \in \mathcal{P}_2(\mathbb{R})$, but $t \notin \operatorname{Span}\{1 - t\}$. Then $\{1 - t, t\}$ is linearly independent. Is $\operatorname{Span}\{1 - t, t\} = \mathcal{P}_2(\mathbb{R})$? No (for instance $t^2 \notin \operatorname{Span}\{1 - t, t\}$). Then the set $\{1 - t, t, t^2\}$ is independent in $\mathcal{P}_2(\mathbb{R})$. Since $\dim \mathcal{P}_2(\mathbb{R}) = 3$, we know that $\{1 - t, t, t^2\}$ is a basis.

Exercises.

3.4.1. Prove that the complex numbers 1 and $1 + i$ are linearly independent over $\mathbb{R}$.

3.4.2. Suppose F is a field.

(a) Prove that $\dim_F F^n = n$.

(b) Prove that $F^n \cong F^m$ if and only if $m = n$.

3.4.3. This exercise concerns Example 3.4.17.

(a) Show that sin and cos are linearly independent (over $\mathbb{R}$) elements of the set W. *Hint*: Write down an arbitrary linear combination and suppose it is equal to the zero function. Then evaluate the functions at carefully selected points that allow you to conclude the coefficients in your linear combination must be zero.

(b) Show that f and g are linearly independent.

3.4.4 ([Ber14, Ex. 3.5.5]). Let α be a real number. Prove that the vectors

$$u = (\cos\alpha, \sin\alpha) \quad \text{and} \quad v = (-\sin\alpha, \cos\alpha)$$

are a basis of $\mathbb{R}^2$.

3.4.5 ([Ber14, Ex. 3.5.6]). True or false (explain): If x_1, x_2, x_3 is a basis of V, then so is $x_1, x_1 + x_2, x_1 + x_2 + x_3$.

3.4.6 ([Ber14, Ex. 3.5.7]). (a) In $\mathbb{R}^2$, find the coordinates of the vector $(2, 3)$ with respect to the basis $\left(\frac{1}{2}\sqrt{3}, \frac{1}{2}\right)$, $\left(-\frac{1}{2}, \frac{1}{2}\sqrt{3}\right)$.

(b) In $\mathbb{R}^n$, find the coordinates of the vector $(a_1, \ldots, a_n)$ with respect to the canonical basis $e_1, \ldots, e_n$ (see Example 1.4.8).

3.4.7 ([Ber14, Ex. 3.5.8]). If $x_1 = (1, 2, 0)$, $x_2 = (2, 1, 0)$, and $x_3 = (a, b, 1)$, where a and b are any real numbers, prove that x_1, x_2, x_3 is a basis of $\mathbb{R}^3$.

3.4.8. Suppose $T\colon V \to W$ is a linear map, with V finite dimensional. Prove that $\dim T(U) \leq \dim U$ for any subspace U of V.

3.4.9. Find a basis of the subspace

$$U = \{(x, y, z) \in \mathbb{R}^3 \mid x + 2y + 3z = 0\}$$

of $\mathbb{R}^3$. What is the dimension of U? *Note:* This is really a MAT 1341 question. It's here to refresh your memory.

3.4.10. Suppose $v_1, \ldots, v_n$ (where $n \geq 2$) is a basis of a vector space V. Choose $r \in \{1, \ldots, n-1\}$ and define

$$M = \operatorname{Span}\{v_1, \ldots, v_r\},$$
$$N = \operatorname{Span}\{v_{r+1}, \ldots, v_n\}.$$

Show that $V = M \oplus N$.

3.4.11. Suppose that U is a subspace of a finitely generated vector space V. Show that U has a complement. That is, show that there exists some subspace W of V such that $U \oplus W = V$.

3.4.12 ([Ber14, Ex. 3.5.13]). Prove that if $x_1, \ldots, x_n$ is a basis of V and $a_1, \ldots, a_n$ are nonzero scalars, then $a_1x_1, \ldots, a_nx_n$ is also a basis of V.

3.4.13 ([Ber14, Ex. 3.5.14]). Prove that if x_1, x_2, x_3 is a basis of V and a_1, a_2, a_3 are nonzero scalars, then the list

$$a_1x_1, \quad a_1x_1 + a_2x_2, \quad a_1x_1 + a_2x_2 + a_3x_3$$

is also a basis of V. *Hint*: Combine Exercises 3.4.5 and 3.4.12.

3.4.14. Suppose that F is a finite field with q elements. Let V be an n-dimensional vector space over F. How many elements does V have? Remember to justify your answer.

3.4.15 ([Ber14, Ex. 3.5.20]). Let V be a finite-dimensional complex vector space. Prove that $\dim_{\mathbb{R}} V = 2 \cdot \dim_{\mathbb{C}} V$. *Hint*: As complex vector spaces, $V \cong \mathbb{C}^n$ for some n.

3.5 The Dimension Theorem

Theorem 3.5.1 (Dimension Theorem)**.** *If V is a finite-dimensional vector space and $T\colon V \to W$ is a linear map, then*

$$\dim V = \dim T(V) + \dim(\operatorname{Ker} T).$$

Proof. By Corollary 2.2.4, $\operatorname{Ker} T$ is a subspace of V, and hence is finite-dimensional by Theorem 3.4.21. Choose a basis $B' = \{v_1, \ldots, v_k\}$ of $\operatorname{Ker} T$. By Lemma 3.4.19, we can extend this to a basis $B = \{v_1, \ldots, v_k, v_{k+1}, v_n\}$ of V.

Suppose $w \in T(V)$. Then $w = T(v)$ for some $v \in V$. Since B is a basis for V, we can write

$$v = \sum_{i=1}^{n} c_i v_i, \quad c_1, \ldots, c_n \in F.$$

Then

$$w = T(v) = T\left(\sum_{i=1}^{n} c_i v_i\right) = \sum_{i=1}^{n} c_i T(v_i) = \sum_{i=k+1}^{n} c_i T(v_i).$$

Therefore,

$$T(V) = \operatorname{Span}\{T(v_{k+1}), \ldots, T(v_n)\}.$$

We claim that the $\{T(v_{k+1}), \ldots, T(v_n)\}$ is linearly independent and hence forms a basis of $T(V)$. Indeed, suppose

$$c_{k+1}T(v_{k_1}) + \cdots + c_n T(v_n) = 0$$

for some $c_{k+1}, \ldots, c_n \in F$. Then

$$T(c_{k+1}v_{k+1} + \cdots + c_n v_n) = c_{k+1}T(v_{k_1}) + \cdots + c_n T(v_n) = 0,$$

and so $c_{k+1}v_{k+1} + \cdots + c_n v_n \in \operatorname{Ker} T$. Since B' is a basis for $\operatorname{Ker} T$, we have

$$c_{k+1}v_{k+1} + \cdots + c_n v_n = c_1 v_1 + \cdots + c_k v_k$$

for some $c_1, \ldots, c_k \in F$. But then

$$-c_1 v_1 - \cdots - c_k v_k + c_{k+1}v_{k+1} + \cdots + c_n v_n = 0.$$

Since B is a basis, it is linearly independent. Hence $c_1 = c_2 = \cdots = c_n = 0$.

Finally, counting the number of basis elements, we have

$$\dim V = n, \quad \dim(\operatorname{Ker} T) = k, \quad \dim T(V) = n - k.$$

Hence $\dim V = \dim T(V) + \dim(\operatorname{Ker} T)$, as desired. □

Definition 3.5.2 (Rank and Nullity)**.** If $T\colon V \to W$ is a linear map, we define the *rank* of T to be

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

and the *nullity* of T to be

$$\operatorname{null} T = \dim(\operatorname{Ker} T).$$

We can now rephrase Theorem 3.5.1 as

$$\operatorname{rank} T + \operatorname{null} T = \dim V, \tag{3.2}$$

where V is the domain of the linear map T. The equation (3.2) is sometimes call the *Rank-Nullity Theorem*.

Corollary 3.5.3. *If $T\colon V \to W$ is linear, and $\dim V = \dim W < \infty$, then*

$$T \text{ is an isomorphism} \iff T \text{ is injective} \iff T \text{ is surjective.}$$

Proof. Obviously, if T is an isomorphism, it is injective and surjective (by definition). Suppose T is injective. Then $\operatorname{Ker} T = \{\mathbf{0}\}$ and so

$$\dim T(V) = \dim V = \dim W$$

by the Dimension Theorem. So $T(V) = W$ (since $T(V)$ is a subspace of W, we can apply Theorem 3.4.21). Finally, suppose T is surjective. Then $T(V) = W$ and so $\operatorname{rank} T = \dim W = \dim V$. By the Dimension Theorem, $\operatorname{null} T = 0$ and so T is injective, hence an isomorphism. □

Theorem 3.5.4. *If $V_1, V_2, \ldots, V_n$ are finite-dimensional vector spaces, then so is $V = V_1 \times \cdots \times V_n$ and*

$$\dim V = \dim V_1 + \dim V_2 + \cdots + \dim V_n.$$

Proof. For each $i = 1, \ldots, n$, let B_i be a basis of V_i, and let

$$B_i' = \{(0, \ldots, 0, v, 0, \ldots, 0) \mid v \in B_i\} \subseteq V,$$

where the v appears in the i-th position. We leave it as an exercise (Exercise 3.5.1) to show that

$$B := B_1' \cup B_2' \cup \cdots \cup B_n'$$

is a basis of V. Since each B_i has $\dim V_i$ elements, B has $\sum_{i=1}^{n} \dim V_i$ elements, and the theorem follows. □

Theorem 3.5.5. *A system of m homogeneous linear equations in n unknowns, where $n > m$, always has a nontrivial solution.*

Proof. As you saw in MAT 1341, such a system is equivalent to a matrix equation

$$Ax = \mathbf{0},$$

where A is the coefficient matrix, and hence is $m \times n$. The map

$$T\colon F^n \to F^m, \quad T(x) = Ax,$$

is linear (since it is multiplication by a matrix). The set of solutions is precisely the kernel of T. By Theorem 3.5.1, we have

$$\dim T(F^n) + \dim \operatorname{Ker} T = \dim F^n = n \implies \dim \operatorname{Ker} T = n - \dim T(F^n) \geq n - m > 0.$$

Here we used that $T(F^n) \subseteq F^m$ and so $\dim T(F^n) \leq \dim F^m = m$. So $\dim \operatorname{Ker} T > 0$, which means that $\operatorname{Ker} T$ is not the zero vector space and hence there are nonzero elements in the kernel of T (which correspond to nontrivial solutions to the homogeneous system). □

Theorem 3.5.6. *Suppose $T\colon V \to W$ is a linear map between finite-dimensional vector spaces.*

(a) If T is injective, then $\dim V \leq \dim W$.

(b) If T is surjective, then $\dim V \geq \dim W$.

(c) If T is bijective, then $\dim V = \dim W$.

Proof. (a) If T is injective, then $\operatorname{Ker} T = \{\mathbf{0}\}$. Thus

$$\dim V = \dim T(V) + \dim \operatorname{Ker} T = \dim T(V) + 0 \leq \dim W.$$

(b) If T is surjective, then $T(V) = W$. Thus

$$\dim V = \dim T(V) + \dim \operatorname{Ker} T = \dim W + \dim \operatorname{Ker} T \geq \dim W.$$

(c) This follows immediately from the previous two parts.

□

Exercises.

3.5.1. Show that B, as defined in the proof of Theorem 3.5.4 is a basis of V.

3.5.2 ([Ber14, Ex. 3.6.1]). Does there exists a linear map $T\colon \mathbb{R}^7 \to \mathbb{R}^3$ whose kernel is 3-dimensional?

3.5.3. For this exercise, the notation $U \xrightarrow{T} V$ means that T is a mapping from U to V (that is, it means the same thing as $T\colon U \to V$).

Let

$$\{\mathbf{0}\} \xrightarrow{T_0} V_1 \xrightarrow{T_1} V_2 \xrightarrow{T_2} V_3 \xrightarrow{T_3} \{\mathbf{0}\}$$

be linear maps satisfying:

$$\operatorname{Im} T_{i-1} = \operatorname{Ker} T_i \quad \text{for all } i = 1, 2, 3.$$

(a) Show that T_1 is injective and that T_2 is surjective.

(b) Show that $\sum_{i=1}^{3} (-1)^i \dim V_i = 0$.

3.5.4 ([Ber14, Ex. 3.6.14]). Let $T\colon V \to W$ and $S\colon W \to U$ be linear maps, with V finite dimensional.

(a) If S is injective, then $\operatorname{Ker} ST = \operatorname{Ker} T$ and $\operatorname{rank}(ST) = \operatorname{rank}(T)$.

(b) If T is surjective, then $\operatorname{Im} ST = \operatorname{Im} S$ and $\operatorname{null}(ST) - \operatorname{null}(S) = \dim V - \dim W$.

3.5.5 ([Ber14, Ex. 3.7.3]). Let V be a finite-dimensional vector space and let $S, T \in \mathcal{L}(V)$. Prove the following statements.

(a) $\operatorname{rank}(S + T) \leq \operatorname{rank} S + \operatorname{rank} T$. *Hint*: If M and N are two subspaces of V, use the map $M \times N \to M + N$, $(x, y) \mapsto x + y$, to show that $\dim(M + N) \leq \dim M + \dim N$.

(b) $\operatorname{rank}(ST) \leq \operatorname{rank} S$, and $\operatorname{rank}(ST) \leq \operatorname{rank} T$.

(c) $\operatorname{null}(S+T) \geq \operatorname{null} S + \operatorname{null} T - \dim V$.

3.5.6 ([Ber14, Ex.3.7.6]). Let $V = \mathcal{P}(\mathbb{R})$ be the vector space of all real polynomial functions (see Example 1.2.8). Let $S, T \in \mathcal{L}(V)$ be the linear maps such that $Tp = p'$ and Sp is the antiderivative of p with constant term zero. Then $TS = I$. Is T bijective?

3.5.7 ([Ber14, Ex.3.7.7]). Let V be any vector space (not necessarily finite dimensional) and suppose $R, S, T \in \mathcal{L}(V)$ are such that $ST = I$ and $TR = I$. Prove that T is bijective and $R = S = T^{-1}$. *Hint*: Look at $(ST)R$.

3.6 Dimensions of spaces of linear maps

Theorem 3.6.1. *Suppose V is an n-dimensional vector space and W is an arbitrary (possibly infinite-dimensional) vector space. If $\{v_1, \dots, v_n\}$ is a basis of V and $w_1, \dots, w_n$ are* any *vectors in W, then there exists a unique linear map $T\colon V \to W$ such that $Tv_i = w_i$ for all $i = 1, \dots, n$.*

Proof. Existence: We know by Theorem 2.1.4 that the maps $R\colon F^n \to V$ and $S\colon F^n \to W$ defined by

$$R(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n,$$
$$S(a_1, \dots, a_n) = a_1 w_1 + \cdots + a_n w_n,$$

are linear. It is clear that if $\{e_1, \dots, e_n\}$ is the canonical basis of F^n, then

$$Re_i = v_i, \quad Se_i = w_i.$$

Since $\{v_1, \dots, v_n\}$ is a basis of V, R is bijective by Theorem 3.4.4. So R is invertible and R^{-1} is linear by Theorem 2.4.9. Therefore, the map $T := SR^{-1}$ is linear. Since

$$Tv_i = SR^{-1}v_i = Se_i = w_i,$$

the map T has the desired properties.

Uniqueness: Suppose $T_1, T_2\colon V \to W$ are two linear maps with the given property. Then

$$(T_1 - T_2)(v_i) = T_1 v_i - T_2 v_i = w_i - w_i = \mathbf{0} \quad \forall\, i = 1, \dots, n.$$

Since $\{v_1, \dots, v_n\}$ is a basis for V, this means that $T_1 - T_2 = 0$ (the zero map). Thus $T_1 = T_2$. □

Corollary 3.6.2. *If V and W are finite-dimensional vector spaces, then $\mathcal{L}(V, W)$ is also finite dimensional and*

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Proof. Let $n = \dim V$ and let $\{v_1, \dots, v_n\}$ be a basis of V. Define a map

$$\Phi\colon \mathcal{L}(V, W) \to W^n = W \times W \times \cdots \times W, \quad \Phi(T) = (Tv_1, \dots, Tv_n).$$

Then Φ is linear (Exercise 3.6.1) and bijective by Theorem 3.6.1. Hence Φ is a vector space isomorphism and so

$$\dim \mathcal{L}(V, W) = \dim W^n = n(\dim W) = (\dim V)(\dim W).$$

(In the first equality, we used Theorem 3.5.4). □

Exercises.

3.6.1. Prove that the map Φ defined in the proof of Corollary 3.6.2 is linear.

3.6.2. Suppose V and W are finite-dimensional vector spaces over a field F, and $\dim \mathcal{L}(V, W) = 11$. Show that either $V \cong F$ or $W \cong F$.

3.6.3 ([Ber14, Ex. 3.8.2]). True or false (explain): $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^3) \cong \mathbb{R}^6$.

3.6.4 ([Ber14, Ex. 3.8.3]). If V and W are finite-dimensional vector spaces, prove that $\mathcal{L}(V, W) \cong \mathcal{L}(W, V)$. *Hint*: Don't look for a map $\mathcal{L}(V, W) \to \mathcal{L}(W, V)$.

3.6.5 ([Ber14, Ex. 3.8.7]). Prove that if V is a finite-dimensional vector space and $T \in \mathcal{L}(V)$, then there exists a positive integer r and scalars $a_0, a_1, \ldots, a_r$ (not all zero) such that

$$a_0 v + a_1 T v + a_2 T^2 v + \cdots + a_r T^r v = \mathbf{0}, \quad \forall\ v \in V.$$

3.6.6 ([Ber14, Ex. 3.8.8]). Let $x_1, \ldots, x_n$ be a basis of V and, for each pair of indices $i, j \in \{1, \ldots, n\}$, let $E_{i,j} \in \mathcal{L}(V)$ be the linear mapping given by

$$E_{i,j} x_k = \begin{cases} x_i & \text{if } j = k, \\ \mathbf{0} & \text{if } j \neq k. \end{cases}$$

Prove the following statements:

(a) $E_{i,j} E_{j,k} = E_{i,k}$;

(b) $E_{i,j} E_{h,k} = \mathbf{0}$ if $j \neq h$;

(c) $E_{1,1} + E_{2,2} + \cdots + E_{n,n} = I$.

3.7 Dual spaces

Recall (Definition 2.1.10) that if V is a vector space over a field F then the *dual space* of linear forms is

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

Remark 3.7.1. Some references use the notation V' instead of V^*.

Theorem 3.7.2. *If* $\dim V < \infty$, *then* $\dim V^* = \dim V$.

Proof. This follows from Corollary 3.6.2:

$$\dim V^* = \dim \mathcal{L}(V, F) = (\dim V)(\dim F) = (\dim V) \cdot 1 = \dim V. \qquad \square$$

Proposition 3.7.3 (Existence of dual bases)**.** *Suppose* $\{v_1, \ldots, v_n\}$ *is a basis of a vector space* V. *Then there exists a basis* $\{f_1, \ldots, f_n\}$ *of* V^* *such that*

$$f_i(v_j) = \delta_{ij} := \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

(The symbol δ_{ij} *is called the* Kronecker delta*.)*

Proof. For $i = 1, \ldots, n$, define

$$f_i \colon V \to F, \quad f_i \left(\sum_{j=1}^{n} c_j v_j \right) = c_i. \tag{3.3}$$

Then each f_i is linear (Exercise 3.7.1) and so $f_i \in V^*$. Moreover, we have

$$f_i(v_j) = f_i(0v_1 + \cdots + 0v_{j-1} + 1v_j + 0v_{j+1} + \cdots + 0v_n) = \delta_{ij}.$$

Since we know from Theorem 3.7.2 that $\dim V^* = n$, to show that $\{f_1, \ldots, f_n\}$ is a basis, it is enough to show that this set is linearly independent (Theorem 3.4.14). Now, for $c_1, \ldots, c_n \in F$,

$$\begin{aligned} \sum_{j=1}^{n} c_j f_j = 0 &\implies \sum_{j=1}^{n} c_j f_j(v_i) = 0 \ \forall \ i = 1, \ldots, n \\ &\implies \sum_{j=1}^{n} c_j \delta_{ji} = 0 \ \forall \ i = 1, \ldots, n \\ &\implies c_i = 0 \ \forall \ i = 1, \ldots, n. \end{aligned}$$

Thus $\{f_1, \ldots, f_n\}$ is linearly independent and so we're done. $\square$

Definition 3.7.4 (Dual basis)**.** We call $\{f_1, \ldots, f_n\}$ the *basis of* V^* *dual to* $\{v_1, \ldots, v_n\}$.

Example 3.7.5. If $\{e_1, \ldots, e_n\}$ is the canonical basis of F^n, then the dual basis of $(F^n)^*$ is $\{f_1, \ldots, f_n\}$ where

$$f_i(a_1, \ldots, a_n) = a_i$$

is the i-th *coordinate function*.

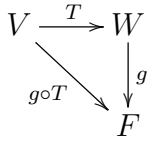
If $f \colon \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x, y, z) = x - 2y + 3z$ (so $f \in (\mathbb{R}^3)^*$), then

$$f = f_1 - 2f_2 + 3f_3.$$

Theorem 3.7.6. *Suppose $T\colon V \to W$ is a linear map. Then the map $T^*\colon W^* \to V^*$ defined by*

$$T^*g = g \circ T, \quad g \in W^*$$

is linear.



Proof. First note that, for $g \in W^* = \mathcal{L}(W, F)$, the composition $g \circ T$ is indeed a linear map from V to F by Theorem 2.3.13, that is, we have $g \circ T \in V^*$.

We need to show T^* is linear. Suppose $g, h \in W^*$ and c, d are scalars. Then for all $v \in V$, we have

$$\begin{aligned}(T^*(cg + dh))(v) &= ((cg + dh)T)(v)\\ &= (cg + dh)(Tv)\\ &= c(g(Tv)) + d(h(Tv))\\ &= c(gT)(v) + d(hT)(v)\\ &= c(T^*g)(v) + d(T^*h)(v)\\ &= (cT^*g + dT^*h)(v).\end{aligned}$$

Thus $T^*(cg + dh) = cT^*g + dT^*h$. Therefore T^* is linear. □

Definition 3.7.7 (Transpose)**.** The map T^* is called the *transpose* of the map T.

Remark 3.7.8. We will see later than the transpose of a linear map is closely related to the transpose of a matrix (which you learned about in MAT 1341).

Theorem 3.7.9. *If $T\colon V \to W$ is a linear map, then*

$$\operatorname{Ker} T^* = \{g \in W^* \mid g(w) = 0 \ \forall\ w \in T(V)\}.$$

Proof. If $g \in W^*$, then

$$\begin{aligned}g \in \operatorname{Ker} T^* \iff T^*g = 0 \iff gT = 0 \iff (gT)(v) = 0 \ \forall\ v \in V\\ \iff g(Tv) = 0 \ \forall\ v \in V \iff g(w) = 0 \ \forall\ w \in T(V).\end{aligned}$$ □

Definition 3.7.10 (Annihilator)**.** If M is a subspace of V, then the *annihilator* of M in V^* is

$$M^\circ := \{f \in V^* \mid f(w) = 0 \ \forall\ w \in M\} = \{f \in V^* \mid f = 0 \text{ on } M\}.$$

Note that M° is a subspace of V^*.

Using the above terminology, the result of Theorem 3.7.9 is

$$\operatorname{Ker} T^* = T(V)^\circ = (\operatorname{Im} T)^\circ. \tag{3.4}$$

Theorem 3.7.11. *Suppose V and W are finite dimensional vector spaces. If $T\colon V \to W$ is a linear map, then* $\operatorname{Im} T^* = (\operatorname{Ker} T)^\circ$.

Proof. Let $f \in \operatorname{Im} T^*$. Then $f = T^* g$ for some $g \in W^*$. Thus

$$f(v) = (T^* g)(v) = gTv \ \forall\ v \in V.$$

So

$$v \in \operatorname{Ker} T \implies Tv = 0 \implies f(v) = g(0) = 0.$$

Therefore $\operatorname{Im} T^* \subseteq (\operatorname{Ker} T)^\circ$.

Now suppose $f \in (\operatorname{Ker} T)^\circ$. We want to find a $g \in W^*$ such that $f = T^* g$. Let $\{v_1, \dots, v_k\}$ be a basis of $\operatorname{Ker} T$ and extend this to a basis $\{v_1, \dots, v_k, u_1, \dots, u_l\}$ of V. We know that $\{Tu_1, \dots, Tu_l\}$ is a basis of $\operatorname{Im} T$. (We showed this in our proof of the Dimension Theorem (Theorem 3.5.1).) Extend this to a basis of $\{Tu_1, \dots, Tu_l, w_1, \dots, w_p\}$ of W. Now define $g \in W^*$ as follows. For scalars $c_1, \dots, c_l, a_1, \dots, a_p$, define

$$g(c_1 Tu_1 + \cdots + c_l Tu_l + a_1 w_1 + \cdots + a_p w_p) = c_1 f(u_1) + \cdots + c_l f(u_l) = f(c_1 u_1 + \cdots + c_l u_l).$$

Then g is linear by Theorem 3.6.1. We will show that $T^* g = f$. Let $x \in V$ and write $x = v + u$ with $v \in \operatorname{Ker} T$ and $u \in \operatorname{Span}\{u_1, \dots, u_l\}$. Then

$$(T^* g)(x) = gT(v + u) = gT(u) = f(u).$$

On the other hand,

$$f(x) = f(v + u) = f(v) + f(u) = f(u),$$

since $f \in (\operatorname{Ker} T)^\circ$. Hence $T^* g = f$. Thus $(\operatorname{Ker} T)^\circ \subseteq \operatorname{Im} T^*$ and so $(\operatorname{Ker} T)^\circ = \operatorname{Im} T^*$. □

In fact, Theorem 3.7.11 is true without the assumptions that V and W be finite dimensional. However, the proof of the general result uses the notion of a quotient space (or more general facts about bases), which we will not cover in this course. See Appendix B.2.

Theorem 3.7.12. *If U is a subspace of V and* $\dim V < \infty$, *then*

$$\dim U^\circ = \dim V - \dim U = \dim V^* - \dim U^*.$$

Proof. Let $j\colon U \to V$ be the inclusion map (i.e. $j(u) = u$ for all $u \in U$). Then we have

$$\operatorname{Ker} j^* = (\operatorname{Im} j)^\circ \quad \text{and} \quad \operatorname{Im} j^* = (\operatorname{Ker} j)^\circ.$$

Now, $\operatorname{Im} j = U$ and $\operatorname{Ker} j = \{\mathbf{0}\}$. Thus

$$\operatorname{Ker} j^* = U^\circ \quad \text{and} \quad \operatorname{Im} j^* = \{\mathbf{0}\}^\circ = U^*$$

(so j^* is surjective). By the Dimension Theorem (Theorem 3.5.1), we have

$$\dim \operatorname{Ker} j^* + \dim \operatorname{Im} j^* = \dim V^*.$$

Thus

$$\dim U^\circ + \dim U^* = \dim V^* = \dim V.$$

Since $\dim U = \dim U^*$, the result follows. □

Corollary 3.7.13. *If $T\colon V \to W$ is linear, and $\dim V, \dim W < \infty$, then*

(a) $\operatorname{rank} T = \operatorname{rank} T^*$,

(b) *T is surjective $\iff$ T^* is injective, and*

(c) *T is injective $\iff$ T^* is surjective.*

Proof. (a) We have

$$\begin{aligned}
\operatorname{rank} T^* &= \dim W^* - \dim(\operatorname{Ker} T^*) && \text{(by the Dimension Theorem)}\\
&= \dim W^* - \dim(\operatorname{Im} T)^\circ && \text{(by (3.4))}\\
&= \dim W^* - \big(\dim W - \dim(\operatorname{Im} T)\big) && \text{(by Theorem 3.7.12)}\\
&= \dim \operatorname{Im} T && \text{(since } \dim W = \dim W^* \text{ by Theorem 3.7.2)}\\
&= \operatorname{rank} T.
\end{aligned}$$

(b) We have

$$\begin{aligned}
T \text{ is surjective} &\iff \operatorname{rank} T = \dim W\\
&\iff \operatorname{rank} T^* = \dim W^* && \text{(by part (a) and Theorem 3.7.2)}\\
&\iff \dim \operatorname{Ker} T^* = 0 && \text{(by the Dimension Theorem)}\\
&\iff T^* \text{ is injective.}
\end{aligned}$$

(c) We have

$$\begin{aligned}
T \text{ is injective} &\iff \operatorname{null} T = 0\\
&\iff \operatorname{rank} T = \dim V && \text{(by the Dimension Theorem)}\\
&\iff \operatorname{rank} T^* = \dim V^* && \text{(by part (a) and Theorem 3.7.2)}\\
&\iff T^* \text{ is surjective.}
\end{aligned}$$

□

Exercises.

3.7.1. Show that the maps f_i defined by (3.3) are linear.

3.7.2 ([Ber14, Ex. 3.6.7]). Let V be an n-dimensional vector space and let f be a nonzero linear form on V. Prove that $\dim \operatorname{Ker}(f) = n - 1$. Conversely, show that every $(n-1)$-dimensional linear subspace of V is the kernel of a linear form.

3.7.3 ([Ber14, Ex. 3.6.9]). Let V be a vector space and suppose $f_1, \dots, f_n$ are linear forms on V such that

$$(\operatorname{Ker} f_1) \cap \cdots \cap (\operatorname{Ker} f_n) = \{\mathbf{0}\}.$$

Prove that V is finite-dimensional and $\dim V \le n$. *Hint*: If F is the field of scalars, consider a suitable mapping $V \to F^n$.

3.7.4 ([Ber14, Ex. 3.6.10]). Let V be the set of all vectors $(x, y, z) \in \mathbb{R}^3$ such that $2x-3y+z = 0$. Prove that V is a subspace of $\mathbb{R}^3$ and find its dimension. *Hint*: Use Exercise 3.7.2.

3.7.5 ([Ber14, Ex. 3.9.3]). If V and W are finite-dimensional vector spaces and $T\colon V \to W$ is a linear map, prove that T and T^* have the same nullity if and only if $\dim V = \dim W$.

3.7.6 ([Ber14, Ex. 3.9.5]). Let $V = \mathcal{P}(\mathbb{R})$ be the vector space of real polynomial functions (Example 1.2.8). Every nonnegative integer k defined a linear form φ_k on V given by $\varphi_k(p) = p^{(k)}(0)$, where $p^{(k)}$ denotes the k-th derivative of p. If $D\colon V \to V$ is the differentiation map $Dp = p'$, show that $D^*\varphi_k = \varphi_{k+1}$. *Hint*: $\varphi_k(p) = (D^k p)(0)$.

3.7.7 ([Ber14, Ex. 3.9.7]). Suppose V and W are finite-dimensional vector spaces and $T\colon V \to W$ is a linear map. Prove that T is bijective if and only if T^* is bijective. *Hint*: Use Corollary 3.7.13.

3.7.8 ([Ber14, Ex. 3.9.10]). Let V and W be finite-dimensional vector spaces. Show that the mapping

$$\Phi\colon \mathcal{L}(V, W) \to \mathcal{L}(W^*, V^*), \quad \Phi(T) = T^*$$

is linear and bijective.

3.7.9 ([Ber14, Ex. 3.9.13]). Let V be a finite-dimensional vector space, M a subspace of V, and M° the annihilator of M in V^*. Prove the following:

(a) $M^\circ = \{\mathbf{0}\}$ if and only if $M = V$;

(b) $M^\circ = V^*$ if and only if $M = \{\mathbf{0}\}$.

Chapter 4

Matrices

In this chapter we look at matrices. Some of this material will be familiar from MAT 1341. However, we will cover the topic of matrices in more detail here. In particular, we discuss the matrix of a linear map, which generalizes the "standard matrix" of a map $\mathbb{R}^n \to \mathbb{R}^m$ seen in MAT 1341. We will then recall the procedure of gaussian elimination (row reduction) and the concept of the rank of a matrix. The material in this section roughly corresponds to [Tre, §2.2, §2.7, §2.8].

4.1 The matrix of a linear map

Choosing bases allows one to associated a matrix to any linear map between finite-dimensional vector spaces, as we now explain.

Definition 4.1.1 (The matrix of a linear map)**.** Suppose V and W are finite-dimensional vector spaces over a field F, $n = \dim V$, $m = \dim W$, and $T\colon V \to W$ is a linear map. Let $B = \{v_1, \ldots, v_n\}$ be an ordered basis of V and let $D = \{w_1, \ldots, w_m\}$ be an ordered basis of W. For each $j = 1, \ldots, n$, write Tv_j as a linear combination of $w_1, \ldots, w_m$:

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i, \quad j = 1, \ldots, n.$$

Then the $m \times n$ matrix $[a_{ij}]$ is called the *matrix of T relative to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$* and is denoted $[T]_B^D$.

Remark 4.1.2. (a) It is important that the bases be *ordered*. Changing the order of vectors in the bases will change the matrix.

(b) We will sometimes use the notation $[T]$ instead of $[T]_B^D$ when we've fixed the bases B and D and there is no chance of confusion.

(c) A given linear map can have different matrices (if you pick different bases).

Recall that if V is an n-dimensional space, and $B = \{v_1, \ldots, v_n\}$ is an ordered basis of

V, then we have an isomorphism

$$C_B \colon V \to F^n, \quad C_B\left(\sum_{i=1}^{n} c_i v_i\right) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in F^n.$$

So C_B is the *coordinate* function that takes the coordinates of a vector relative to the basis B. If D is an ordered basis of W and $T \in \mathcal{L}(V, W)$, we have the following diagram.

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle C_B}\downarrow\uparrow{\scriptstyle C_B^{-1}} & & {\scriptstyle C_D}\downarrow\uparrow{\scriptstyle C_D^{-1}} \\ F^n & \xrightarrow[C_D T C_B^{-1}]{} & F^m \end{array}$$

We know from MAT 1341 that a linear map $F^n \to F^m$ corresponds to multiplication by a matrix. (In MAT 1341, F was $\mathbb{R}$ or $\mathbb{C}$, but the argument is the same in general.) So $C_D T C_B^{-1}$ corresponds to multiplication by some matrix, and this is the matrix $[T]_B^D$. This gives us an isomorphism

$$\mathcal{L}(V, W) \to \mathcal{L}(F^n, F^m), \quad T \mapsto C_D T C_B^{-1},$$

and an isomorphism

$$\mathcal{L}(V, W) \to M_{m,n}(F), \quad T \mapsto [T]_B^D.$$

We often simply identify an $m \times n$ matrix with the corresponding map $F^n \to F^m$ (given by multiplication by that matrix) and so we write $[T]_B^D = C_D T C_B^{-1}$.

Note that

$$C_D(Tv_j) = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

and so

$$[T]_B^D = \begin{bmatrix} C_D(Tv_1) & C_D(Tv_2) & \cdots & C_D(Tv_n) \end{bmatrix}.$$

Example 4.1.3. Let $S \colon F^n \to F^m$ be a linear map and choose the standard bases $B = \{e_1, \ldots, e_n\}$ and $D = \{e_1, \ldots, e_m\}$. Then $C_B \colon F^n \to F^n$ and $C_D \colon F^n \to F^n$ are the identity maps. Thus

$$[T]_B^D = \begin{bmatrix} Se_1 & Se_2 & \cdots & Se_n \end{bmatrix}$$

is the standard matrix of the linear transformation (as you saw in MAT 1341).

Recall (from MAT 1341) that, for a matrix A, $\operatorname{col} A$ denotes the *column space* of A (the span of the columns of A).

Proposition 4.1.4. *With the same notation as above, we have*

(a) $[T]C_B(v) = C_D T(v)$ *for all* $v \in V$,

(b) $C_B(\operatorname{Ker} T) = \operatorname{Ker}[T]$ *or* $\operatorname{Ker} T = C_B^{-1}(\operatorname{Ker}[T])$, *and*

(c) $C_D(\operatorname{Im} T) = \operatorname{Im}[T] = \operatorname{col}[T]$ *or* $\operatorname{Im} T = C_D^{-1}(\operatorname{col}[T])$.

Proof. Recall that we have $[T] = C_D T C_B^{-1}$. Composing on the left with C_B gives (a). Since C_B and C_D are isomorphisms, we have

$$\begin{aligned} v \in \operatorname{Ker}[T] &\iff [T]v = \mathbf{0} \\ &\iff C_D T C_B^{-1} v = \mathbf{0} \\ &\iff T C_B^{-1} v = \mathbf{0} \\ &\iff C_B^{-1} v \in \operatorname{Ker} T \\ &\iff v \in C_B(\operatorname{Ker} T), \end{aligned}$$

which proves (b). Finally, to prove (c), we have

$$\begin{aligned} v \in \operatorname{Im}[T] &\iff [T]w = v \text{ for some } w \in F^n \\ &\iff C_D T C_B^{-1} w = v \text{ for some } w \in F^n \\ &\iff C_D T x = v \text{ for some } x \in V \\ &\iff C_D y = v \text{ for some } y \in \operatorname{Im} T \\ &\iff v \in C_D(\operatorname{Im} T). \end{aligned}$$

□

Example 4.1.5. Let $T\colon \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the linear map given by $T(p) = p' - p''$. Choose ordered bases $B = \{1, t, t^2, t^3\}$ and $D = \{1, t, t^2\}$ of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$ respectively. Then

$$\begin{aligned} [T] &= \begin{bmatrix} C_D T(1) & C_D T(t) & C_D T(t^2) & C_D T(t^3) \end{bmatrix} \\ &= \begin{bmatrix} C_D(0) & C_D(1) & C_D(2t-2) & C_D(3t^2 - 6t) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & -2 & 0 \\ 0 & 0 & 2 & -6 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \end{aligned}$$

Now let's find the kernel and image of T. Using techniques from MAT 1341, we know that

$$\operatorname{Ker}[T] = \operatorname{Span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right), \quad \operatorname{col}[T] = \mathbb{R}^3.$$

Therefore

$$\operatorname{Ker} T = \operatorname{Span}\left\{ C_B^{-1} \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right) \right\} = \operatorname{Span}\{1\},$$

and

$$\operatorname{Im} T = C_D^{-1}(\mathbb{R}^3) = \mathcal{P}_2(\mathbb{R}).$$

Note that if we have a linear map $T\colon V \to V$ from some vector space V to itself, we often use the same bases for V as the domain and as the codomain. But this is not always the case.

Example 4.1.6. Consider the bases $B = \{e_1, e_2\}$ and $D = \{(1,1), (1,-1)\}$ of $\mathbb{R}^2$ and let $I\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the identity map. Then

$$[I]_B^D = \begin{bmatrix} C_D I(e_1) & C_D I(e_2) \end{bmatrix} = \begin{bmatrix} C_D(e_1) & C_D(e_2) \end{bmatrix}.$$

To finish the calculation, we need to find $C_D(e_1)$ and $C_D(e_2)$. In other words, we need to write e_1 and e_2 in the basis D. We have

$$e_1 = (1,0) = \frac{1}{2}(1,1) + \frac{1}{2}(1,-1), \quad e_2 = \frac{1}{2}(1,1) - \frac{1}{2}(1,-1).$$

Thus

$$[I]_B^D = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}.$$

Exercises.

4.1.1. Choose the same bases for $\mathcal{P}_2(\mathbb{R})$ and $\mathcal{P}_3(\mathbb{R})$ as in Example 4.1.5 and let $S\colon \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be the linear map defined by letting $S(p)$ be the antiderivative of p with constant term zero. Find the matrix $[S]$ for S and check that $[S]C_D(p) = C_B S(p)$ for all $p \in \mathcal{P}_2(\mathbb{R})$.

4.1.2 ([Ber14, Ex. 4.2.1]). Lev V be a 3-dimensional vector space, let x_1, x_2, x_3 be a basis of V, and let $T \in \mathcal{L}(V)$ be the linear map given by

$$Tx_1 = x_1, \quad Tx_2 = x_1 + x_2, \quad Tx_3 = x_1 + x_2 + x_3.$$

Find the matrices of T and T^{-1} relative to the basis x_1, x_2, x_3.

4.1.3. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the linear map defined by

$$T(x,y) = (3x + 2y, x - y, 4x + 5y).$$

Find the matrix of T relative to the canonical bases of $\mathbb{R}^2$ and $\mathbb{R}^3$.

4.1.4 ([Ber14, Ex. 4.2.4]). Let $V = \mathcal{P}_3(\mathbb{R})$ be the vector space of real polynomials of degree ≤ 3 (see Example 1.2.8), and let $D\colon V \to V$ be the linear map $Dp = p'$ (the derivative of p). Find the matrix of D relative to the basis $1, t, t^2, t^3$ of V.

4.1.5 ([Ber14, Ex. 4.2.5]). Fix $y \in \mathbb{R}^3$ and define

$$T\colon \mathbb{R}^3 \to \mathbb{R}^3, \quad Tx = x \times y$$

(cross product of vectors). Find the matrix of T relative to the canonical basis of $\mathbb{R}^3$.

4.1.6 ([Ber14, Ex. 4.2.7]). If $A = (a_{i,j})$ is an $n \times n$ matrix, the *trace* of A is

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{i,i}.$$

Prove that $\operatorname{tr}\colon M_n(F) \to F$ is a linear form on the vector space $M_n(F)$ of all $n \times n$ matrices over the field F.

4.1.7 ([Ber14, Ex. 4.2.8]). Let V be an n-dimensional vector space over F with basis $x_1, \ldots, x_n$, and let $f\colon V \to F$ be a linear form on V. Describe the matrix of f relative to the basis $x_1, \ldots, x_n$ of V and the basis 1 of F.

4.1.8 ([Ber14, Ex. 4.2.9]). Recall (from MAT 1341) that if $A = (a_{i,j})$ is an $m \times n$ matrix over F, the *transpose* of A, denoted A^t, is the $n \times m$ matrix $(b_{i,j})$, where $b_{i,j} = a_{j,i}$. Prove that $A \mapsto A^t$ is a vector space isomorphism $M_{m,n}(F) \to M_{n,m}(F)$.

4.1.9 ([Ber14, Ex. 4.2.10]). Let $V = W = \mathbb{R}^2$. Choose the basis $B = \{x_1, x_2\}$ of V, where $x_1 = (2, 3)$, $x_2 = (4, -5)$ and choose the basis $D = \{y_1, y_2\}$ of W, where $y_1 = (1, 1)$, $y_2 = (-3, 4)$. Find the matrix of the identity linear mapping $I\colon V \to W$ with respect to these bases.

4.1.10. Suppose U, V and W are vector spaces (over the same field) with ordered bases B, D, and E respectively. Suppose we have linear maps

$$U \xrightarrow{T} V \xrightarrow{S} W.$$

Show that $[ST]_B^E = [S]_D^E \cdot [T]_B^D$ (where the product on the right hand side is matrix multiplication). Note that we must use the same basis for the intermediate space V in both matrices.

4.1.11. Suppose V is a vector space with two ordered bases B and D. Show that a linear map $T\colon V \to V$ is invertible if and only if the matrix $[T]_B^D$ is invertible. Furthermore, if T is invertible, show that $[T^{-1}]_D^B = \left([T]_B^D\right)^{-1}$. *Hint*: Use Exercise 4.1.10.

4.2 Change of bases and similar matrices

Here we specialize to linear maps $T\colon V \to V$ from a vector space to itself and assume V is finite dimensional. What is the relationship between the matrices of T in different bases? Recall that

$$[T]_B^B = C_B T C_B^{-1}, \quad \text{and so} \quad C_B^{-1}[T]_B^B C_B = T.$$

Now suppose D is another ordered basis of V. Then

$$C_B^{-1}[T]_B^B C_B = T = C_D^{-1}[T]_D^D C_D,$$

and so

$$[T]_D^D = C_D C_B^{-1}[T]_B^B C_B C_D^{-1}.$$

Let $P = C_B C_D^{-1}$. Then $P^{-1} = C_D C_B^{-1}$ and so

$$[T]_D^D = P^{-1}[T]_B^B P.$$

Now, if $B = \{v_1, \ldots, v_n\}$ and $D = \{w_1, \ldots, w_n\}$, then

$$C_D(w_j) = e_j, \quad \text{and so} \quad w_j = C_D^{-1}(e_j).$$

Therefore

$$Pe_j = C_B C_D^{-1}(e_j) = C_B(w_j).$$

Since Pe_j is just the j-th column of P, we see that

The j-th column of P gives the coordinates of w_j in the basis B.

Thus, if $P = [P_{ij}]$, then $w_j = \sum_{i=1}^n P_{ij} v_i$. The matrix P is called the *change of basis matrix* from B to D.

Example 4.2.1. Suppose $V = \mathbb{R}^2$ and $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the linear map given by multiplication by the matrix

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Let $B = \{e_1, e_2\}$ be the standard basis and $D = \{w_1, w_2\}$ where $w_1 = (1, 1)$ and $w_2 = (1, -1)$. Then

$$[T]_B^B = \begin{bmatrix} C_B T(e_1) & C_B T(e_2) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

is the standard matrix of T since B is the standard basis of $\mathbb{R}^2$. However,

$$[T]_D^D = \begin{bmatrix} C_D\left(T\begin{bmatrix}1\\1\end{bmatrix}\right) & C_D\left(T\begin{bmatrix}1\\-1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} C_D\left(\begin{bmatrix}3\\3\end{bmatrix}\right) & C_D\left(\begin{bmatrix}1\\-1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix},$$

a diagonal matrix!

Note that $Pe_j = C_B C_D^{-1}(e_j)$, so

$$Pe_1 = C_B C_D^{-1}(e_1) = C_B(w_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad Pe_2 = C_B C_D^{-1}(e_2) = C_B(w_2) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and

$$P^{-1} = \frac{-1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Therefore

$$P^{-1}[T]_B^B P = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 3 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = [T]_D^D.$$

Remark 4.2.2. In this particular case, w_1 and w_2 are eigenvectors of the matrix $\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$, with eigenvalues 2 and 1, respectively.

$$P^{-1}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}P = \begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}$$

is diagonal and the columns of P are w_1 and w_2 (which are $C_B(w_1)$ and $C_B(w_2)$ since B is the standard basis!), and the diagonal entries of $[T]_D^D$ are the eigenvalues. The whole idea of changing bases is (usually) to *simplify* $[T]_B^B$ as much as possible.

Definition 4.2.3 (Similar matrices)**.** Two $n \times n$ matrices X and Y are *similar* if there is an invertible matrix P such that $P^{-1}XP = Y$.

Examples 4.2.4.

$$\text{The matrix } \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \text{ is similar to } \begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}, \quad \text{with } P = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

$$\text{The matrix } \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \text{ is similar to } \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}, \quad \text{with } P = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

$$\text{The matrix } \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \text{ is similar to } P^{-1}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}P \text{ and } Q\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}Q^{-1} \text{ for any } P \text{ (or } Q).$$

Remark 4.2.5. (a) Similarity is an equivalence relation on matrices (Exercise 4.2.5).

(b) Over an infinite field, a given matrix is *usually* similar to infinitely many matrices. However, there are exceptions. For example, any matrix similar to the identity matrix I_n is of the form $P^{-1}I_nP = P^{-1}P = I_n$. So I_n is only similar to itself. The same is true of cI_n for any scalar c.

The next theorem explains why we used the word 'transpose' in Definition 3.7.7.

Theorem 4.2.6. *Suppose $T\colon V \to W$ is a linear map between finite-dimensional vector spaces and $T^*\colon W^* \to V^*$ is the transpose map. If B and D are ordered basis of V and W respectively and B^* and D^* are the corresponding dual bases, then*

$$\left([T]_B^D\right)^t = [T^*]_{D^*}^{B^*},$$

where A^t denotes the transpose of the matrix A (see Exercise 4.1.8).

Proof. Let

$$B = \{v_1, \ldots, v_n\}, \quad D = \{w_1, \ldots, w_m\}, \quad X = [T]_B^D = [X_{ij}],$$
$$B^* = \{f_1, \ldots, f_n\}, \quad D^* = \{g_1, \ldots, g_m\}, \quad Y = [T^*]_{D^*}^{B^*} = [Y_{ij}].$$

Then, for all $1 \le i \le n$, $1 \le j \le m$, we have

$$(T^*g_j)(v_i) = (g_jT)(v_i) = g_j(Tv_i) = g_j\left(\sum_{k=1}^{m} X_{ki}w_k\right) = X_{ji}.$$

On the other hand,

$$(T^*g_j)(v_i) = \left(\sum_{l=1}^{n} Y_{lj} f_l\right)(v_i) = Y_{ij}.$$

Thus $X_{ji} = Y_{ij}$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$. Therefore $X^t = Y$. □

Let's do one final example on matrices associated to linear maps. Suppose one knows the matrix for a linear map in some bases. How do you compute the action of the linear map?

Example 4.2.7. Choose the ordered basis $B = \{1, t, t^2\}$ of $\mathcal{P}_2(\mathbb{R})$ and $D = \{e_{11}, e_{12}, e_{21}.e_{22}\}$ of $M_2(\mathbb{R})$ (here e_{ij} is the matrix with a one in the (i, j)-position and zeroes everywhere else). Suppose a linear map $T\colon \mathcal{P}_2(\mathbb{R}) \to M_2(\mathbb{R})$ has matrix

$$[T]_B^D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 5 & -1 \end{bmatrix}.$$

What is $T(t^2 - 1)$?

Recall that $[T]_B^D = C_D T C_B^{-1}$. Thus $T = C_D^{-1}[T]_B^D C_B$. So

$$\begin{aligned} T(t^2-1) &= C_D^{-1}[T]_B^D C_B(t^2-1) \\ &= C_D^{-1}[T]_B^D \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= C_D^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= C_D^{-1} \begin{bmatrix} -1 \\ 1 \\ -1 \\ -2 \end{bmatrix} \\ &= -e_{11} + e_{12} - e_{21} - 2e_{22} \\ &= \begin{bmatrix} -1 & 1 \\ -1 & -2 \end{bmatrix} \end{aligned}$$

Exercises.

4.2.1 ([Tre, Ex. 2.8.2]). Consider the vectors

$$(1, 2, 1, 1), \quad (0, 1, 3, 1), \quad (0, 3, 2, 0), \quad (0, 1, 0, 0).$$

(a) Prove that these vectors form a basis in $\mathbb{R}^4$.

(b) Find the change of basis matrix that changes from this basis to the standard basis of $\mathbb{R}^4$.

4.2.2 ([Tre, Ex. 2.8.3]). Find the change of basis matrix that changes from the basis $1, 1+t$ of $\mathcal{P}_1(\mathbb{R})$ to the basis $1-t, 2t$ (see Example 1.2.8).

4.2.3 ([Tre, Ex. 2.8.4]). Consider the linear map

$$T\colon \mathbb{C}^2 \to \mathbb{C}^2, \quad T(x,y) = (3x+y, x-2y).$$

Find the matrix of T in the standard basis and also in the basis $(1,1), (1,2)$.

4.2.4. Let

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad B = \{e_1, e_2\}, \quad D = \{(3,1), (-2,1)\}, \quad T(v) = Av.$$

Find $[T]_B^B$ and $[T]_D^D$.

4.2.5. Fix a positive integer n. Prove that similarity is an equivalence relation on the set of $n \times n$ matrices. (See Definition B.1.1 for the definition of equivalence relation.)

4.3 Gaussian elimination

In this section and the next we review some material from MAT 1341. Recall that you learned how to row reduce a matrix. This procedure is called *Gaussian elimination*. You used the following operations:

Definition 4.3.1 (Elementary row/column operations)**.** The following are called *elementary row operations* on a matrix A with entries in F.

- *Type I*: Interchange two rows of A.
- *Type II*: Multiply any row of A by a nonzero element of F.
- *Type III*: Add a multiple of one row of A to another row of A.

Replacing the word "row" by "column" everywhere above gives the *elementary column operations*.

Definition 4.3.2 (Elementary matrices)**.** An $n \times n$ *elementary matrix* is a matrix obtained by performing an elementary row/column operation on the identity matrix I_n. In particular, we define the following elementary matrices:

- For $1 \le i, j \le n$, $i \ne j$, we let $P_{i,j}$ be the elementary matrix obtained from I_n by interchanging the i-th and j-th rows (equivalently, columns).
- For $1 \le i \le n$ and $a \in F^\times$, we let $M_i(a)$ be the elementary matrix obtained from I_n by multiplying the i-th row (equivalently, column) by a.

- For $i \leq i, j \leq n$, $i \neq j$, and $a \in F$, we let $E_{i,j}(a)$ be the elementary matrix obtained from I_n by adding a times row j to row i (equivalently, adding a times column i to column j).

The *type* of the elementary matrix is the type of the corresponding row/column operation performed on I_n.

Lemma 4.3.3. *(a) Every elementary matrix is invertible and the inverse is an elementary matrix of the same type.*

(b) Performing an elementary row operation on a matrix A is equivalent to multiplying A on the left by the corresponding elementary matrix.

(c) Performing an elementary column operation on a matrix A is equivalent to multiplying A on the right by the corresponding elementary matrix.

Proof. You saw this in MAT 1341, and so we will omit the proof here. □

Definition 4.3.4 (Row-echelon form)**.** A matrix $R \in M_{m,n}(F)$ is in *row-echelon form* if:

(a) all nonzero rows are above all zero rows, and

(b) the leading coefficient (the first nonzero entry from the left) of a nonzero row is strictly to the right of the leading coefficient of the row above it.

Lemma 4.3.5. *Every matrix $A \in M_{m,n}(F)$ can be transformed to a row-echelon form matrix R by performing row operations of type I and III. Equivalently, there exist finitely many elementary matrices $E_1, E_2, \ldots, E_k$ such that $R = E_1 E_2 \cdots E_k A$.*

Proof. You saw this in MAT 1341, and so we will omit the proof here. □

Theorem 4.3.6. *By performing both row and column operations, every matrix $A \in M_{m,n}(F)$ can be transformed to a block matrix of the form*

$$D = \begin{bmatrix} I_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \tag{4.1}$$

for some r with $0 \leq r \leq \min(m, n)$.

Proof. Starting a matrix A, we perform the following steps:

- Use row operations to reduce the matrix to row-echelon form as in Lemma 4.3.5.
- Use row operations of type II to turn the leading entry in each row to a 1.
- Use column operations of type III to eliminate all the entries to the right of each leading entry.
- Use column operations of type I to move the column with a 1 in the top row to the first column, the column with a 1 in the second row to the second column, etc.

□

Exercises.

4.3.1. Using row and column operations, reduce the following matrices to the form indicated in Theorem 4.3.6:

(a) $\begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 3 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$

4.4 The rank of a matrix

Definition 4.4.1 (Rank of a matrix)**.** The *rank* of $A \in M_{m,n}(F)$, denoted $\operatorname{rank}(A)$ is the rank of the linear map

$$T\colon F^n \to F^m$$

given by matrix multiplication on the left by A(equivalently, such that $[T] = A$ with respect to the standard bases of F^n and F^m).

Lemma 4.4.2. *A matrix $A \in M_n(F)$ is invertible if and only if* $\operatorname{rank}(A) = n$.

Proof. Let $T\colon F^n \to F^n$ be the linear map given by matrix multiplication by A. Thus $[T] = A$, with respect to the standard basis of F^n. Then T is invertible if and only if A is invertible. Thus

$$\begin{aligned} \operatorname{rank}(A) = n &\iff \operatorname{rank}(T) = n && \text{(by Definition 4.4.1)} \\ &\iff \dim T(V) = n && \text{(by Definition 3.5.2)} \\ &\iff T \text{ is invertible} && \text{(by Corollary 3.5.3)} \\ &\iff A \text{ is invertible.} && \square \end{aligned}$$

Lemma 4.4.3. *A matrix $A \in M_n(F)$ is invertible if and only if it can be written as a product of elementary matrices.*

Proof. Since elementary matrices are invertible, it is clear that if A can be written as a product of elementary matrices, then A is invertible.

Now suppose that A is invertible. Then, by Lemma 4.4.2, $\operatorname{rank}(A) = n$. Thus, the matrix D of Theorem 4.3.6 is the identity matrix I_n. (You learned in MAT 1341 that you can use row reduction to convert an invertible matrix to the identity matrix.) Thus, there exists elementary matrices $E_1, \ldots, E_p$ and $F_1, \ldots, F_q$ such that

$$E_1 \cdots E_p A F_1 \cdots F_q = I_n.$$

Hence

$$A = E_p^{-1} \cdots E_1^{-1} F_q^{-1} \cdots F_1^{-1}.$$

Since the inverse of an elementary matrices is an elementary matrix, it follows that A can be written as a product of elementary matrices. □

Lemma 4.4.4. *If $A \in M_{m,n}(F)$ and $B \in M_{n,k}(F)$, then*

$$\operatorname{rank}(AB) \leq \operatorname{rank}(A) \quad \textit{and} \quad \operatorname{rank}(AB) \leq \operatorname{rank}(B).$$

Proof. Let $T_A \colon F^n \to F^m$ be the linear map given by matrix multiplication (on the left) by A, and let $T_B \colon F^k \to F^n$ be the linear map given by matrix multiplication (on the left) by B. Then

$$\begin{aligned} \operatorname{rank}(AB) &= \operatorname{rank}(T_A \circ T_B) && \text{(by definition)} \\ &\leq \dim(\operatorname{Im} T_A) && \text{(by Exercise 3.5.5(b))} \\ &= \operatorname{rank}(A) \end{aligned}$$

and

$$\begin{aligned} \operatorname{rank}(AB) &= \operatorname{rank}(T_A \circ T_B) && \text{(by definition)} \\ &\leq \dim(\operatorname{Im} T_B) && \text{(by Exercise 3.5.5(b))} \\ &= \operatorname{rank}(B). && \square \end{aligned}$$

Corollary 4.4.5. *Let $A \in M_{m,n}(F)$. If P and Q are invertible $m \times m$ and $n \times n$ matrices, respectively, then*

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

Proof. By Lemma 4.4.4, we have

$$\operatorname{rank}(PA) \leq \operatorname{rank}(A)$$

and

$$\operatorname{rank}(A) = \operatorname{rank}\left(P^{-1}PA\right) \leq \operatorname{rank}(PA).$$

Hence $\operatorname{rank}(PA) = \operatorname{rank}(A)$. The proof that $\operatorname{rank}(AQ) = \operatorname{rank}(A)$ is similar. □

Lemma 4.4.6. *Let $T \colon V \to W$ be a linear map between finite-dimensional vector spaces. Let B be an ordered basis of V and let D be an ordered basis of W. Then*

$$\operatorname{rank}(T) = \operatorname{rank}\left([T]_B^D\right).$$

Proof. Let $n = \dim V$ and $m = \dim W$. Let $S_V \colon F^n \to V$ be the isomorphism that sends the standard basis of F^n to the basis B, and let $S_W \colon W \to F^m$ be the isomorphism that sends the basis D to the standard basis of F. Let

$$U = S_W \circ T \circ S_V \colon F^n \to F^m$$

be the composition. Then $[T]_B^D = [U]$ (where $[U]$ denotes the matrix of U with respect to the standard bases). Thus

$$\begin{aligned} \operatorname{rank}(T) &= \operatorname{rank}(S_W^{-1} \circ U \circ S_V^{-1}) && (\text{since } U = S_W \circ T \circ S_V) \\ &= \operatorname{rank}(U) && (\text{by Corollary 4.4.5}) \\ &= \operatorname{rank}([U]) && (\text{by Definition 4.4.1}) \\ &= \operatorname{rank}([T]_B^D). && \square \end{aligned}$$

Lemma 4.4.6 tells us that to compute the rank of any linear transformation, you can compute the rank of its matrix in *any* basis. So we reduce the problem to matrix calculations you learned in MAT 1341.

Lemma 4.4.7. *The rank of a matrix A is equal to the dimension of the column space of A.*

Proof. Let $A \in M_n(F)$ and let $T\colon F^n \to F^n$ be the matrix given by matrix multiplication (on the left). Then $T(e_j)$ is the j-th column of A. Thus we have

$$\operatorname{rank}(A) = \dim(\operatorname{Im}(T)) = \dim(\operatorname{Span}\{T(e_1), \ldots, T(e_m)\}) = \dim(\operatorname{col}(A)). \qquad \square$$

Lemma 4.4.8. *Suppose $A \in M_{m,n}(F)$ has row-echelon form R and block matrix D from Theorem 4.3.6. Then*

$$\operatorname{rank}(A) = \operatorname{rank}(R) = \operatorname{rank}(D) = r.$$

Proof. The proof of this lemma is left as Exercise 4.4.1. $\square$

Recall that the *row space* of a matrix is the span of its rows.

Lemma 4.4.9. *Suppose $A \in M_{m,n}(F)$.*

(a) We have $\operatorname{rank}(A) = \operatorname{rank}(A^t)$.

(b) The rank of A is equal to the dimension of the row space of A.

(c) The row space and column space of A have the same dimension.

Proof. (a) Let D be the matrix from Theorem 4.3.6. Thus we have elementary matrices $E_1, \ldots, E_p$ and $F_1, \ldots, F_q$ such that

$$E_1 \cdots E_p A F_1 \cdots F_q = D.$$

Note that $\operatorname{rank}(D^t) = \operatorname{rank}(D)$, since the size of the identity matrix block in D is the same as the size of the identity matrix block in D^t. Thus we have

$$\begin{aligned} \operatorname{rank}(A) &= \operatorname{rank}(D) && (\text{by Lemma 4.4.8}) \\ &= \operatorname{rank}(D^t) \\ &= \operatorname{rank}(F_q^t \cdots F_1^t A^t E_p^t \cdots E_1^t) \\ &= \operatorname{rank}(A^t). && (\text{by Corollary 4.4.5}) \end{aligned}$$

(b) The row space of A is equal to the column space of A^t. Thus, the result follows from part (a) and Lemma 4.4.7.

(c) This follows from part (b) and Lemma 4.4.7.

$\square$

Exercises.

4.4.1. Prove Lemma 4.4.8.

4.4.2. Prove that similar matrices have the same rank.

4.4.3. Suppose A and B are $n \times n$ matrices. Prove that, if AB is invertible, then A and B are both invertible. Do not use determinants, since we have not seen them yet. *Hint*: Use Lemma 4.4.4.

4.4.4. Suppose A is an $n \times m$ matrix with entries in F. Recall (from MAT 1341), that the *null space* of A is

$$\operatorname{Ker} A = \{x \in F^m \mid Ax = \mathbf{0}\}.$$

Prove that $\operatorname{rank}(A) + \dim(\operatorname{Ker} A) = n$.

4.4.5. Is it possible for a matrix $A = M_3(\mathbb{R})$ to have column space spanned by $(1, 1, 1)$ and null space spanned by $(3, -1, 2)$?

Chapter 5

Determinants and multlinear maps

You saw determinants of matrices in MAT 1341, and learned how to compute them. We now revisit this topic in more depth. In particular, we explain *why* the determinant is defined the way it is (or, equivalently, can be computed in the way you learned) by proving that the determinant is the *only* function on matrices satisfying three very natural properties. The material in this section roughly corresponds to [Tre, Ch. 3].

We will assume in this chapter that $2 \neq 0$ in the field F. In particular, we assume $F \neq \mathbb{F}_2$. We do this so that we can use the argument that, for $a \in F$, the equation $a = -a$ implies that $a = 0$. Note that the proof of this goes

$$a = -a \implies 2a = 0 \implies \frac{1}{2} \cdot 2a = \frac{1}{2} \cdot 0 \implies a = 0.$$

But $\frac{1}{2}$ only exists if $2 \neq 0$. There are ways to avoid the assumption $2 \neq 0$ by slightly changing the definitions in this chapter. But, for simplicity, we make this assumption. Of course, there are no issues if F is $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$.

5.1 Multilinear maps

Suppose $V_1, \ldots, V_n$ and W are vector spaces over a field F. A *multilinear map*

$$f \colon V_1 \times \cdots \times V_n \to W$$

is a map that is linear separately in each variable. In other words, for each $i = 1, \ldots, n$, if all the variables other than v_i are held constant, then $f(v_1, \ldots, v_n)$ is a linear function of v_i. If $n = 2$, then we use the term *bilinear* instead of multilinear.

Examples 5.1.1. (a) If $n = 1$, then $f \colon V_1 \to W$ is linear if and only if it is multilinear.

(b) Suppose U, V, W are vector spaces over a field F. The map

$$f \colon \mathcal{L}(V, W) \times \mathcal{L}(U, V) \to \mathcal{L}(U, W), \quad f(S, T) = ST,$$

is bilinear. To see this, we check that, for

$$c, d \in F, \quad S, S_1, S_2 \in \mathcal{L}(V, W), \quad \text{and} \quad T, T_1, T_2 \in \mathcal{L}(U, V),$$

we have

$$f(cS_1 + dS_2, T) = (cS_1 + dS_2)T = cS_1T + dS_2T = cf(S_1, T) + df(S_2, T) \quad \text{and}$$
$$f(S, cT_1 + dT_2) = S(cT_1 + dT_2) = cST_1 + dST_2 = cf(S, T_1) + df(S, T_2).$$

(c) The map

$$f\colon \underbrace{V \times \cdots \times V}_{n \text{ factors}} \to V, \quad f(v_1, \ldots, v_n) = v_1 + \cdots + v_n,$$

is linear, but not multilinear unless $n = 1$ or V is the zero vector space.

To see that f is linear, we check that, for all scalars c, d and vectors

$$v_1, \ldots, v_n, w_1, \ldots, w_n \in V,$$

we have

$$\begin{aligned} f\big(c(v_1, \ldots, v_n) + d(w_1, \ldots, w_n)\big) &= f(cv_1 + dw_1, \ldots, cv_n + dw_n) \\ &= cv_1 + dw_1 + \cdots + cv_n + dw_n \\ &= c(v_1 + \cdots + v_n) + d(w_1 + \cdots + w_n) \\ &= cf(v_1, \ldots, v_n) + df(w_1, \ldots, w_n). \end{aligned}$$

To see that f is not multilinear in general, suppose that $n \geq 2$ and V is not the zero vector space. Let v be a nonzero vector in V. Then

$$f(v, \mathbf{0}, \ldots, \mathbf{0}) = v + \mathbf{0} + \cdots + \mathbf{0} = v \neq \mathbf{0}.$$

But if f were multilinear we would have

$$f(v, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}) = f(v, \mathbf{0}, \ldots, \mathbf{0}, 0 \cdot \mathbf{0}) = 0f(v, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}) = \mathbf{0}.$$

This is a contradiction.

(d) The zero map $V_1 \times \cdots \times V_n \to W$ is multilinear.

A multlinear map

$$f\colon \underbrace{V \times \cdots \times V}_{n \text{ factors}} \to F$$

is called an *n-linear form*. If $n = 2$, this is called a *bilinear form*.

Example 5.1.2. For every bilinear form $f\colon V \times V \to F$, we have

$$f(c_1v_1 + c_2v_2, u) = c_1f(v_1, u) + c_2f(v_2, u) \quad \text{and}$$
$$f(u, c_1v_1 + c_2v_2) = c_1f(u, v_1) + c_2f(u, v_2)$$

for all $c_1, c_2 \in F$ and $u, v_1, v_2 \in V$.

Each row of a matrix $A \in M_n(F)$ can be viewed as an element of the vector space F^n. Therefore, as a vector space, we can identify $M_n(F)$ with

$$\underbrace{F^n \times \cdots \times F^n}_{n \text{ factors}},$$

where the first factor of F^n corresponds to the first row of the matrix, etc. Then a map $f\colon M_n(F) \to F$ corresponds to a map

$$f\colon \underbrace{F^n \times \cdots \times F^n}_{n \text{ factors}} \to F,$$

and it makes sense to ask whether or not this map is n-linear.

Examples 5.1.3. (a) Suppose that, for each $i = 1, \ldots, n$, $f_i\colon V_i \to F$ is a linear form. Then

$$f\colon V_1 \times \cdots \times V_n \to F, \quad f(v_1, \ldots, v_n) = f_1(v_1) \cdots f_n(v_n),$$

is an n-linear form. See Exercise 5.1.2.

(b) The function

$$f\colon M_n(F) \to F, \quad f(A) = a_{1,1}a_{2,2} \cdots a_{n,n}, \quad A = (a_{i,j}),$$

is an n-linear form. However, if $n \geq 2$, it is *not* a linear form. See Exercise 5.1.2.

(c) The function

$$\operatorname{tr}\colon M_n(F) \to F, \quad \operatorname{tr}(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}, \quad A = (a_{i,j}),$$

is *not* an n-linear form. However, it is a linear form. The value $\operatorname{tr}(A)$ is called the *trace* of A. See Exercise 5.1.2.

As the above examples illustrate, a map can be linear but not multilinear and vice versa. In fact, when $n \geq 2$, a map is only both linear and multilinear when it is the zero map. See Exercise 5.1.5.

Exercises.

5.1.1. Prove that the cross product map

$$f\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3, \quad f(u, v) = u \times v$$

is bilinear.

5.1.2. Prove the statements made in Examples 5.1.3.

5.1.3. Suppose that

$$f\colon V_1 \times \cdots \times V_n \to W$$

is a multilinear map. Prove that $f(v_1, \ldots, v_n) = 0$ whenever $v_i = 0$ for some $1 \leq i \leq n$.

5.1.4. Let V be a vector space with dual space V^*. Show that the map

$$f\colon V \times V^* \to F, \quad f(v, \varphi) = \varphi(v),$$

is bilinear.

5.1.5. Suppose V_1, V_2, W are vector spaces over F. Prove that

$$f\colon V_1 \times V_2 \to W$$

is the zero map if and only if f is both linear and bilinear.

5.2 The determinant

Given $A \in M_n(F)$ for $n \geq 2$, we let $A_{i,j}$ denote the matrix obtained from A by deleting the i-th row and j-th column. We will denote (i, j) entry of the matrix A by $a_{i,j}$.

Definition 5.2.1 (Determinant)**.** We define a function

$$\det\colon M_n(F) \to F$$

recursively (with respect to n) as follows. For $n = 1$ we define

$$\det(A) := a_{1,1}.$$

For $n \geq 2$, we define

$$\det(A) := \sum_{i=1}^{n} (-1)^{i+1} \det\left(A_{i,1}\right) \cdot a_{i,1}. \tag{5.1}$$

The function det is called the *determinant function*. The value $\det(A)$ is called the *determinant* of A, and is also denoted $\det A$ or $|A|$. The scalar $(-1)^{i+j} \det(A_{i,j})$ is called the *cofactor* of the entry of A in position (i, j). The formula (5.1) is called *cofactor expansion along the first column of* A.

Example 5.2.2. For 2×2 matrices, we have

$$\det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

Recall that I_n is the identity matrix.

Lemma 5.2.3. *We have* $\det(I_n) = 1$.

Proof. We prove the result by induction on n. Since

$$\det(I_1) = \det[1] = 1,$$

the result holds for $n = 1$.

Now assume that $\det(I_n) = 1$ for some $n \geq 1$. Noting that $(I_{n+1})_{1,1} = I_n$, we use (5.1) to compute

$$\det(I_{n+1}) = (-1)^2 \det(I_n) \cdot 1 + (-1)^3 \det(I_{n+1})_{2,1} \cdot 0 + \cdots + (-1)^{n+2} \det(I_{n+1})_{n+1,1} \cdot 0 = 1.$$

□

Recall from Section 5.1 that $\det\colon M_n(F) \to F$ can be viewed as a function

$$\det\colon \underbrace{F^n \times \cdots \times F^n}_{n \text{ factors}} \to F.$$

Theorem 5.2.4. *The determinant function* $\det\colon M_n(F) \to F$ *is* n*-linear.*

Proof. We prove the result by induction on n. The base case $n = 1$ is obvious.

Now assume that $n > 1$ and that $\det\colon M_{n-1}(F) \to F$ is $(n-1)$-linear. Let $A = (a_{ij}) \in M_n(F)$. Suppose that for the r-th row of A we have

$$(a_{r,1}, \ldots, a_{r,n}) = \ell(b_1, \ldots, b_n) + k(c_1, \ldots, c_n), \ \ell, k \in F.$$

In other words, we have

$$a_{r,i} = \ell b_i + k c_i, \ 1 \leq i \leq n.$$

Let B denote the matrix obtained from A by replacing the r-th row by $(b_1, \ldots, b_n)$ and let C denote the matrix obtained from A by replacing the r-th row by $(c_1, \ldots, c_n)$.

We have

$$\begin{aligned}
\det(A) &\overset{(5.1)}{=} \sum_{i=1}^{n} (-1)^{i+1} \det(A_{i,1}) \cdot a_{i,1} \\
&= \sum_{\substack{1 \leq i \leq n \\ i \neq r}} (-1)^{i+1} \det(A_{i,1}) \cdot a_{i,1} + (-1)^{r+1} \det(A_{r,1}) \cdot a_{r,1} \\
&= \sum_{\substack{1 \leq i \leq n \\ i \neq r}} (-1)^{i+1} \left(\ell \det(B_{i,1}) + k \det(C_{i,1})\right) \cdot a_{i,1} + (-1)^{r+1} \det(A_{r,1}) \cdot (\ell b_{r,1} + k c_{r,1}) \\
&= \ell \sum_{i=1}^{n} (-1)^{i+1} \det(B_{i,1}) \cdot b_{i,1} + k \sum_{i=1}^{n} (-1)^{i+1} \det(C_{i,1}) \cdot c_{i,1} \\
&\overset{(5.1)}{=} \ell \det(B) + k \det(C),
\end{aligned}$$

where we used the induction hypothesis in the third equality. This completes the proof of the induction step. □

Example 5.2.5. The determinant of a 2×2 matrix is a bilinear form

$$\det\colon F^2 \times F^2 \to F$$

given by

$$\det\big((x_1, x_2), (y_1, y_2)\big) = \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = x_1 y_2 - x_2 y_1.$$

The following lemma will be generalized in Lemma 5.3.1. However, our proof of Lemma 5.3.1 will depend indirectly on the following special case. Thus, we need to prove it independently to avoid circular logic.

Lemma 5.2.6. *Suppose the matrix B is obtained from $A \in M_n(F)$, $n \geq 2$, by interchanging two neighbouring rows. Then*

$$\det B = -\det A.$$

Proof. The proof of this lemma is left as Exercise 5.2.1. □

Lemma 5.2.7. *If a matrix $A \in M_n(F)$, $n \geq 2$, has two identical rows, then* $\det(A) = 0$.

Proof. By successively swapping neighbouring rows, we can turn A into a matrix B where the two identical rows are neighbouring. By Lemma 5.2.6, we have $\det(A) = \pm \det(B)$. Now, since swapping the two neighbouring identical rows of B leaves B unchanged, Lemma 5.2.6 tells us that

$$\det(B) = -\det(B).$$

Thus $\det(A) = \pm \det(B) = 0$. □

Exercises.

5.2.1. Prove Lemma 5.2.6. *Hint*: Use (5.1) and induction on n.

5.2.2. If A is an $n \times n$ matrix and c is a scalar, how are the determinants $\det(A)$ and $\det(cA)$ related?

5.2.3 ([Tre, Ex. 3.3.2]). How are the determinants of A and B related if

(a)
$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}, \quad B = \begin{bmatrix} 2a_1 & 3a_2 & 5a_3 \\ 2b_1 & 3b_2 & 5b_3 \\ 2c_1 & 3c_2 & 5c_3 \end{bmatrix};$$

(b)
$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}, \quad B = \begin{bmatrix} 3a_1 & 4a_2 + 5a_1 & 5a_3 \\ 3b_1 & 4b_2 + 5b_1 & 5b_3 \\ 3c_1 & 4c_2 + 5c_1 & 5c_3 \end{bmatrix}.$$

5.2.4 ([Tre, Ex. 3.3.4]). A square $(n \times n)$ matrix is called *skew-symmetric* (or *antisymmetric*) if $A^t = -A$. Prove that if A is skew-symmetric and n is odd, then $\det(A) = 0$. Is this true for even n?

5.2.5 ([Tre, Ex. 3.3.8]). Show that

$$\det \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix} = (c-a)(c-b)(b-a).$$

This is a particular case of the so-called *Vandermonde determinant*.

5.3 Characterizing properties of the determinant

From Theorem 5.2.4, Lemma 5.2.7, and Lemma 5.2.3, we see that the determinant has the following properties:

(D1) The function det is n-linear.

(D2) We have $\det(A) = 0$ whenever A has two identical rows.

(D3) We have $\det(I_n) = 1$.

We will soon show that these properties *uniquely characterize* the determinant. In other words, the determinant is the *only* map $M_n(F) \to F$ satisfying properties (D1)–(D3). To do this, we will prove some facts about the determinant using only properties (D1)–(D3). Therefore, these facts are true for any other map with these properties.

We begin by analyzing the effects of the elementary row operations on the determinant.

Lemma 5.3.1 (Determinant under row operation of type I)**.** *Suppose the matrix B is obtained from $A \in M_n(F)$ by interchanging two rows. Then*

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

In particular $\det(P_{i,j}) = -1$.

Proof. Suppose B is obtained from A by interchanging rows i and j. Let u and v be the i-th and j-th rows of A, respectively. Define the following matrices:

- C is the matrix whose i-th and j-th rows are both $u + v$,
- M is the matrix whose i-th and j-th rows are both u,
- N is the matrix whose i-th and j-th rows are both v.

Then we have

$$\begin{aligned} 0 &= \det(C) && \text{(by (D2))} \\ &= \det(A) + \det(B) + \det(M) + \det(N) && \text{(by (D1))} \\ &= \det(A) + \det(B) + 0 + 0. && \text{(by (D2))} \end{aligned}$$

Thus $\det(B) = -\det(A)$. Since $P_{i,j}$ is obtained from the identity matrix by interchanging the i-th and j-th rows, we have

$$\det(P_{i,j}) = -\det(I_n) = -1. \qquad \square$$

Lemma 5.3.2 (Determinant under row operation of type II)**.** *Assume that a matrix B is obtained from a matrix A by multiplying a row of A by a scalar $a \in F$. Then*

$$\det(B) = a \cdot \det(A).$$

In particular, $\det(M_i(a)) = a$*, for* $a \in F$*.*

Proof. The first assertion follows immediately from (D1). Since $M_i(a)$ is obtained from the identity matrix by multiplying the i-th row by a, we have

$$\det(M_i(a)) = a \det(I_n) = a \cdot 1 = a. \qquad \square$$

Lemma 5.3.3 (Determinant under row operation of type III)**.** *Let $a \in F$. Assume that a matrix B is obtained from a matrix A by adding a times the j-th row to the i-th row ($i \neq j$). Then*

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

In particular, $\det(E_{i,j}(a)) = 1$*.*

Proof. Let C be the matrix obtained from A by replacing its i-th row with the j-th row of A. In particular, C has two identical rows (namely, its i-th and j-th rows are the same). Then

$$\begin{aligned} \det(B) &= \det(A) + a \det(C) && \text{(by (D1))} \\ &= \det(A) + 0. && \text{(by (D2))} \end{aligned}$$

$\square$

Lemma 5.3.4. *If $A \in M_n(F)$ has $\operatorname{rank}(A) < n$, then $\det(A) = 0$.*

Proof. By Lemma 4.3.5, we can use row operations of type I and III to transform the matrix A into a matrix B in row-echelon form. By Lemmas 5.3.1 and 5.3.3, we have $\det(A) = \pm\det(B)$. If $\operatorname{rank}(A) < n$, then B has a zero row. Therefore, $\det(B) = 0$ by (D1) (see Exercise 5.1.3). $\square$

Theorem 5.3.5 (The determinant is multiplicative)**.** *For any $A, B \in M_n(F)$, we have*

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

Proof. First suppose that $\operatorname{rank}(A) < n$. Then $\operatorname{rank}(AB) \leq \operatorname{rank}(A) < n$ by Lemma 4.4.4. Hence

$$\det(A)\det(B) = 0 \cdot \det(B) = 0 = \det(AB),$$

and we're done.

Now suppose $\operatorname{rank}(A) = n$. Lemmas 5.3.1, 5.3.2, and 5.3.3 imply that the theorem holds whenever A is an elementary matrix. Since A is an invertible matrix (by Lemma 4.4.2), it can be written as a product $A = E_1 E_2 \cdots E_k$ of elementary matrices by Lemma 4.4.3. Then

$$\begin{aligned}
\det(AB) &= \det(E_1 E_2 \cdots E_k B) \\
&= \det(E_1)\det(E_2 \cdots E_k B) \\
&\vdots \\
&= \det(E_1)\det(E_2)\cdots\det(E_k)\det(B) \\
&= \det(E_1 E_2)\det(E_3)\cdots\det(E_k)\det(B) \\
&\vdots \\
&= \det(E_1 E_2 \cdots E_k)\det(B) \\
&= \det(A)\det(B).
\end{aligned}$$

□

The following theorem states that the properties (D1)–(D3) uniquely characterize the determinant.

Theorem 5.3.6 (Characterization of the determinant)**.** *Suppose $\delta\colon M_n(F) \to F$ is a function such that*

(a) δ is n-linear,

(b) $\delta(A) = 0$ whenever A has two identical rows,

(c) $\delta(I_n) = 1$.

(In other words, suppose δ has properties (D1)–(D3).) Then $\delta(A) = \det(A)$ for all $A \in M_n(F)$.

Proof. First note that the results of Lemmas 5.3.1, 5.3.2, 5.3.3, 5.3.4, and Theorem 5.3.5 hold for δ, since their proofs only used properties (D1)–(D3).

First, suppose A is not invertible (i.e. $\operatorname{rank}(A) < n$). Then, by Lemma 5.3.4 (and its analogue for δ), we have

$$\det(A) = 0 = \delta(A).$$

Now suppose A is invertible. Then we can write A as a product $A = E_1 E_2 \cdots E_k$ of elementary matrices. Then we have

$$\det(A) = \det(E_1)\cdots\det(E_k) = \delta(E_1)\cdots\det(E_k) = \delta(E_1 \cdots E_k) = \delta(A).$$

□

Exercises.

5.3.1 ([Tre, Ex. 3.3.5]). A square matrix A is called *nilpotent* if $A^k = \mathbf{0}$ for some positive integer k. Show that if A is a nilpotent matrix, then $\det(A) = 0$.

5.3.2. Prove that if A and B are similar matrices, then $\det(A) = \det(B)$.

5.4 Other properties of the determinant

We conclude this chapter by proving a few more useful properties of the determinant.

Lemma 5.4.1. *A matrix A is invertible if and only if $\det(A) \neq 0$. Furthermore, if A is invertible, then*

$$\det\left(A^{-1}\right) = \det(A)^{-1}.$$

Proof. If $A \in M_n(F)$ is not invertible, then $\operatorname{rank}(A) < n$, and so $\det(A) = 0$ by Lemma 5.3.4.

Now suppose A is invertible. Then we have

$$\det(A)\det\left(A^{-1}\right) = \det\left(AA^{-1}\right) = \det(I_n) = 1.$$

Thus $\det(A) \neq 0$ and $\det\left(A^{-1}\right) = \det(A)^{-1}$. □

Lemma 5.4.2. *If $A \in M_n(F)$, then*

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

Proof. If A is not invertible, then $\operatorname{rank}(A) = \operatorname{rank}(A^t) < n$ by Lemmas 4.4.9(a) and 4.4.2. Thus, by Lemma 5.3.4, $\det(A) = 0 = \det(A^t)$.

Now suppose A is invertible. Then we can write A as a product $A = E_1E_2\cdots E_k$ of elementary matrices by Lemma 4.4.3. Note that $\det(E) = \det(E^t)$ for any elementary matrix E (Exercise 5.4.1). Thus,

$$\begin{aligned}\det(A^t) &= \det(E_k^t \cdots E_2^t E_1^t)\\ &= \det(E_k^t)\cdots\det(E_2^t)\det(E_1^t)\\ &= \det(E_1)\det(E_2)\cdots\det(E_k)\\ &= \det(A).\end{aligned}$$

□

Lemma 5.4.3. *If A is a triangular matrix, then*

$$\det(A) = a_{1,1}a_{2,2}\cdots a_{n,n}$$

is the product of the elements on the diagonal of A.

Proof. The proof of the case where A is upper triangular is left as Exercise 5.4.3. Then the lower triangular case follows from Lemma 5.4.2. □

Exercises.

5.4.1. Verify directly (i.e. without using Lemma 5.4.2) that $\det(E) = \det(E^t)$ for any elementary matrix E.

5.4.2. A real square matrix Q is called *orthogonal* if $Q^tQ = I$ (see Definition 6.2.6). Prove that is Q is an orthogonal matrix, then $\det(Q) = \pm 1$.

5.4.3. Prove Lemma 5.4.3 in the case that A is upper triangular. *Hint*: Use (5.1) and induction on n.

5.4.4. (a) Suppose that the matrix B is obtained from $A \in M_n(F)$ by interchanging two columns. Prove that $\det(B) = -\det(A)$.

(b) Suppose that the matrix B is obtained from $A \in M_n(F)$ by multiplying a column of A by a scalar a. Prove that $\det(B) = a \cdot \det(A)$.

(c) Suppose that a is a scalar and that the matrix B is obtained from $A \in M_n(F)$ by adding a times the j-th column to the i-th column. Prove that $\det(B) = \det(A)$.

5.4.5. Prove that the determinant can be computed by cofactor expansion along any row or column. More precisely, if $A \in M_n(F)$ with $n \geq 2$, prove that

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} \det(A_{i,j}) \cdot a_{i,j} \quad \text{for all } 1 \leq j \leq n, \tag{5.2}$$

and that

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} \det(A_{i,j}) \cdot a_{i,j} \quad \text{for all } 1 \leq i \leq n. \tag{5.3}$$

Hint: To prove (5.2), let B be the matrix obtained from A by interchanging the first and j-th columns. Relate $\det(A)$ to $\det(B)$ and compute $\det(B)$ using (5.1).

5.4.6 ([Tre, Ex. 3.3.9]). Let A be a square matrix. Show that the block triangular matrices

$$\begin{bmatrix} I & M \\ \mathbf{0} & A \end{bmatrix}, \quad \begin{bmatrix} A & M \\ \mathbf{0} & I \end{bmatrix}, \quad \begin{bmatrix} I & \mathbf{0} \\ M & A \end{bmatrix}, \quad \begin{bmatrix} A & \mathbf{0} \\ M & I \end{bmatrix}$$

all have determinant equal to $\det(A)$. Here M is an arbitrary matrix.

5.4.7 ([Tre, Ex. 3.3.10]). Use Exercise 5.4.6 to show that, if A and C are square matrices, then

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

Here B is an arbitrary matrix. *Hint*:

$$\begin{bmatrix} A & B \\ \mathbf{0} & C \end{bmatrix} = \begin{bmatrix} I & B \\ \mathbf{0} & C \end{bmatrix} \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix}.$$

5.4.8 ([Tre, Ex. 3.3.11]). Let A be an $m \times n$ matrix, and let B be an $n \times m$ matrix. Prove that

$$\det \begin{bmatrix} \mathbf{0} & A \\ -B & I \end{bmatrix} = \det(AB).$$

Hint: While it is possible to transform the matrix by row operations to a form where the determinant is easy to compute, the easiest way is to multiply on the right by

$$\begin{bmatrix} I & \mathbf{0} \\ B & I \end{bmatrix}.$$

5.4.9 ([Tre, Ex. 3.5.5]). Let D_n be the determinant of the matrix

$$\begin{bmatrix} 1 & -1 & & & \\ 1 & 1 & -1 & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & 1 & 1 \end{bmatrix},$$

where all the unspecified entries are zero. Using cofactor expansion, show that $D_n = D_{n-1} + D_{n-2}$. This (together with a computation of D_1 and D_2) implies that D_n is the n-th Fibonacci number.

Chapter 6

Inner product spaces

In this chapter we look at vector spaces with additional structure. That additional structure is the notion of an *inner product*, which can be thought of as a generalization of the dot product (for $\mathbb{R}^n$) to arbitrary vector spaces. In particular, when one has an inner product, one can speak of the *length* of a vector. The material in this chapter roughly corresponds to [Tre, Ch. 5].

6.1 Definitions

Definition 6.1.1 (Inner product space)**.** A *(real) inner product space* is a vector space over $\mathbb{R}$ together with a map

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}, \quad (x, y) \to \langle x, y \rangle$$

satisfying

(a) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0 \iff v = 0$ for all $v \in V$,

(b) $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$, and

(c) $\langle c_1 v_1 + c_2 v_2, w \rangle = c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle$ for all $v_1, v_2, w \in V$ and $c_1, c_2 \in \mathbb{R}$.

The real number $\langle v, w \rangle$ is called the *inner product of v and w*.

Remark 6.1.2. Properties (b) and (c) imply that

$$\langle v, c_1 w_1 + c_2 w_2 \rangle = c_1 \langle v, w_1 \rangle + c_2 \langle v, w_2 \rangle$$

for all $v, w_1, w_2 \in V$ and $c_1, c_2 \in \mathbb{R}$. So $\langle \cdot, \cdot \rangle$ is bilinear.

Example 6.1.3. On $\mathbb{R}^n$,

$$\begin{aligned} \langle x, y \rangle &:= x \cdot y \qquad \text{(dot product)} \\ &= x^t y \qquad \text{(matrix product)} \end{aligned}$$

is an inner product, called the *canonical inner product* on $\mathbb{R}^n$.

Example 6.1.4. On $C[a,b] := \{f\colon [a,b] \to \mathbb{R} \mid f \text{ is continuous}\}$,

$$\langle f, g\rangle := \int_a^b f(t)g(t)\,dt$$

is an inner product. Property (a) is the only difficult property to check. It relies on the fact that if h is a continuous real-valued function on $[a,b]$ and $h(t) \geq 0$ for all $t \in [a,b]$, then

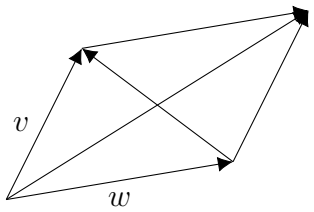
$$\int_a^b h(t)\,dt = 0 \iff h(t) = 0 \;\forall\; t \in [a,b].$$

Definition 6.1.5 (Norm)**.** If $(V, \langle\cdot,\cdot\rangle)$ is an inner product space, then the *norm* of $v \in V$ is defined to be

$$\|v\| := \sqrt{\langle v, v\rangle}.$$

Theorem 6.1.6. *If V is an inner product space, then*

(a) $\|v\| = 0 \iff v = 0$ *for all* $v \in V$,
(b) $\|cv\| = |c|\,\|v\|$ *for all* $c \in \mathbb{R}$ *and* $v \in V$,
(c) $\langle v, w\rangle = \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right)$ *for all* $v, w \in V$ *(*polarization*), and*
(d) $\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2$ *for all* $v, w \in V$ *(*parallelogram law*).*



Proof. The proof of these statements is left as Exercise 6.1.1. □

Theorem 6.1.7 (Cauchy-Schwartz Inequality)**.** *If $(V, \langle\cdot,\cdot\rangle)$ is an inner product space, then*

$$|\langle v, w\rangle| \leq \|v\|\,\|w\| \quad \forall\; v, w, \in V,$$

and equality holds above if and only if $\{v, w\}$ is linearly dependent.

Proof. Let

$$p(t) = \langle v + tw, v + tw\rangle, \quad t \in \mathbb{R}.$$

Then $p(t) \geq 0$ for all $t \in \mathbb{R}$. Now

$$p(t) = \langle v, v\rangle + t\langle v, w\rangle + t\langle w, v\rangle + t^2\langle w, w\rangle = \|v\|^2 + 2t\langle v, w\rangle + t^2\|w\|^2,$$

and so p is a quadratic polynomial. Since $p(t) \geq 0$ for all $t \in \mathbb{R}$, it can have at most one root. Recall that if a polynomial $p(t) = at^2 + bt + c$ has at most one root, then $b^2 - 4ac \leq 0$ (with equality if and only if p has exactly one root). Thus

$$b^2 - 4ac = 4\langle v, w\rangle^2 - 4\|v\|^2\|w\|^2 \leq 0 \implies |\langle v, w\rangle| \leq \|v\|\|w\|.$$

Equality holds it and only if there exists a *unique* $t_0 \in \mathbb{R}$ such that

$$0 = p(t_0) = \langle v + t_0 w, v + t_0 w \rangle,$$

which implies that $v + t_0 w = 0$, and so $\{v, w\}$ is linearly dependent. □

Example 6.1.8. In the setting of Example 6.1.4, we have

$$\left| \int_a^b f(t)g(t)\,dt \right| \leq \sqrt{\int_a^b f(t)^2\,dt \cdot \int_a^b g(t)^2\,dt}.$$

This is not an obvious fact at all. (Try proving it directly using methods from calculus—it's hard!)

Corollary 6.1.9 (Triangle inequality)**.** *If V is an inner product space, then for all $v, w \in V$,*

$$\|v + w\| \leq \|v\| + \|w\|.$$

This is called the triangle inequality*.*

Proof. We have

$$\|v+w\|^2 = \langle v+w, v+w \rangle = \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \leq \|v\|^2 + 2\|v\|\,\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.$$

Taking square roots gives the triangle inequality. □

Exercises.

6.1.1. Prove Theorem 6.1.6. *Hint:* Replace the norms by inner products. For the last two parts, expand both sides of each equation.

6.1.2 ([Ber14, 5.1.1]). Show that if v and w are vectors in an inner product space such that

$$\|v\|^2 = \|w\|^2 = \langle v, w \rangle,$$

then $v = w$.

6.1.3. For vectors u, v in an inner product space, show that

$$\langle u, v \rangle = 0 \iff \|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

6.1.4 ([Ber14, 5.1.3]). For nonzero vectors v, w in an inner product space, prove that $\|v+w\| = \|v\| + \|w\|$ if and only if $v = cw$ for some $c > 0$. *Hint*: Inspect the proof of Corollary 6.1.9, then apply the second assertion of Theorem 6.1.7.

6.1.5 ([Ber14, Ex. 5.1.6]). Suppose that $r_1, \dots, r_n$ are positive real numbers. Prove that

$$\langle x, y \rangle := \sum_{i=1}^{n} r_i a_i b_i,$$

for $x = (a_1, \dots, a_n)$, $y = (b_1, \dots, b_n)$ defines an inner product on $\mathbb{R}^n$ (different from the canonical inner product, unless $r_1 = r_2 = \cdots = r_n = 1$).

6.1.6 ([Ber14, Ex. 5.1.7]). In the inner product space of Example 6.1.4, prove that if f and g are continuously differentiable (that is, the derivatives f' and g' exist and are differentiable), then

$$\langle f, g' \rangle + \langle f', g \rangle = f(b)g(b) - f(a)g(a).$$

Hint: Integration by parts.

6.1.7 ([Ber14, Ex. 5.1.9]). Suppose V is an inner product space and $T\colon V \to V$ is a linear map. Prove that

$$\langle Tu, Tv \rangle = \frac{1}{4}\left(\|T(u+v)\|^2 - \|T(u-v)\|^2\right)$$

for all vectors u and v. *Hint*: Use Theorem 6.1.6.

6.2 Orthogonality

Definition 6.2.1 (Orthogonal and orthonormal)**.** If V is an inner product space and $v, w \in V$, we say v and w are *orthogonal*, and write $v \perp w$, if $\langle v, w \rangle = 0$. We say a set $\{v_1, \dots, v_n\}$ of vectors in V is orthogonal if

$$\langle v_i, v_j \rangle = \begin{cases} 0, & i \neq j, \\ \|v_i\|^2 \neq 0, & i = j. \end{cases}$$

In other words, they are nonzero and pairwise orthogonal. We say the set $\{u_1, \dots, u_m\}$ is *orthonormal* if it is orthogonal and $\|u_i\| = 1$ for all $i = 1, \dots, m$ (i.e. each u_i is a *unit vector*).

Theorem 6.2.2. *If $\{v_1, \dots, v_n\}$ is orthogonal, then it is linearly independent.*

Proof. Suppose

$$c_1 v_1 + \cdots + c_n v_n = \mathbf{0}$$

for some scalars $c_1, \dots, c_n$. Then for each $j = 1, \dots, n$,

$$0 = \langle \mathbf{0}, v_j \rangle = \left\langle \sum_{i=1}^{n} c_i v_i, v_j \right\rangle = \sum_{i=1}^{n} c_i \langle v_i, v_j \rangle = c_j \|v_j\|^2.$$

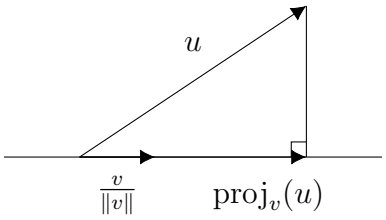
Since $\|v_j\|^2 \neq 0$, this implies $c_j = 0$. Since this is true for all $j = 1, \dots, n$, we see that the set is linearly independent. □

Definition 6.2.3 (Projection)**.** Suppose V is an inner product space. For $v \in V$, $v \neq 0$, define the *projection* map $\text{proj}_v \colon V \to V$ by

$$\text{proj}_v\, u = \frac{\langle u, v\rangle}{\|v\|^2} v.$$

Note

$$\text{proj}_v\, u = \left\langle u, \frac{v}{\|v\|} \right\rangle \frac{v}{\|v\|}.$$



Theorem 6.2.4. *If U is a finite-dimensional subspace of an inner product space V, then U has an orthogonal basis (hence an orthonormal basis as well).*

The proof of this theorem involves a process called the *Gram-Schmidt algorithm*.

Proof. Let $\{v_1, \ldots, v_k\}$ be any basis of U. We produce an orthogonal basis $\{w_1, \ldots, w_k\}$ of U as follows.

$$\begin{aligned}
w_1 &= v_1 \\
w_2 &= v_2 - \text{proj}_{w_1} v_2 \\
w_3 &= v_3 - \text{proj}_{w_1} v_3 - \text{proj}_{w_2} v_3 \\
&\vdots \\
w_k &= v_k - \sum_{l=1}^{k-1} \text{proj}_{w_l} v_k.
\end{aligned}$$

We claim that $\{w_1, \ldots, w_k\}$ is orthogonal. Then it follows from Theorem 6.2.2 that it is linearly independent, and hence a basis of U (since U has dimension k).

We prove the claim by induction. For $1 \leq n \leq k$, let $P(n)$ be the assertion that

(a) $\text{Span}\{w_1, \ldots, w_n\} = \text{Span}\{v_1, \ldots, v_n\}$, and

(b) $\{w_1, \ldots, w_n\}$ is orthogonal.

We know that $P(1)$ is true because $\text{Span}\{w_1\} = \text{Span}\{v_1\}$ (since $w_1 = v_1$) and $\{w_1\} = \{v_1\}$ is orthogonal since $w_1 = v_1 \neq 0$.

Now we show that, for $1 \leq n < k$, $P(n) \implies P(n+1)$. So assume $P(n)$ is true. In other words, $\text{Span}\{w_1, \ldots, w_n\} = \text{Span}\{v_1, \ldots, v_n\}$, and $\{w_1, \ldots, w_n\}$ are orthogonal. Then, for $1 \leq i \leq n$, we have

$$\langle w_{n+1}, w_i\rangle = \langle v_{n+1}, w_i\rangle - \left\langle \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l\rangle}{\|w_l\|^2} w_l, w_i \right\rangle$$

$$\begin{aligned}
&= \langle v_{n+1}, w_i\rangle - \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l\rangle}{\|w_l\|^2}\langle w_l, w_i\rangle \\
&= \langle v_{n+1}, w_i\rangle - \frac{\langle v_{n+1}, w_i\rangle}{\|w_i\|^2}\langle w_i, w_i\rangle \\
&= 0.
\end{aligned}$$

Moreover, if $w_{n+1} = 0$, then

$$v_{n+1} = \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l\rangle}{\|w_l\|^2} w_l \in \operatorname{Span}\{w_1, \ldots, w_n\},$$

and so, by the induction hypothesis $P(n)$(a), $v_{n+1} \in \operatorname{Span}\{v_1, \ldots, v_n\}$. But this is impossible since $\{v_1, \ldots, v_{n+1}\}$ is linearly independent. Thus $w_{n+1} \neq 0$ and so $\{w_1, \ldots, w_{n+1}\}$ is orthogonal. So $P(n+1)$(b) holds.

Now,

$$w_{n+1} \in \operatorname{Span}\{w_1, \ldots, w_n, v_{n+1}\} = \operatorname{Span}\{v_1, \ldots, v_{n+1}\},$$

since $P(n)$(a) holds, so

$$\operatorname{Span}\{w_1, \ldots, w_{n+1}\} \subseteq \operatorname{Span}\{v_1, \ldots, v_{n+1}\}.$$

Moreover,

$$v_{n+1} = w_{n+1} + \sum_{l=1}^{n} \operatorname{proj}_{w_l} v_{n+1} \in \operatorname{Span}\{w_1, \ldots, w_n, w_{n+1}\},$$

so that

$$\operatorname{Span}\{v_1, \ldots, v_{n+1}\} \subseteq \operatorname{Span}\{w_1, \ldots, w_{n+1}\}.$$

Thus $P(n+1)$(a) also holds. This completes the proof by induction.

We can then form an orthonormal basis

$$\left\{ \frac{w_1}{\|w_1\|}, \ldots, \frac{w_k}{\|w_k\|} \right\}. \qquad \square$$

Example 6.2.5. Take $V = \mathcal{P}_2(\mathbb{R})$ with the inner product

$$\langle p, q\rangle = \int_{-1}^{1} p(t)q(t)\,dt, \quad p, q \in \mathcal{P}_2(\mathbb{R}).$$

Note that, after restricting the domain of the polynomial functions to $[-1, 1]$, $\mathcal{P}_2(\mathbb{R})$ is a subspace of the inner product space $C[-1, 1]$ (since polynomials are continuous) and the above is the restriction of the inner product on $C[-1, 1]$ (see Example 6.1.4) to $\mathcal{P}_2(\mathbb{R})$.

Let $\{1, t, t^2\}$ be the standard basis of $\mathcal{P}_2(\mathbb{R})$. We'll find an orthonormal basis using the Gram-Schmidt algorithm. We have

$$\begin{aligned}
w_1 &= 1, \\
w_2 &= t - \operatorname{proj}_1 t = t - \frac{\langle t, 1\rangle}{\|1\|^2} \cdot 1
\end{aligned}$$

Now,

$$\langle t, 1\rangle = \int_{-1}^{1} t \cdot 1 \, dt = \frac{1}{2}t^2 \Big|_{-1}^{1} = \frac{1}{2} - \frac{1}{2} = 0.$$

Therefore,

$$\begin{aligned} w_2 &= t - 0 = t \\ w_3 &= t^2 - \frac{\langle t^2, 1\rangle}{\|1\|^2}1 - \frac{\langle t^2, t\rangle}{\|t\|^2}t. \end{aligned}$$

Since

$$\begin{aligned} \langle t^2, 1\rangle &= \int_{-1}^{1} t^2 \, dt = \frac{1}{3}t^3 \Big|_{-1}^{1} = \frac{1}{3} - \frac{-1}{3} = \frac{2}{3}, \\ \|1\|^2 = \langle 1, 1\rangle &= \int_{-1}^{1} 1 \cdot 1 \, dt = 2, \\ \langle t^2, t\rangle &= \int_{-1}^{1} t^3 \, dt = \frac{1}{4}t^4 \Big|_{-1}^{1} = 0. \end{aligned}$$

Thus,

$$w_3 = t^2 - \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = t^2 - \frac{1}{3}.$$

Hence,

$$\left\{1, t, t^2 - \frac{1}{3}\right\}$$

is an orthogonal basis of $\mathcal{P}_2(\mathbb{R})$.

Now, if we wanted an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$, we compute

$$\|1\| = \sqrt{2}, \quad \|t\|^2 = \int_{-1}^{1} t^2 \, dt = \frac{2}{3} \implies \|t\| = \frac{\sqrt{6}}{3},$$

and

$$\|w_3\|^2 = \int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)^2 dt = \int_{-1}^{1} \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9}\right) dt = \frac{1}{5}t^2 - \frac{2}{9}t^3 + \frac{1}{9}t \Big|_{-1}^{1} = \frac{2}{5} - \frac{4}{9} + \frac{2}{9} = \frac{8}{45}.$$

Let

$$u_1 = \frac{w_1}{\|w_1\|} = \frac{\sqrt{2}}{2}, \quad u_2 = \frac{w_2}{\|w_2\|} = \frac{\sqrt{6}}{2}t, \quad u_3 = \frac{w_3}{\|w_3\|} = \frac{3\sqrt{10}}{4}\left(t^2 - \frac{1}{3}\right).$$

Then $\langle u_i, u_j\rangle = \delta_{ij}$.

Orthogonal bases are practical, since if $\{u_1, \ldots, u_k\}$ is an orthogonal basis for U, and $v \in U$, then

$$v = \frac{\langle v, u_1\rangle}{\|u_1\|^2}u_1 + \frac{\langle v, u_2\rangle}{\|u_2\|^2}u_2 + \cdots + \frac{\langle v, u_k\rangle}{\|u_k\|^2}u_k.$$

(See Exercise 6.2.4.) In other words,

$$v = \sum_{i=1}^{k} c_i u_i, \quad \text{where} \quad c_i = \frac{\langle v, u_i \rangle}{\|u_i\|^2}$$

are the *Fourier coefficients* of v with respect to $u_1, \ldots, u_k$. This is convenient, since it saves us time. We don't need to solve a linear system for the coefficients $c_1, \ldots, c_n$ as we normally would need to. Computing Fourier coefficients is usually less work (once the orthogonal basis is known).

Definition 6.2.6 (Orthogonal matrix)**.** A matrix $A \in M_n(F)$ is called *orthogonal* if $A^t A = I$.

Lemma 6.2.7. *A matrix $A \in M_n(\mathbb{R})$ is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$ (with the dot product as the inner product).*

Proof. Let $A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ (i.e. v_i is the i-th column of A). Then

$$A^t A = \begin{bmatrix} v_1^t \\ \vdots \\ v_n^t \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

So the (i, j) entry of $A^t A$ is $v_i^t v_j$. Thus

$$A^t A = I_n \iff v_i^t v_j = \delta_{ij} \iff \{v_1, \ldots, v_n\} \text{ is orthonormal.}$$

Since F^n is n-dimensional and orthonormal sets are linearly independent, the result follows. □

Definition 6.2.8 (Orthogonal complement)**.** Let U be a subset of an inner product space V. Then we define

$$U^\perp := \{v \in V \mid \langle v, u \rangle = 0 \ \forall \ u \in U\}.$$

If U is a *subspace* (as opposed to just a subset), then $U^\perp$ is called the *orthogonal complement* to U. The notation $U^\perp$ is read "U perp".

Theorem 6.2.9. *Suppose V is an inner product space and U is a finite-dimensional subspace of V. Then*

(a) $U^\perp$ is also a subspace of V, and
(b) $V = U \oplus U^\perp$.

The proof will use the important idea of "duality" in inner product spaces.

Lemma 6.2.10. *Suppose V is an inner product space. Define $\varphi \colon V \to V^*$ by*

$$\varphi(v)(w) = \langle v, w \rangle.$$

If V is finite dimensional, then φ is an isomorphism.

Proof. Suppose $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in V$. Then for all $w \in V$,

$$\begin{aligned}\varphi(c_1v_1 + c_2v_2)(w) &= \langle c_1v_1 + c_2v_2, w\rangle \\ &= c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle \\ &= (c_1\varphi(v_1))(w) + c_2\varphi(v_2)(w).\end{aligned}$$

Thus

$$\varphi(c_1v_1 + c_2v_2) = c_1\varphi(v_1) + c_2\varphi(v_2),$$

and so φ is linear. Since $\dim V = \dim V^*$, it suffices to show that φ is injective. Suppose $\varphi(v) = 0$ for some $v \in V$. Then, in particular,

$$0 = \varphi(v)(v) = \langle v, v\rangle \|v\|^2,$$

and so $v = 0$. Hence φ is injective and therefore an isomorphism. □

Corollary 6.2.11 (Riesz Representation Theorem)**.** *If V is a finite-dimensional inner product space, then for all $f \in V^*$, there exists a unique $v \in V$ such that*

$$f(w) = \langle v, w\rangle, \quad \forall\ w \in V.$$

In other words, all linear forms are given by taking the inner product with some fixed vector of V.

Proof. This follows from the fact that φ in Lemma 6.2.10 is bijective. □

Proof of Theorem 6.2.9. The proof of part (a) is left as Exercise 6.2.1. We will prove part (b). So we need to show that $U \cap U^\perp = \{\mathbf{0}\}$ and $U + U^\perp = V$. Suppose $v \in U \cap U^\perp$. Then

$$\langle v, u\rangle = 0 \quad \forall\ u \in U,$$

since $v \in U^\perp$. But since we also have $v \in U$, we have $\langle v, v\rangle = 0$. Thus $v = \mathbf{0}$. So $U \cap U^\perp = \{\mathbf{0}\}$.

Now we show $U + U^\perp = V$. Let $v \in V$ and define

$$f\colon U \to \mathbb{R}, \quad f(u) = \langle v, u\rangle \quad \forall\ u \in U.$$

Then $f \in U^*$. Now, U is a finite-dimensional inner product space (we simply restrict the inner product on V to U). Therefore, by the Riesz Representation Theorem for U, there exists a $\tilde{u} \in U$ such that

$$f(u) = \langle \tilde{u}, u\rangle \quad \forall\ u \in U.$$

Therefore,

$$\langle \tilde{u}, u\rangle = \langle v, u\rangle\ \forall\ u \in U \implies \langle v - \tilde{u}, u\rangle = 0\ \forall\ u \in U \implies v - \tilde{u} \in U^\perp.$$

So

$$v = \tilde{u} + (v - \tilde{u}) \in U + U^\perp.$$

Since v was arbitrary, we have $V = U + U^\perp$. □

Corollary 6.2.12. *If U is a subspace of a finite-dimensional inner product space V, then*

$$\dim U^\perp = \dim V - \dim U.$$

Corollary 6.2.13. *Recall that if U is a finite-dimensional subspace of an inner product space V, then $V = U \oplus U^\perp$ (Theorem 6.2.9(b)). Define*

$$\operatorname{proj}_U \colon V \to V, \quad \operatorname{proj}_U(u+x) = u, \quad u \in U, \quad x \in U^\perp.$$

This map has the following properties:

(a) $\operatorname{Im} \operatorname{proj}_U = U$,

(b) $\operatorname{Ker} \operatorname{proj}_U = U^\perp$,

(c) $\langle v - \operatorname{proj}_U v, u\rangle = 0$ *for all* $v \in V$, $u \in U$,

(d) $\|v - \operatorname{proj}_U v\| \le \|v - u\|$ *for all* $u \in U$, *and equality holds if and only if* $u = \operatorname{proj}_U v$.

Proof. Statements (a) and (b) are easy. If $v = u + x = \operatorname{proj}_U v + x$ (for $u \in U$, $x \in U^\perp$), then $v - \operatorname{proj}_U v = x \in U^\perp$, so (c) holds.

It remains to show (d). Let $\tilde{u} = \operatorname{proj}_U v$, so that $v = \tilde{u} + x$, with $x \in U^\perp$. Now let u be *any* vector in U. Then

$$\begin{aligned} \|v-u\|^2 &= \|v - \tilde{u} + \tilde{u} - u\|^2 \\ &= \|x + (\tilde{u} - u)\|^2 \\ &= \|x\|^2 + 2\langle x, \tilde{u} - u\rangle + \|\tilde{u} - u\|^2 \\ &= \|x\|^2 + \|\tilde{u} - u\|^2. \end{aligned}$$

Thus

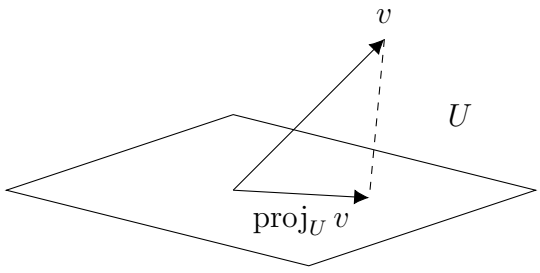
$$\|v-u\|^2 = \|v-\tilde{u}\|^2 + \|\tilde{u}-u\|^2$$

and so

$$\|v-u\|^2 \ge \|v-\tilde{u}\|^2.$$

Moreover, equality holds if and only if $\|\tilde{u} - u\|^2 = 0$, i.e. $u = \tilde{u} = \operatorname{proj}_U v$. □

The map proj_U is called *orthogonal projection* to U. The vector $\operatorname{proj}_U v$ is the "best approximation" to v by a vector in U.



Remark 6.2.14. If $\{u_1, \dots, u_k\}$ is any orthogonal basis for U, then an explicit formula for proj_U is

$$\operatorname{proj}_U v = \frac{\langle v, u_1\rangle}{\|u_1\|^2}u_1 + \frac{\langle v, u_2\rangle}{\|u_2\|^2}u_2 + \cdots + \frac{\langle v, u_k\rangle}{\|u_k\|^2}u_k.$$

Indeed, if v' denotes the formula on the right side, then $v' \in U$, and a short computation shows that $v - v' \in U^\perp$. By uniqueness of the decomposition $v = \tilde{u} + x$, $\tilde{u} \in U$, $x \in U^\perp$, we conclude that $v' = \operatorname{proj}_U v$.

Corollary 6.2.15. *If U is a subspace of a finite-dimensional inner product space V, then $\left(U^\perp\right)^\perp = U$.*

Proof. If $u \in U$, then $\langle u, v\rangle = 0$ for all $v \in U^\perp$. Thus, by definition, $u \in \left(U^\perp\right)^\perp$. So $U \subseteq \left(U^\perp\right)^\perp$. Moreover,

$$\dim\left(U^\perp\right)^\perp = \dim V - \dim U^\perp = \dim V - (\dim V - \dim U) = \dim U.$$

Thus $U = \left(U^\perp\right)^\perp$. □

Corollary 6.2.16. *Suppose U is a subspace of a finite-dimensional inner product space V. Then the isomorphism $\varphi\colon (U \oplus U^\perp) = V \to V^*$ of Lemma 6.2.10 satisfies $\varphi\left(U^\perp\right) = U^0$.*

Proof. We have

$$\varphi(v) \in U^0 \iff \varphi(v)(u) = 0 \ \forall\ u \in U \iff \langle v, u\rangle = 0 \ \forall\ u \in U \iff v \in U^\perp.$$ □

Exercises.

6.2.1. Suppose U is a subset of an inner product space V.

(a) Show that $U^\perp$ is a subspace of V.

(b) Show that $U^\perp = V$ if and only if $U = \varnothing$ or $U = \{\mathbf{0}\}$. *Note*: We do not make any assumptions here about U or V being finite-dimensional (or even that U is a subspace).

6.2.2. Prove that vectors u, v in an inner product space are orthogonal if and only if $\|u+v\| = \|u - v\|$.

6.2.3. For vectors u, v in an inner product space, prove that $\|u\| = \|v\|$ if and only if $u + v$ and $u - v$ are orthogonal.

6.2.4. Prove that if $\{u_1, \dots, u_k\}$ is an orthogonal basis for an inner product space U, and $v \in U$, then

$$v = \frac{\langle v, u_1\rangle}{\|u_1\|^2}u_1 + \frac{\langle v, u_2\rangle}{\|u_2\|^2}u_2 + \cdots + \frac{\langle v, u_k\rangle}{\|u_k\|^2}u_k.$$

6.2.5 ([Ber14, 5.1.2]). (*Bessel's inequality*) In an inner product space V, if $x_1, \ldots, x_n$ are pairwise orthogonal unit vectors (that is, $\|x_i\| = 1$ for all i, and $x_i \perp x_j$ when $i \neq j$) then

$$\sum_{i=1}^{n} |\langle x, x_i \rangle|^2 \leq \|x\|^2 \quad \text{for all } x \in V.$$

Hint: Define $c_i = \langle x, x_i \rangle$, $y = c_1 x_1 + \cdots + c_n x_n$, and $z = x - y$. Show that $y \perp z$ and apply Exercise 6.1.3 to $x = y + z$ to conclude that $\|x\|^2 \geq \|y\|^2$.

6.2.6 ([Ber14, Ex. 5.1.5]). Suppose that $x_1, \ldots, x_n$ are vectors in an inner product space that are pairwise orthogonal, that is, $\langle x_i, x_j \rangle$ for $i \neq j$.

(a) Prove that $\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2$.

(b) Deduce an alternative proof of Theorem 6.2.2.

6.2.7 ([Ber14, Ex. 5.2.2]). Let X be a set and let $V = \mathcal{F}(X, \mathbb{R})$ (see Example 1.2.5). Let $W \subseteq V$ be the set of all functions $f \in V$ whose support

$$\{x \in X \mid f(x) \neq 0\}$$

is a finite subset of X. Prove the following statements:

(a) W is a subspace of V.

(b) The formula

$$\langle f, g \rangle = \sum_{x \in X} f(x) g(x)$$

defines an inner product on W. (Note that, even though the set X may be infinite, the sum above has at most finitely many nonzero terms, and so is well defined.)

(c) The formula

$$\varphi(f) = \sum_{x \in X} f(x)$$

defines a linear form on W.

(d) If X is infinite, there does not exist $g \in W$ such that $\varphi(f) = \langle f, g \rangle$ for all $f \in W$.

6.2.8 ([Ber14, Ex. 5.2.4]). If V is any inner product space (not necessarily finite dimensional), then $U = (U^\perp)^\perp$ for every finite-dimensional subspace U. *Hint*: The proof of Corollary 6.2.15 is not applicable here, since we do not assume that V is finite dimensional. The problem is to show that $(U^\perp)^\perp \subseteq U$. If $x \in (U^\perp)^\perp$ and $x = y + z$, with $y \in U$ and $z \in U^\perp$, as in Theorem 6.2.9, then $z = x - y \in (U^\perp)^\perp$, so $z \perp z$.

6.2.9 ([Ber14, Ex. 5.2.6]). Suppose U and V are subspaces of a finite-dimensional inner product space. Prove the following:

(a) $(U + V)^\perp = U^\perp \cap V^\perp$;

(b) $(U \cap V)^\perp = U^\perp + V^\perp$.

Hint: Use Corollary 6.2.15.

6.2.10 ([Ber14, Ex. 5.2.7]). Consider $\mathbb{R}^3$ with the canonical inner product. If

$$V = \text{Span}\{(1,1,1),(1,-1,1)\},$$

find $V^\perp$.

6.3 Adjoints

Recall (Lemma 6.2.10) that if V is a finite-dimensional inner product space, then we have an isomorphism

$$\varphi_V \colon V \to V^*, \quad \varphi(v)(w) = \langle v, w\rangle, \quad \forall v, w \in V. \tag{6.1}$$

Definition 6.3.1 (Adjoint of a linear map)**.** If $T\colon V \to W$ is a linear map between finite-dimensional inner product spaces, the *adjoint* of T is the linear map $T^\star\colon W \to V$ defined by

$$T^\star = \varphi_V^{-1} \circ T^* \circ \varphi_W,$$

where $T^*\colon W^* \to V^*$ is the transpose map of T from Definition 3.7.7.

$$\begin{array}{ccc} V & \xleftarrow{T^\star} & W \\ {\scriptstyle \varphi_V^{-1}}\uparrow & & \downarrow{\scriptstyle \varphi_W} \\ V^* & \xleftarrow[T^*]{} & W^* \end{array}$$

Proposition 6.3.2. *If $T\colon V \to V$ is a linear map on a finite-dimensional inner product space, then*

$$\langle Tv, w\rangle = \langle v, T^\star w\rangle \quad \forall\ v, w, \in V. \tag{6.2}$$

Proof. Let $\varphi = \varphi_V\colon V \to V^*$ be the isomorphism (6.1). Then $T^\star = \varphi^{-1} \circ T^* \circ \varphi$ and so $\varphi \circ T^\star = T^* \circ \varphi$. Now, for all $v, w \in V$, we have

$$(\varphi \circ T^\star)(w)(v) = (\varphi(T^\star w))(v) = \langle T^\star w, v\rangle = \langle v, T^\star w\rangle,$$

while

$$(T^* \circ \varphi)(w)(v) = T^*(\varphi(w))(v) = \varphi(w)(Tv) = \langle w, Tv\rangle = \langle Tv, w\rangle.$$

Thus

$$\langle Tv, w\rangle = \langle v, T^\star w\rangle \quad \forall\ v, w, \in V. \qquad \square$$

Proposition 6.3.3. *If V is a finite-dimensional inner product space, then property (6.2) characterizes the adjoint. In other words, if $S\colon V \to V$ satisfies*

$$\langle Tv, w\rangle = \langle v, Sw\rangle \quad \forall\ v, w \in V,$$

then $S = T^\star$.

Proof. We have

$$\begin{aligned}
\langle Tv, w\rangle = \langle v, Sw\rangle \quad \forall\ v, w \in V &\implies \langle v, Sw\rangle = \langle v, T^\star w\rangle \ \forall\ v, w \in V \qquad \text{(by (6.2))} \\
&\implies \langle v, (S - T^\star)w\rangle = 0 \ \forall\ v, w \in V \\
&\implies (S - T^*)w = 0 \ \forall\ w \in V \\
&\implies S = T^\star.
\end{aligned}$$

For the second $\implies$ above, the forward implication $\Rightarrow$ follows by taking $v = (S - T^\star)w$. This gives $\langle (S-T^\star)w, (S-T^\star)w\rangle = 0$, which implies that $(S-T^\star)w = \mathbf{0}$ by Definition 6.1.1(a). □

Remark 6.3.4. Often in the literature, the same symbol is used for the transpose and the adjoint. This is because the isomorphism $\varphi\colon V \to V^*$ is so 'natural' for an inner product space that we 'erase' it in the equation $\varphi \circ T^\star = T^* \circ \varphi$. Moreover, as we will see next, in the classic example of $\mathbb{R}^n$ with the dot product, the adjoint corresponds to the transpose of a matrix.

Theorem 6.3.5. *Fix $A = M_n(\mathbb{R})$ (i.e. A is an $n \times n$ matrix with real entries). Suppose $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(v) = Av$ for all $v \in \mathbb{R}^n$. Then if $\mathbb{R}^n$ is given the dot product, we have*

$$T^\star(v) = A^t v \quad \forall\ v \in \mathbb{R}^n.$$

Proof. Let $v, w \in \mathbb{R}^n$. Then

$$\begin{aligned}
\langle Tv, w\rangle &= (Av) \cdot w && \text{(dot product)} \\
&= (Av)^t w && \text{(matrix product)} \\
&= (v^t A^t)w \\
&= v^t A^t w \\
&= v \cdot (A^t w) \\
&= \langle v, A^t w\rangle.
\end{aligned}$$

Therefore, by Proposition 6.3.3, $T^\star(v) = A^t v$ for all $v \in \mathbb{R}^n$. □

Definition 6.3.6 (Orthogonal transformation, isometry)**.** An *orthogonal transformation*, or *isometry*, is a linear map $T\colon V \to V$ on an inner product space V that preserves the inner product:

$$\langle u, v\rangle = \langle Tu, Tv\rangle, \quad \text{for all } u, v \in V.$$

Corollary 6.3.7. *A linear map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation if and only if its matrix, relative to the canonical basis of $\mathbb{R}^n$, is an orthogonal matrix.*

Proof. Let A be the matrix of T relative to the canonical basis of $\mathbb{R}$. Then $Tv = Av$ for all $v \in \mathbb{R}^n$. We have

$$\begin{aligned}
T \text{ is orthogonal} &\iff \langle Tu, Tv\rangle = \langle u, v\rangle \text{ for all } u, v \in V \\
&\iff \langle u, T^\star Tv\rangle = \langle u, v\rangle \text{ for all } u, v \in V \\
&\iff \langle u, A^t Av\rangle = \langle u, v\rangle \text{ for all } u, v \in V
\end{aligned}$$

$$\begin{aligned}
&\iff \langle u, A^tAv - v\rangle = 0 \text{ for all } u, v \in V \\
&\iff A^tAv - v = \mathbf{0} \text{ for all } v \in V && \text{(as in proof of Prop. 6.3.3)} \\
&\iff A^tAv = v \text{ for all } v \in V \\
&\iff A^tA = I \\
&\iff A \text{ is orthogonal.} && \square
\end{aligned}$$

See Exercise 6.3.5 for a generalization of Corollary 6.3.7.

Exercises.

6.3.1 ([Ber14, Ex. 5.3.1]). Consider $\mathbb{R}^3$ and $\mathbb{R}^2$ as inner product spaces with the canonical inner products. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the linear map defined by

$$T(a, b, c) = (2a - c, 3b + 4c).$$

(a) Find the matrix of $T^\star$ relative to the canonical orthonormal bases.

(b) Find the matrix of $T^\star$ relative to the basis $\frac{1}{2}(1, \sqrt{3}), \frac{1}{2}(-\sqrt{3}, 1)$ of $\mathbb{R}^2$ and the canonical basis e_1, e_2, e_3 of $\mathbb{R}^3$.

6.3.2 ([Ber14, Ex. 5.3.3]). Suppose that U and V are inner product spaces, and $T\colon U \to V$, $S\colon V \to U$ are maps such that

$$\langle Tu, v\rangle = \langle u, Sv\rangle \quad \text{for all } u \in U,\ v \in V.$$

Prove that S and T are linear (and $S = T^\star$ when U and V are finite dimensional).

6.3.3 ([Ber14, Ex. 5.3.4]). For A, B in the vector space $M_n(\mathbb{R})$ of $n \times n$ real matrices, define

$$\langle A, B\rangle = \operatorname{tr}(AB^t).$$

Prove that $M_n(\mathbb{R})$ is an inner product space and find the formula for $\|A\|$ in terms of its entries $a_{i,j}$.

6.3.4 ([Ber14, Ex. 5.3.7]). Suppose U and V are finite-dimensional inner product spaces, $T\colon U \to V$ is a linear map, A is a subset of U, and B is a subset of V. Prove the following:

(a) $T(A)^\perp = (T^\star)^{-1}(A^\perp)$;

(b) $T^\star(B)^\perp = T^{-1}(B^\perp)$;

(c) $T(U)^\perp = \operatorname{Ker} T^\star$, therefore $V = T(U) \oplus \operatorname{Ker} T^*$;

(d) $T^\star(V)^\perp = \operatorname{Ker} T$, therefore $U = T^\star(V) \oplus \operatorname{Ker} T$;

(e) $T(A) \subseteq B \implies T^\star(B^\perp) \subseteq A^\perp$.

6.3.5. Suppose B is an orthonormal basis of a finite-dimensional inner product space V and $T\colon V \to V$ is a linear map. Show that T is an orthogonal transformation if and only if $[T]_B^B$ is an orthogonal matrix. (This generalizes Corollary 6.3.7.)

6.3.6 ([Ber14, Ex. 12.4.3]). Suppose V and W are finite-dimensional inner product spaces and $T \in \mathcal{L}(V, W)$.

(a) Prove that T and $T^\star$ have the same rank.

(b) Prove that T and $T^\star T$ have the same kernel and same rank.

Chapter 7

Diagonalization

In this final chapter we study the topics of eigenvectors, eigenvalues, and diagonalization. You saw these concepts in MAT 1341. Here we discuss them in more detail. In particular, we discuss the diagonalizability of symmetric matrices in some depth. This chapter roughly corresponds to [Tre, Ch. 4].

Throughout this chapter, V is a vector space over a field F. If V is finite dimensional, $T \in \mathcal{L}(V)$, and B is an ordered basis for V, we will write $[T]_B$ for $[T]_B^B$.

7.1 Eigenvectors, eigenvalues, and diagonalization

Diagonal matrices are particularly easy to work with. Thus, if T is a linear operator on a finite-dimensional vector space V, we would like to know when we can find a basis B for V such that $[T]_B$ is diagonal. If such a basis exists, how can we find it?

Definition 7.1.1 (Diagonalizable operator, diagonalizable matrix)**.** A linear operator $T \in \mathcal{L}(V)$ is said to be *diagonalizable* if there is an ordered basis B for V such that $[T]_B$ is a diagonal matrix. A matrix $A \in M_n(F)$ is *diagonalizable* if the linear operator $F^n \to F^n$, $v \mapsto Av$, is diagonalizable.

Lemma 7.1.2. *Suppose $A \in M_n(F)$. If A is diagonalizable, then any matrix similar to A is diagonalizable. Moreover, A is diagonalizable if and only if it is similar to a diagonal matrix.*

Proof. This follows from the discussion in Section 4.2, where we saw that if T is the linear operator

$$T\colon F^n \to F^n, \quad v \mapsto Av,$$

then the matrices $[T]_B$ for T, for ordered bases B, are precisely the matrices similar to A. (In particular, A is the matrix for T in the standard basis.) □

Definition 7.1.3 (Eigenvector, eigenvalue)**.** Suppose $T \in \mathcal{L}(V)$. A nonzero vector $v \in V$ is called an *eigenvector* of T if there exists a scalar $\lambda \in F$ such that

$$Tv = \lambda v.$$

The scalar λ is called the *eigenvalue* corresponding to the eigenvector v.

If $A \in M_n(F)$, a nonzero vector $v \in F^n$ is called an *eigenvector* of A if

$$Av = \lambda v.$$

The scalar λ is called the *eigenvalue* corresponding to the eigenvector v. Note that if we view A as a linear operator $F^n \to F^n$, $v \mapsto Av$, then this definition coincides with the previous one.

Theorem 7.1.4. *Suppose V is finite dimensional.*

(a) A linear operator $T \in \mathcal{L}(V)$ is diagonalizable if and only if there exists a basis for V consisting of eigenvectors of T.

(b) If $T \in \mathcal{L}(V)$ is diagonalizable and $B = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of V consisting of eigenvectors of T, then $[T]_B$ is a diagonal matrix, and its j-th diagonal entry is the eigenvalue λ_j corresponding to v_j, for $j = 1, \ldots, n$.

(c) A matrix $A \in M_n(F)$ is diagonalizable if and only if there exists a basis of F^n consisting of eigenvectors of A.

(d) Suppose $A \in M_n(F)$ is diagonalizable and $B = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of V consisting of eigenvectors of A. Let

$$P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

be the matrix whose i-th column is v_i. Then $P^{-1}AP$ is a diagonal matrix.

Proof. Suppose T is diagonalizable. Then there exists an ordered basis $B = \{v_1, \ldots, v_n\}$ of V such that $[T]_B$ is diagonal. Let λ_j be the j-th diagonal entry of $[T]_B$. Then, for $j = 1, \ldots, n$, we have

$$Tv_j = C_B^{-1}[T]_B C_B v_j = C_B^{-1}[T]_B e_j = C_B^{-1}\lambda_j e_j = \lambda_j v_j.$$

Thus v_j is an eigenvector with corresponding eigenvalue λ_j. So B is a basis for V consisting of eigenvectors of T. This proves the forward implication in part (a).

Now suppose $B = \{v_1, \ldots, v_n\}$ is a basis of V consisting of eigenvectors of T. Then, for $j = 1, \ldots, n$,

$$Tv_j = \lambda_j v_j,$$

where λ_j is the eigenvalue corresponding to v_j. Thus $[T]_B$ is a diagonal matrix, and its j-th diagonal entry is λ_j. This proves part (b) and the reverse implication in part (a).

Part (c) follows immediately from (a).

To prove part (d), suppose $A \in M_n(F)$ is diagonalizable and $B = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of V consisting of eigenvectors of A. Define

$$T\colon F^n \to F^n, \quad Tv = Av.$$

By part (b), T is diagonalizable and $[T]_B$ is diagonal. Let $D = \{e_1, \ldots, e_n\}$ be the standard basis of F^n. By our discussion of change of bases in Section 4.2, we have

$$[T]_B = P^{-1}[T]_D P = P^{-1}AP. \qquad \square$$

Example 7.1.5. Let

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \in M_2(\mathbb{R}), \quad v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Then

$$Av_1 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -2v_1.$$

Thus v_1 is an eigenvector of A with eigenvalue $\lambda_1 = -2$. Similarly,

$$Av_2 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} = 5v_2.$$

Thus v_2 is an eigenvector of A with eigenvalue 5. Since the set $B = \{v_1, v_2\}$ is linearly independent (neither vector is a multiple of the other), it is a basis for $\mathbb{R}^2$. Hence A is diagonalizable and

$$P^{-1}AP = \begin{bmatrix} -2 & 0 \\ 0 & 5 \end{bmatrix} \qquad \text{where} \quad P = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}.$$

Example 7.1.6. Recall that $C^\infty(\mathbb{R})$ is the vector space of all infinitely differentiable functions $\mathbb{R} \to \mathbb{R}$. Consider the linear operator

$$T\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}), \quad Tf = f'.$$

Let $c \in \mathbb{R}$ and define $f(x) = ce^{\lambda x}$. Then $f \in C^\infty(\mathbb{R})$ and

$$Tf = \lambda f.$$

So f is an eigenvector for T with eigenvalue λ. Thus, operators on infinite-dimensional vector spaces can also have eigenvectors.

Definition 7.1.7 (Eigenspace)**.** For $\lambda \in F$ and $T \in \mathcal{L}(V)$, we define

$$E_\lambda := \{v \in V \mid Tv = \lambda v\}.$$

Then λ is an eigenvalue if and only if $E_\lambda \neq \{\mathbf{0}\}$. If λ is an eigenvalue, then E_λ is called the *eigenspace* of T corresponding to λ. We leave it as Exercise 7.1.1 to show that E_λ is a subspace of V. Note that E_λ is the union of the set of all eigenvectors corresponding to λ and $\{\mathbf{0}\}$. The zero vector is never an eigenvector (by definition), but it is contained in any eigenspace.

Example 7.1.8. Consider A as in Example 7.1.5. We know that -2 and 5 are eigenvalues of A. Then E_{-2} consists of all vectors $v \in \mathbb{R}^2$ such that $Av = -2v$. Now

$$Av = -2v \iff (A + 2I)v = 0.$$

In MAT 1341, you learned how to solve matrix equations. Solving the above equation, we find that

$$E_{-2} = \operatorname{Span}\{v_1\}.$$

Similarly, we find that

$$E_5 = \operatorname{Span}\{v_2\}.$$

Proposition 7.1.9. *Suppose $A \in M_n(F)$. A scalar $\lambda \in F$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.*

Proof. For $\lambda \in F$, we have

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff Av = \lambda v \text{ for some } v \neq 0 \\ &\iff (A - \lambda I)v = 0 \text{ for some } v \neq 0 \\ &\iff A - \lambda I \text{ is not invertible} \\ &\iff \det(A - \lambda I) = 0. \end{aligned}$$

□

The polynomial $\det(A - xI)$ is called the *characteristic polynomial* of A. So Proposition 7.1.9 says that the eigenvalues of A are the roots of the characteristic polynomial of A. Note that the degree of the characteristic polynomial of $A \in M_n(F)$ is n.

If P is invertible, then

$$\begin{aligned} \det(A - xI) &= \det(P^{-1}) \det(A - xI) \det(P) \qquad (\text{since } \det(P^{-1}) = \det(P)^{-1}) \\ &= \det(P^{-1}(A - xI)P) \\ &= \det(P^{-1}AP - xI). \end{aligned}$$

Therefore, similar matrices have the same characteristic polynomial. Recall from Section 4.2 that if $T \in \mathcal{L}(V)$ and B, D are basis of V, then $[T]_B$ and $[T]_D$ are similar. Therefore, we can define the *characteristic polynomial* of T to be the characteristic polynomial of the matrix $[T]_B$, and this definition is independent of the choice of the basis B. Note that the degree of the characteristic polynomial of T is the dimension of V.

Example 7.1.10. Suppose

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in M_2(F).$$

Then the characteristic polynomial of A is

$$\det(A - xI) = \det \begin{bmatrix} -x & -1 \\ 1 & -x \end{bmatrix} = x^2 + 1.$$

(a) If $F = \mathbb{R}$, then there are no zeros. Thus, A has no eigenvalues, and hence no eigenvectors. So A is not diagonalizable over $\mathbb{R}$.

(b) Now suppose $F = \mathbb{C}$. Then the zeros of the characteristic polynomial are

$$\lambda_1 = i \qquad \text{and} \qquad \lambda_2 = -i.$$

To find the eigenspace corresponding to eigenvalue λ we need to solve the system

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

When $\lambda = \lambda_1 = i$, we solve

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The solution space for this system is

$$E_i = \left\{ \begin{bmatrix} iz \\ z \end{bmatrix} \mid z \in \mathbb{C} \right\} = \operatorname{Span}_{\mathbb{C}} \left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}.$$

Similarly, we find that

$$E_{-i} = \left\{ \begin{bmatrix} -iz \\ z \end{bmatrix} \mid z \in \mathbb{C} \right\} = \operatorname{Span}_{\mathbb{C}} \left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\}.$$

Since $\{(i, 1), (-i, 1)\}$ is a basis for $\mathbb{C}^2$, the matrix A is diagonalizable over $\mathbb{C}$.

(c) Now suppose $F = \mathbb{F}_2$. Then $\lambda_1 = 1$ is the only root of the characteristic polynomial, hence the only eigenvalue. To find the corresponding eigenspace, we solve

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and find that

$$E_1 = \operatorname{Span}_{\mathbb{F}_2} \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Therefore, it is not possible to find a basis consisting of eigenvectors. So A is not diagonalizable over $\mathbb{F}_2$.

Exercises.

7.1.1. Suppose $\lambda \in F$ and $T \in \mathcal{L}(V)$. Show that E_λ is an subspace of V.

7.1.2. Recall that $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is the real vector space of all functions $\mathbb{R} \to \mathbb{R}$. Define

$$T \colon \mathcal{F}(\mathbb{R}, \mathbb{R}) \to \mathcal{F}(\mathbb{R}, \mathbb{R}), \quad (Tf)(x) = f(-x), \ f \in \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

Show that $\mathcal{F}(\mathbb{R}, \mathbb{R}) = E_1 \oplus E_{-1}$. *Hint*: Look at Exercise 1.5.10.

7.1.3. Let λ be an eigenvalue of $T \in \mathcal{L}(V)$, and let E_λ be the corresponding eigenspace. Show that if $S \in \mathcal{L}(V)$ and $ST = TS$, then $S(E_\lambda) \subseteq E_\lambda$.

7.1.4. Suppose V is finite dimensional and $T \in \mathcal{L}(V)$ is diagonalizable. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T.

(a) Prove that $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$.

(b) Suppose $S \in \mathcal{L}(V)$. Prove that $ST = TS$ if and only if $S(E_{\lambda_i}) \subseteq E_{\lambda_i}$ for all $i = 1, \ldots, k$.

7.1.5 ([Ber14, Ex. 8.2.6]). Let $T \in \mathcal{L}(V)$ be a linear operator such that V has a basis $v_1, \ldots, v_n$ consisting of eigenvectors for T, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. (We do not assume that the λ_i are distinct.) Prove that every eigenvalue of T is equal to one of the λ_i. *Hint*: If $Tv = \lambda v$, express v as a linear combination of the v_i, then apply T.

7.2 Criteria for diagonalization

In this section we investigate diagonalization in further detail and give a test for diagonalizability. Throughout this section, the vector space V is finite dimensional.

Lemma 7.2.1. *Suppose $T \in \mathcal{L}(V)$. If $v_1, \ldots, v_k$ are eigenvectors of T corresponding to distinct eigenvalues, then these vectors are linearly independent.*

Proof. We prove the result by induction on k. If $k = 1$, then $\{v_1\}$ is linearly independent since $v_1 \neq 0$.

Suppose $k \geq 2$ and that the theorem holds for $k-1$ eigenvectors. Let $v_1, \ldots, v_k$ be eigenvectors corresponding to distinct eigenvalues and suppose

$$a_1 v_1 + \cdots + a_k v_k = \mathbf{0}. \tag{7.1}$$

Applying the operator $T - \lambda_k I$ to both sides, we obtain

$$a_1(\lambda_1 - \lambda_k)v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = \mathbf{0}.$$

By the induction hypothesis, the vectors $v_1, \ldots, v_{k-1}$ are linearly independent. Thus

$$a_1(\lambda_1 - \lambda_k) = \cdots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Since the λ_i are all distinct by assumption, we have $\lambda_i - \lambda_k \neq 0$ for $i = 1, \ldots, k-1$. Thus

$$a_1 = \cdots = a_{k-1} = 0.$$

Then (7.1) gives $a_k v_k = \mathbf{0}$. Since $v_k \neq 0$, we have $a_k = 0$. So the vectors $v_1, \ldots, v_k$ are linearly independent, completing the proof of the induction step. □

Corollary 7.2.2. *If a linear operator T on an n-dimensional vector space has n distinct eigenvalues, then T is diagonalizable.*

Proof. Let $v_1, \ldots, v_n$ be eigenvectors of T corresponding to the n distinct eigenvalues. Then, by Lemma 7.2.1, these vectors are linearly independent. Since there are n of them, they form a basis. □

Definition 7.2.3 (Algebraic multiplicity)**.** Let λ be an eigenvalue of a matrix A (or a linear operator T), and let $f(x)$ be the characteristic polynomial of A (respectively, of T). Then the *algebraic multiplicity* of λ is the largest positive integer m_λ such that $(x - \lambda)^{m_\lambda}$ is a factor of $f(x)$, that is, such that $f(x) = (x - \lambda)^{m_\lambda} g(x)$ for some polynomial $g(x)$.

Definition 7.2.4 (Geometric multiplicity)**.** Let λ be an eigenvalue of a matrix A (or a linear operator T). The *geometric multiplicity* of λ is $\dim E_\lambda$.

Theorem 7.2.5. *Let λ be an eigenvalue of $T \in \mathcal{L}(V)$ with algebraic multiplicity m_λ. Then*

$$1 \leq \dim E_\lambda \leq m_\lambda.$$

Proof. Choose a basis $\{v_1, \ldots, v_p\}$ for E_λ and extend it to a basis

$$B = \{v_1, \ldots, v_p, \ldots, v_n\}$$

of V. Let $A = [T]_B$. Since $Av_i = \lambda v_i$ for $i = 1, \ldots, p$, we see that A is of block form

$$A = \begin{bmatrix} \lambda I_p & B \\ 0 & C \end{bmatrix}.$$

The characteristic polynomial of A is

$$f(x) = \det(A - xI_n) = \det\big((\lambda - x)I_p\big) \det(C - xI_{n-p}) = (\lambda - x)^p g(x),$$

where $g(x) = \det(C - xI_{n-p})$. Thus $(\lambda - x)^p$ is a factor of $f(x)$, and so $\dim E_\lambda = p \leq m_\lambda$. □

Example 7.2.6. Let $F = \mathbb{R}$, and consider the matrix

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

Then

$$\det(A - xI) = \det \begin{bmatrix} -x & 1 & 0 \\ 0 & -x & 2 \\ 0 & 0 & -x \end{bmatrix} = -x^3.$$

Thus A has only one eigenvalue $\lambda = 0$, with algebraic multiplicity 3. To find E_0 we solve

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} v = \mathbf{0}$$

and find the solution set

$$E_0 = \left\{ \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\}.$$

So $\dim E_0 = 1 < 3$. So there is no basis for $\mathbb{R}^3$ consisting of eigenvectors, so T is not diagonalizable.

Lemma 7.2.7. *Suppose that $T \in \mathcal{L}(V)$ and $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T. For each $i = 1, \ldots, k$, let $u_i \in E_{\lambda_i}$. If $u_1 + \cdots + u_k = \mathbf{0}$, then $u_i = \mathbf{0}$ for all $i = 1, \ldots, k$.*

Proof. If any of the u_i are nonzero, then $u_1 + \cdots + u_k$ cannot be zero by Lemma 7.2.1. □

Lemma 7.2.8. *Suppose that $T \in \mathcal{L}(V)$ and $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T. For each $i = 1, \ldots, k$, let S_i be a linearly independent subset of E_{λ_i}. Then $S = S_1 \cup \cdots \cup S_k$ is a linearly independent subset of V.*

Proof. For each $i = 1, \ldots, k$, let $S_i = \{v_{i,1}, \ldots, v_{i,n_i}\}$. Suppose

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{i,j} v_{i,j} = \mathbf{0}$$

for some $a_{i,j} \in F$. For each $i = 1, \ldots, k$, let

$$w_i = \sum_{j=1}^{n_i} a_{i,j} v_{i,j}.$$

Then $w_1 + \cdots + w_k = \mathbf{0}$ and $w_i \in E_{\lambda_i}$. By Lemma 7.2.7, we have $w_i = \mathbf{0}$ for all i. Since each S_i is linearly independent, we have $a_{i,j} = 0$ for all j. □

Theorem 7.2.9 (Test for diagonalization)**.** *Suppose T is a linear operator on an n-dimensional vector space V over a field F. Then T is diagonalizable if and only if both of the following conditions hold:*

(a) The characteristic polynomial of T is a product of linear factors

$$c(\lambda_1 - x)^{m_{\lambda_1}} \cdots (\lambda_k - x)^{m_{\lambda_k}},$$

where $c \in F$, $c \neq 0$, the $\lambda_1, \ldots, \lambda_k$ are distinct roots (eigenvalues), and $m_{\lambda_1}, \ldots, m_{\lambda_k}$ are the respective algebraic multiplicities. (Note that $m_{\lambda_1} + \cdots + m_{\lambda_k} = n$.)

(b) For each eigenvalue λ of T, we have $m_\lambda + \operatorname{rank}(T - \lambda I) = n$ (equivalently, $\dim E_\lambda = m_\lambda$).

Proof. $\Rightarrow$: Suppose T is diagonalizable. Then there exists an ordered basis B of V such that $D := [T]_B$ is diagonal. By Theorem 7.1.4, the diagonal entries of D are the eigenvalues of T corresponding to the eigenvectors in the basis B. (These eigenvalues are not necessarily distinct.) Then the characteristic polynomial of T is

$$\det(D - xI_n) = (\lambda_1 - x)^{m_{\lambda_1}} \cdots (\lambda_k - x)^{m_{\lambda_k}},$$

where λ_i, $i = 1, \ldots, k$, are the distinct eigenvalues m_{λ_i} is the number of occurrences of λ_i on the diagonal of D.

For each $i = 1, \ldots, k$, we have m_{λ_i} elements of the basis B that are eigenvectors corresponding to λ_i. Since these vectors are linearly independent, we have $\dim E_{\lambda_i} \geq m_{\lambda_i}$. By Theorem 7.2.5, we have $\dim E_{\lambda_i} \leq m_{\lambda_i}$. Thus $\dim E_{\lambda_i} = m_{\lambda_i}$ for each $i = 1, \ldots, k$.

Finally, for each $i = 1, \ldots, k$, we have

$$\begin{aligned} \operatorname{rank}(T - \lambda_i I_n) &= \dim \operatorname{Im}(T - \lambda_i I_n) \\ &= n - \dim \ker(T - \lambda_i I_n) \qquad \text{(by the Dimension Theorem)} \\ &= n - \dim E_{\lambda_i} \\ &= n - m_{\lambda_i}. \end{aligned}$$

□

$\Leftarrow$: Suppose the two conditions hold. For each $i = 1, \ldots, k$, choose a basis B_i of E_{λ_i}. Then $B = B_1 \cup \cdots \cup B_k$ is linearly independent by Lemma 7.2.8. In addition, B has n elements. Thus B is a basis of V consisting of eigenvectors. Therefore, by Theorem 7.1.4, T is diagonalizable.

A field F is said to be *algebraically closed* if every nonconstant polynomial in $\mathcal{P}(F)$ factors as a product of linear (i.e. degree one) polynomials in $\mathcal{P}(F)$. Thus, if we work over an algebraically closed field, condition (a) of Theorem 7.2.9 is always satisfied. Thus, T is diagonalizable if and only if condition (b) is satisfied.

The field $\mathbb{R}$ is not algebraically closed. For example, x^2+1 does not factor as a product of linear polynomials over $\mathbb{R}$. However, $\mathbb{C}$ is algebraically closed. This is the *Fundamental Theorem of Algebra*, whose proof is beyond the scope of this course.

Exercises.

7.2.1. Suppose that $T \in \mathcal{L}(V)$ has characteristic polynomial $-x^3+3x^2-2x$. Prove that T is diagonalizable.

7.2.2 ([Ber14, Ex. 8.3.3]). Suppose $T \in \mathcal{L}(V)$, M is a nonzero subspace of V such that $T(M) \subseteq M$, and $R = T|_M \colon M \to M$ is the restriction of T to M. Let f_T and f_R be the characteristic polynoimals of T and R, respectively. Show that f_R divides f_T, that is, $f_T = f_R g$ for some polynomial g. *Hint*: Let $v_1, \ldots, v_m$ be a basis of M and consider the matrix of T relative to a basis $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ of V.

7.2.3. Suppose $a, b \in F$. Show that the matrix

$$\begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$$

is diagonalizable if and only if $b = 0$.

7.2.4. If $f(x) = a_0 + a_1 x + \cdots + a_m x^m \in \mathcal{P}(F)$ and A is square matrix, we define

$$f(A) = a_0 + a_1 A + \cdots + a_m A^m.$$

(a) If $A, P \in M_n(F)$, with P invertible, and $f \in \mathcal{P}(F)$, prove that $f(P^{-1}AP) = P^{-1} f(A) P$.

(b) If $D \in M_n(F)$ is diagonal with diagonal entries $d_1, \ldots, d_n$, prove that $f(D)$ is the diagonal matrix with entries $f(d_1), \ldots, f(d_n)$.

(c) Suppose A is a diagonalizable matrix with characteristic polynomial f. Prove that $f(A)$ is the zero matrix.

7.3 Self-adjoint operators and symmetric matrices

In this section and the next we discuss the diagonalization of self-adjoint operators and symmetric matrices. In particular, we will see that *all* self-adjoint operators and symmetric matrices are diagonalizable. Throughout this section V is a finite-dimensional inner product space.

Definition 7.3.1 (Self-adjoint)**.** A linear map $T\colon V \to V$ is *self-adjoint* if

$$T = T^\star.$$

Recall that a matrix A is *symmetric* if $A = A^t$.

Example 7.3.2. Consider $\mathbb{R}^n$ with the dot product, and let $A \in M_n(\mathbb{R})$. Then we have the corresponding linear map

$$T\colon \mathbb{R}^n \to \mathbb{R}^n, \quad T(v) = Av, \ v \in \mathbb{R}^n.$$

We saw in Theorem 6.3.5 that the adjoint is

$$T^\star\colon \mathbb{R}^n \to \mathbb{R}^n, \quad T^\star(v) = A^t v, \ v \in \mathbb{R}^n.$$

Thus T is self-adjoint if and only if $A = A^t$, that is, A is symmetric.

The following lemma generalizes Example 7.3.2.

Lemma 7.3.3. *If B is an orthonormal basis for V, and $C_B\colon V \to \mathbb{R}^n$ is the coordinate isomorphism, then*

(a) $\langle v, w\rangle = C_B(v) \cdot C_B(w)$ (dot product) for all $v, w \in V$, and

(b) if $T\colon V \to V$ is linear, then $[T^\star]_B = ([T]_B)^t$. So T is self-adjoint if and only if $[T]_B$ is symmetric.

Proof. Let $B = \{w_1, \ldots, w_n\}$ and $v, w \in V$. Denote

$$C_B(v) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \qquad C_B(w) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Then

$$\begin{aligned} \langle v, w\rangle &= \left\langle \sum_{i=1}^n a_i w_i, \sum_{j=1}^n b_j w_j \right\rangle \\ &= \sum_{i,j=1}^n a_i b_j \langle w_i, w_j\rangle \\ &= \sum_{i=1}^n a_i b_i \\ &= \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \\ &= C_B(v) \cdot C_B(w). \end{aligned}$$

It remains to prove the second statement. For all $v, w \in V$, we have

$$\begin{aligned}
\langle Tv, w\rangle &= C_B(Tv) \cdot C_B(w) \\
&= ([T]_B C_B(v)) \cdot C_B(w) && \text{(since } [T]_B = C_B T C_B^{-1}\text{)} \\
&= ([T]_B C_B(v))^t \, C_B(w) \\
&= C_B(v)^t \, ([T]_B)^t \, C_B(w) \\
&= C_B(v) \cdot \left(([T]_B)^t \, C_B(w)\right).
\end{aligned}$$

Therefore

$$C_B(v) \cdot \left(([T]_B)^t \, C_B(w)\right) = \langle Tv, w\rangle = \langle v, T^\star w\rangle = C_B(v) \cdot C_B(T^\star w).$$

Since this holds for all v, we have

$$x \cdot \left(([T]_B)^t \, C_B(w)\right) = x \cdot C_B(T^\star w) \quad \forall\ x \in \mathbb{R}^n \text{ and } \forall\ w \in V.$$

Therefore

$$([T]_B)^t \, C_B(w) = C_B(T^\star w) \quad \forall\ w \in V,$$

and so

$$([T]_B)^t \, C_B = C_B T^\star,$$

which implies

$$([T]_B)^t = [T^\star]_B. \qquad \square$$

Corollary 7.3.4. *If D is an orthonormal basis of V, and $\{y_1, \ldots, y_n\}$ is an orthonormal basis of $\mathbb{R}^n$ (with the dot product), then $B = \{C_D^{-1}(y_1), \ldots, C_D^{-1}(y_n)\}$ is also an orthonormal basis of V.*

Proof. By Lemma 7.3.3(a), we have

$$\langle C_D^{-1}(y_i), C_D^{-1}(y_j)\rangle = y_i \cdot y_j = \delta_{ij}. \qquad \square$$

Lemma 7.3.5. *Suppose T is a self-adjoint operator. If v and w are eigenvectors of T corresponding to distinct eigenvalues, then $v \perp w$. In particular, if v and w are eigenvectors of a symmetric matrix A corresponding to distinct eigenvalues, then $v \perp w$.*

Proof. Suppose v and w correspond to eigenvalues λ and μ, respectively, and that $\lambda \neq \mu$. Then we have

$$\begin{aligned}
\lambda\langle v, w\rangle &= \langle \lambda v, w\rangle \\
&= \langle Tv, w\rangle \\
&= \langle v, T^\star w\rangle && \text{(by 6.2)} \\
&= \langle v, Tw\rangle \\
&= \langle v, \mu w\rangle \\
&= \mu\langle v, w\rangle.
\end{aligned}$$

Thus

$$(\lambda - \mu)\langle v, w\rangle = 0.$$

Since $\lambda \neq \mu$, this implies that $\langle v, w\rangle = 0$.

The second assertion follows by considering the linear transformation $v \mapsto Av$. $\square$

Exercises.

7.3.1 ([Ber14, Ex. 12.4.1]). Let U be a subspace of a finite-dimensional inner product space V, and let $P = \mathrm{proj}_U \colon V \to V$ be the orthogonal projection onto U (see Corollary 6.2.13). Show that $P^\star = P = P^2$.

7.3.2 ([Ber14, Ex. 12.4.2]). Let V be a finite-dimensional inner product space and suppose $T \in \mathcal{L}(V)$ satisfies $T^2 = T$ and $T^\star = T$. Define $U = T(V)$. Prove the following statements:

(a) $U = \{v \in V : Tv = v\} = \ker(I - T)$.

(b) $U^\perp = \ker T$.

(c) $T = \mathrm{proj}_U$.

7.4 Diagonalization of self-adjoint operators

We now discuss the diagonalization of self-adjoint operators and symmetric matrices. Throughout this section V is a finite-dimensional inner product space.

Recall that for a complex number $z = a + bi$, $a, b \in \mathbb{R}$, its *complex conjugate* is

$$\bar{z} = a - bi.$$

Its *norm* is

$$|z| = \sqrt{\bar{z}z} = \sqrt{a^2 + b^2}.$$

For a matrix

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \in M_{m,n}(\mathbb{C})$$

we define its *conjugate transpose*

$$A^\dagger = \begin{bmatrix} \overline{a_{1,1}} & \overline{a_{2,1}} & \cdots & \overline{a_{m,1}} \\ \overline{a_{1,2}} & \overline{a_{2,2}} & \cdots & \overline{a_{m,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1,n}} & \overline{a_{2,n}} & \cdots & \overline{a_{m,n}} \end{bmatrix} \in M_{n,m}(\mathbb{C}).$$

Note that

$$A^\dagger = A^t \quad \text{if } A \in M_n(\mathbb{R}).$$

If

$$v = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{C}^n,$$

we have

$$v^\dagger v = \sum_{i=1}^{n} \bar{a}_i a_i = \sum_{i=1}^{n} |a_i|^2 \geq 0.$$

(Here we are identifying a 1×1 matrix with its entry.) It follows that

$$v^\dagger v = 0 \iff v = \mathbf{0}.$$

Proposition 7.4.1. *Every symmetric matrix $A \in M_n(\mathbb{R})$ has a real eigenvector $v \in \mathbb{R}^n$ corresponding to a real eigenvalue.*

Proof. Let $A \in M_n(\mathbb{R})$ be a symmetric matrix and consider the characteristic polynomial

$$f(x) = \det(A - xI).$$

We can view $f(x)$ as an element of $\mathcal{P}(\mathbb{C})$. By the Fundamental Theorem of Algebra, $f(x)$ has a root $\lambda \in \mathbb{C}$. So if we view A as an element of $M_n(\mathbb{C})$, it has an eigenvector $v \in \mathbb{C}^n$ corresponding to λ. So we have

$$\lambda v^\dagger v = v^\dagger(\lambda v) = v^\dagger A v.$$

Taking the conjugate transpose of both sides, we get

$$\bar{\lambda} v^\dagger v = (v^\dagger A v)^\dagger = v^\dagger A^\dagger v = v^\dagger A v = v^\dagger(\lambda v) = \lambda v^\dagger v,$$

where we have used the fact that $A^\dagger = A^t = A$ since A is real symmetric. So

$$(\lambda - \bar{\lambda}) v^\dagger v = \mathbf{0}.$$

Since $v \neq \mathbf{0}$, it follows that $\bar{\lambda} = \lambda$, and so $\lambda \in \mathbb{R}$.

It remains to show that $Ax = \lambda x$ for some nonzero x in $\mathbb{R}^n$ (as opposed to $\mathbb{C}^n$). We can write

$$v = (a_1 + ib_1, \ldots, a_n + ib_n) = \underbrace{(a_1, \ldots, a_n)}_{u} + i \underbrace{(b_1, \ldots, c_n)}_{w},$$

where $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$. Then we have

$$\lambda u + i\lambda w = \lambda v = Av = A(u + iw) = Au + iAw.$$

Comparing real and imaginary parts, we see that $Au = \lambda u$ and $Aw = \lambda w$. Since $v \neq \mathbf{0}$, at least one of u and w is nonzero. So at least one of them is a real eigenvector corresponding to the eigenvalue λ. □

Corollary 7.4.2. *Every self-adjoint linear operator on V has a real eigenvalue.*

Proof. Choose an orthonormal basis B of V. Then, by Lemma 7.3.3, the matrix $[T]_B$ is symmetric. By Lemma 7.4.1, $[T]_B$ has an eigenvalue λ corresponding to eigenvector v. Then

$$T C_B^{-1}(v) = C_B^{-1}[T]_B v = \lambda C_B^{-1}(v).$$

So $C_B^{-1}(v)$ is an eigenvector of T corresponding to eigenvalue λ. □

Theorem 7.4.3. *Suppose V is a finite-dimensional inner product space. The linear operator $T\colon V \to V$ is* self-adjoint *if and only if there is an orthonormal basis of V consisting of eigenvectors of T. Equivalently (by Theorem 7.1.4) T is self-adjoint if and only if there exists an orthonormal basis B such that $[T]_B$ is diagonal.*

Proof. The "if" part follows from Lemma 7.3.3 and the fact that diagonal matrices are symmetric. So it remains to prove the "only if" part.

Suppose for a moment that we can find an orthonormal basis B of V such that $[T]_B$ is upper triangular. By Lemma 7.3.3, $[T]_B$ is also symmetric. So $[T]_B$ must in fact be diagonal. Therefore, it suffices to find an orthonormal basis B of V such that $[T]_B$ is upper triangular. We do this by induction on $n = \dim V$.

First suppose $n = 1$ and choose a basis $\{v_1\}$ of V. Then we must have $Tv_1 = \lambda v_1$ for some λ. Hence v_1 is an eigenvector, and

$$\left\{ \frac{v_1}{\|v_1\|} \right\}$$

is an orthonormal basis of V consisting of eigenvectors. Since the matrix for T in this basis is 1×1, it is clearly upper triangular.

Now suppose $n \geq 2$ and that the result holds for vector spaces of dimension $n - 1$. By Corollary 7.4.2, T has a real eigenvector v_1 corresponding to a real eigenvalue λ_1. Let $W = \{v_1\}^\perp$, and let $v_2, \ldots, v_n$ be an orthonormal basis for W. Then $B = \{v_1, \ldots, v_n\}$ is an orthonormal basis for V and

$$[T]_B = \begin{bmatrix} \lambda & * & \cdots & * \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{bmatrix}.$$

The matrix A defines a linear transformation

$$W \to W, \quad v \mapsto Av.$$

By the induction hypothesis, there exists an orthonormal basis $\{v_2, \ldots, v_n\}$ of W with respect to which the matrix of this operator is upper triangular. Since $[T]_B$ is upper triangular when A is, we are done. □

Corollary 7.4.4. *A matrix $A \in M_n(\mathbb{R})$ is symmetric if and only if there exists an orthogonal matrix C and a diagonal matrix D such that $C^tAC = D$.*

Proof. For the "if" part note that if C is orthogonal and D is diagonal with $C^tAC = D$, then $A = CDC^t$ and so

$$A^t = \left(CDC^t\right)^t = CD^tC^t = CDC^t = A.$$

To prove the "only if part", suppose $A \in M_n(\mathbb{R})$ is symmetric. Define

$$T\colon \mathbb{R}^n \to \mathbb{R}^n, \quad Tv = Av, \quad v \in \mathbb{R}^n.$$

Then $A = [T]_B$, where B is the canonical basis of $\mathbb{R}^n$. (So B is an orthonormal basis.) Since A is symmetric, Lemma 7.3.3 implies that T is self-adjoint. Thus, by Theorem 7.4.3, there exists an orthonormal basis $B' = \{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that $D = [T]_{B'}$ is diagonal. Let

$$C = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

be the change of basis matrix from B to B'. Since the columns of C form an orthonormal basis, the matrix C is orthogonal by Lemma 6.2.7. Thus $C^tC = I$, and so $C^{-1} = C^t$. Then, as in Section 4.2, we have

$$D = [T]_{B'} = C^{-1}[T]_B C = C^t A C. \qquad \square$$

To *orthogonally diagonalize* a matrix A is to find an orthogonal matrix C and a diagonal matrix D such that $C^tAC = D$.

Example 7.4.5. Consider the matrix

$$A = \begin{bmatrix} 5 & 4 \\ 4 & -1 \end{bmatrix}.$$

Let's find a diagonal matrix D and an orthogonal matrix C such that $C^tAC = D$.

First we find the eigenvalues of A. We set

$$\begin{aligned} 0 = \det(A - \lambda I) &= \begin{vmatrix} 5-\lambda & 4 \\ 4 & -1-\lambda \end{vmatrix} \\ &= (5-\lambda)(-1-\lambda) - 16 = \lambda^2 - 4\lambda - 21 = (\lambda - 7)(\lambda + 3). \end{aligned}$$

Thus the eigenvalues are 7 and -3. For the eigenvalue $\lambda = -3$, we find the eigenspace E_{-3} by solving

$$(A - \lambda I)x = \mathbf{0} \implies (A + 3I)x = \mathbf{0}.$$

We do this by row reducing

$$\left[\begin{array}{cc|c} 8 & 4 & 0 \\ 4 & 2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right],$$

and so

$$E_{-3} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right\}.$$

Since we want C to be an orthogonal matrix, its columns must form an orthonormal basis (of eigenvectors). So we need a unit vector in E_{-1}. Take

$$v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Now, we could repeat this process with the eigenvalue 7. However, we can save ourselves some effort by using Theorem 7.4.3 to conclude that A (which is symmetric) must have an orthonormal basis. So we take

$$v_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

(we're using here the fact that $(-b, a)$ is orthogonal to (a, b)). Then $\{v_1, v_2\}$ is an orthonormal basis of eigenvectors,

$$C = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

is an orthogonal matrix, and

$$C^t A C = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 0 & 7 \end{bmatrix}.$$

Example 7.4.6. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $Tv = Av$, $v \in \mathbb{R}^3$, where

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Let's find an orthogonal matrix C such that $C^t A C = D$ is diagonal.

First we find the eigenvalues of A.

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 4 & 2 & 3-\lambda \end{vmatrix} \overset{\text{exercise}}{=} -(\lambda+1)^2(\lambda-8).$$

So the eigenvalues are -1 (with algebraic multiplicity 2) and 8 (with algebraic multiplicity 1). Now,

$$E_{-1} = \operatorname{Ker}(A + I) = \operatorname{Ker} \begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} = \operatorname{Ker} \begin{bmatrix} 2 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{Span} \left\{ \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We convert the basis of the eigenspace into an orthogonal basis using the Gram-Schmidt algorithm:

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} \rightsquigarrow \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -4/5 \\ -2/5 \\ 1 \end{bmatrix} \right\}$$

or (since we can scale each vector by anything we like)

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ -5 \end{bmatrix} \right\}$$

Similarly,

$$E_8 = \operatorname{Ker}(A - 8I) = \operatorname{Ker} \begin{bmatrix} -5 & 2 & 4 \\ 2 & -8 & 2 \\ 4 & 2 & -5 \end{bmatrix} = \operatorname{Ker} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 1/2 \\ 1 \end{bmatrix} \right\}.$$

Thus

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \right\}$$

is a basis for E_8.

We now form the matrix C we're after by using the basis vectors we've found (after dividing by their norm in order to make them unit vectors) as the columns. Thus

$$C = \begin{bmatrix} -\sqrt{5}/5 & 4\sqrt{5}/15 & 2/3 \\ 2\sqrt{5}/5 & 2\sqrt{5}/15 & 1/3 \\ 0 & -\sqrt{5}/3 & 2/3, \end{bmatrix}$$

and

$$C^t AC = D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

Remark 7.4.7. The only thing new here (over what you did in MAT 1341) is that we're dealing with symmetric matrices and we're not looking for just *any* basis of eigenvectors, but rather an *orthonormal* basis of eigenvectors. So in eigenspaces of dimension greater than 1, we must use the Gram-Schmidt algorithm to get an orthonormal basis of the eigenspace. Eigenvectors corresponding to different eigenvalues are *automatically* orthogonal by Lemma 7.3.5.

Exercises.

7.4.1 ([Tre, Ex. 6.2.4]). Orthogonally diagonalize the matrix

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}.$$

Find all square roots of A, that is, find all matrices B such that $B^2 = A$.

7.4.2 ([Tre, Ex. 6.2.4]). Orthogonally diagonalize the matrix

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}.$$

Among all square roots of A, find one that has positive eigenvalues.

7.4.3 ([Tre, Ex. 6.2.8]). Let A be an $m \times n$ matrix. Prove the following statements:

(a) $A^t A$ is symmetric.

(b) All eigenvalues of $A^t A$ are nonnegative.

(c) $A^t A + I$ is invertible.

7.5 Rigid motions

Recall from Definition 6.3.6 that $T \in \mathcal{L}(V)$ is an isometry if it preserves the inner product. In this final section we discuss a related but more general type of map called a rigid motion. Throughout this section, V is an inner product space.

The following theorem gives several equivalent characterizations of isometries.

Theorem 7.5.1. *Suppose V is a finite-dimensional and $T \in \mathcal{L}(V)$. The following statements are equivalent:*

(a) $TT^\star = T^\star T = I$.

(b) $\langle Tv, Tw\rangle = \langle v, w\rangle$ *for all* $v, w \in V$.

(c) If $\{v_1, \ldots, v_n\}$ *is an orthonormal basis for* V*, then* $\{Tv_1, \ldots, Tv_n\}$ *is an orthonormal basis.*

(d) $\|Tv\| = \|v\|$ *for all* $v \in V$.

Proof. (a) $\implies$ (b): Suppose (a) holds. Then, for $v, w \in V$, we have

$$\langle Tv, Tw\rangle = \langle v, T^\star Tw\rangle = \langle v, Iw\rangle = \langle v, w\rangle.$$

(b) $\implies$ (a): Suppose (b) holds. Then, for all $v, w \in V$, we have

$$\langle v, (T^\star T - I)w\rangle = \langle v, T^\star Tw\rangle - \langle v, w\rangle = \langle Tv, Tw\rangle - \langle v, w\rangle = 0.$$

Thus, taking $v = (T^\star T - I)w$, we see that $(T^\star T - I)w = \mathbf{0}$. Therefore $T^\star Tw = Iw = w$ for all $w \in V$. So $T^\star T = I$.

(b) $\implies$ (c): Suppose (b) holds and $B = \{v_1, \ldots, v_n\}$ is an orthonormal basis for V. Then

$$\langle Tv_i, Tv_j\rangle = \langle v_i, v_j\rangle = \delta_{i,j},$$

and so $\{Tv_1, \ldots, Tv_n\}$ is an orthonormal basis.

(c) $\implies$ (d): Suppose (c) holds and let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Any $v \in V$ can be written as

$$v = \sum_{i=1}^{n} a_i v_i, \quad a_1, \ldots, a_n \in \mathbb{R}.$$

Then

$$\|v\|^2 = \left\langle \sum_{i=1}^{n} a_i v_i, \sum_{j=1}^{n} a_j v_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle v_i, v_j\rangle = \sum_{i=1}^{n} a_i^2.$$

Now, we also have

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

By assumption, $\{Tv_1, \dots, Tv_n\}$ is an orthonormal basis. Thus

$$\|Tv\|^2 = \left\langle \sum_{i=1}^{n} a_i Tv_i, \sum_{j=1}^{n} a_j Tv_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \langle Tv_i, Tv_j\rangle = \sum_{i=1}^{n} a_i^2.$$

Hence $\|v\| = \|Tv\|$.

(d) $\implies$ (b): Suppose (d) holds. By the polarization identity (Theorem 6.1.6(c)), for all $v, w \in V$, we have

$$\begin{aligned}\langle Tv, Tw\rangle &= \frac{1}{4}\big(\|Tv + Tw\|^2 - \|Tv - Tw\|^2\big)\\ &= \frac{1}{4}\big(\|T(v + w)\|^2 - \|T(v - w)\|^2\big)\\ &= \frac{1}{4}\big(\|v + w\|^2 - \|v - w\|^2\big)\\ &= \langle v, w\rangle.\end{aligned}$$

□

Definition 7.5.2 (Rigid motion)**.** A function $f\colon V \to V$ is called a *rigid motion* if

$$\|f(x) - f(x)\| = \|x - y\| \quad \text{for all } x, y \in V.$$

If we define the *distance* between x and y to be $\|x - y\|$, then rigid motions are maps that preserve distance.

Example 7.5.3. Any isometry $T\colon V \to V$ is a rigid motion since

$$\|Tx - Ty\| = \|T(x - y)\| = \|x - y\| \quad \text{for all } x, y \in V.$$

Example 7.5.4 (Translations). A function $\rho\colon V \to V$ is called a *translation* if there exists a vector $v \in V$ such that

$$\rho(u) = u + v \quad \text{for all } u \in V.$$

Then, for all $x, y \in V$, we have

$$\|\rho(x) - \rho(y)\| = \|x + v - (y - v)\| = \|x - y\|.$$

So translations are rigid motions. However, a translation is *not* a linear operator unless $v = \mathbf{0}$ since $\rho(\mathbf{0}) = v$.

Lemma 7.5.5. *If $f, g\colon V \to V$ are rigid motions, then so is the composition $f \circ g$.*

Proof. For all $x, y \in V$, we have

$$\begin{aligned}\|(f \circ g)(x) - (f \circ g)(y)\| &= \|f(g(x)) - f(g(y))\| \\ &= \|g(x) - g(y)\| && \text{(since } f \text{ is a rigid motion)}\\ &= \|x - y\| && \text{(since } g \text{ is a rigid motion)}\end{aligned}$$

So $f \circ g$ is a rigid motion. □

Theorem 7.5.6. *Suppose V is finite-dimensional and $f\colon V \to V$ is a rigid motion. Then there exists a unique orthogonal operator $T \in \mathcal{L}(V)$ and a unique translation $\rho\colon V \to V$ such that $f = \rho \circ T$. In other words, every rigid motion can be written in a unique way as an isometry followed by a translation.*

Proof. Define

$$\begin{aligned} T\colon V \to V, &\quad T(u) = f(u) - f(\mathbf{0}), \quad u \in V, \\ \rho\colon V \to V, &\quad \rho(u) = u + f(\mathbf{0}), \quad u \in V. \end{aligned}$$

Then $f = \rho \circ T$. We will show that T is an isometry.

Since T is the composition of f and translation by $-f(0)$, it is a rigid motion by Lemma 7.5.5. It is also clear that $T(\mathbf{0}) = \mathbf{0}$. For all $u \in V$ we have

$$\begin{aligned} \|T(u)\| &= \|T(u) - T(\mathbf{0})\| \\ &= \|f(u) - f(\mathbf{0})\| \\ &= \|u - \mathbf{0}\| && \text{(since } f \text{ is a rigid motion)} \\ &= \|u\|. \end{aligned}$$

Now, for $u, v \in V$, we have

$$\begin{aligned} \|T(u) - T(v)\|^2 &= \langle T(u) - T(v), T(u) - T(v)\rangle \\ &= \|T(u)\| - 2\langle T(u), T(v)\rangle + \|T(v)\|^2 \\ &= \|u\|^2 - 2\langle T(u), T(v)\rangle + \|v\|^2. \end{aligned}$$

On the other hand, we have

$$\|u - v\|^2 = \langle u - v, u - v\rangle = \|u\| - 2\langle u, v\rangle + \|v\|^2.$$

Since T is a rigid motion, we have $\|T(u) - T(v)\|^2 = \|u - v\|^2$. Thus

$$\langle T(u), T(v)\rangle = \langle u, v\rangle \quad \text{for all } u, v \in V. \tag{7.2}$$

Now we show that T is linear. For $u, v \in V$, we have

$$\begin{aligned} \|T(u+v) - T(u) - T(v)\|^2 &= \|T(u+v) - T(u)\| - 2\langle T(u+v) - T(u), T(v)\rangle + \|T(v)\|^2 \\ &= \|u + v - u\|^2 - 2\langle T(u+v), T(v)\rangle + 2\langle T(u), T(v)\rangle + \|v\|^2 \\ &= 2\|v\|^2 - 2\langle u + v, v\rangle + 2\langle u, v\rangle \\ &= 0. \end{aligned}$$

Thus $T(u+v) = T(u) + T(v)$. In addition, for $v \in V$ and $c \in \mathbb{R}$, we have

$$\begin{aligned} \|T(cv) - cT(v)\|^2 &= \|T(cv)\|^2 - 2\langle T(cv), cT(v)\rangle + \|cT(v)\|^2 \\ &= \|cv\|^2 - 2c\langle T(cv), T(v)\rangle + c^2\|T(v)\|^2 \\ &= 2c^2\|v\|^2 - 2c\langle cv, v\rangle \\ &= 2c^2\|v\|^2 - 2c^2\|v\|^2 \end{aligned}$$

$$= 0.$$

So $T(cv) = cT(v)$. Hence T is linear and, by (7.2), it is orthogonal.

It remains to prove uniqueness. Suppose $U, T \in \mathcal{L}(V)$ are orthogonal, $u, v \in V$, and

$$f(x) = T(x) + u = U(x) + v \quad \text{for all } x \in V.$$

Taking $x = \mathbf{0}$, we see that $u = v$. Then $T(x) = U(x)$ for all $x \in V$, so $T = U$. □

Exercises.

7.5.1. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be an orthogonal operator, and let $A = [T]_B \in M_2(\mathbb{R})$, where $B = \{e_1, e_2\}$ is the standard basis for $\mathbb{R}^2$. Then $\{Te_1, Te_2\}$ is also an orthonormal basis for $\mathbb{R}^2$ by Theorem 7.5.1. In particular, Te_1 is a unit vector, and so there is a unique angle θ, $0 \le \theta < 2\pi$, such that $Te_1 = (\cos\theta, \sin\theta)$.

(a) Find all possibilities for Te_2.

(b) Show that one of the following statements must hold:

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

  (ii) $\det A = -1$ and T is a reflection about a line in the origin.

(c) Prove that any rigid motion in $\mathbb{R}^2$ is either a rotation followed by a translation or a reflection in a line through the origin followed by a translation.

Appendix A

A taste of abstract algebra

In this optional appendix, we explore some of the properties that binary operations on sets can have. We also give the precise definition of a field omitted in Section 1.1. The abstract point of view taken here allows one to prove results in a very general setting, and then obtain more specific results as special cases. For example, the fact that additive inverses in fields and additive inverses in vectors are both unique follows from a more general statement about inverses in a monoid.

A.1 Operations on sets

Definition A.1.1 (Operation on a set)**.** Suppose E is a set. We say that $\star$ is an *operation on* E if for every $x, y \in E$, $x \star y$ is a well-defined element of E. A pair $(E, \star)$, where E is a set and $\star$ is an operation on E is called a *set with operation*, or *magma*.

So, loosely speaking, an operation $\star$ on a set E is a "rule" that assigns an element $x \star y$ of E to every pair of elements (x, y) of E. More precisely, $\star$ is a map from $E \times E = \{(x, y) \mid x, y \in E\}$ to E.

Remark A.1.2. The term *magma* is not particularly well-known, even among mathematicians. We will use it simply because it is shorter than "set with operation".

Examples A.1.3. (a) $(\mathbb{Z}, +)$ is a magma.

(b) $(\mathbb{Z}, -)$ is a magma. However, subtraction is *not* an operation on the set $\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$ since $x - y$ is not an element of $\mathbb{N}$ for all $x, y \in \mathbb{N}$. Thus, $(\mathbb{N}, -)$ is *not* a magma.

(c) If "$\div$" denotes ordinary division of numbers, then $\div$ is not an operation on $\mathbb{R}$ because $x \div y$ is not defined when $y = 0$. However, if we let $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, then $(\mathbb{R}^*, \div)$ is a magma.

Definition A.1.4. Suppose that $(E, \star)$ is a magma.

(a) We say that $\star$ is *commutative* if $x \star y = y \star x$ for all $x, y \in E$.

(b) We say that $\star$ is *associative* if $(x \star y) \star z = x \star (y \star z)$ for all $x, y, z \in E$.

Example A.1.5. Consider the set $E = \{1, 2, 3\}$ and the operation $\star$ defined by the table:

$\star$	1	2	3
1	3	2	3
2	1	1	2
3	2	2	2

The table tells us that $1 \star 2 = 2$ and $2 \star 1 = 1$. Note that $\star$ is indeed an operation on E, because it satisfies the requirement of Definition A.1.1. So $(E, \star)$ is a magma. Observe that $1 \star 2 \neq 2 \star 1$, so this operation is not commutative.

Examples A.1.6. (a) In the magma $(\mathbb{R}, +)$, the operation is commutative and associative.

(b) In the magma $(\mathbb{R}, \cdot)$, the operation is commutative and associative.

(c) In the magma $(\mathbb{R}, -)$, the operation is *not* commutative and *not* associative. To see that it is not commutative, we note that, for instance, $3 - 5 \neq 5 - 3$. To prove it is not associative, we note that, for example, $(1 - 3) - 5 \neq 1 - (3 - 5)$.

Remark A.1.7. To show that a magma is *not* associative or commutative, you only need to find *one* counterexample, but to show that a magma *is* associative or commutative, you have to show that the property is satisfied for *all* elements—you *cannot* just give a particular example. For instance, in Example A.1.5, we have $2 \star 3 = 2 = 3 \star 2$ but $\star$ is not a commutative operation (as we noted in the example).

Exercises.

A.1.1. Let $E = \{1, 2, 3\}$ and define $\star$ by the table:

$\star$	1	2	3
1	3	4	3
2	1	1	2
3	2	2	2

Is $(E, \star)$ a magma? If so, is it commutative and/or associative?

A.2 Use of parentheses

Suppose $(E, \star)$ is a magma. In expressions like $a \star (b \star (c \star d))$, can we omit the parentheses? The answer depends on whether or not $\star$ is associative.

If $\star$ is an associative operation then we can omit all parentheses. The reason is that it doesn't matter whether by $a \star b \star c$ we mean $(a \star b) \star c$ or $a \star (b \star c)$, since these two quantities

are equal. We may also write longer expressions like $a \star b \star c \star d$ without parentheses, because the possible interpretations

$$((a \star b) \star c) \star d, \quad (a \star (b \star c)) \star d, \quad (a \star b) \star (c \star d), \quad a \star ((b \star c) \star d), \quad a \star (b \star (c \star d)),$$

all give the same result, so there is no ambiguity.

The general rule is that parentheses must be used when the operation is not associative. There is one exception to this rule: with the non-associative operation of subtraction, parentheses can be omitted in certain cases. Namely, people have adopted the convention that in expressions such as $a-b-c-d$ (without parentheses), the leftmost operation is evaluated first. In other words, the convention says that $a - b - c - d$ should always be interpreted as meaning $((a - b) - c) - d$. So, if you want to write $((a - b) - c) - d$, then you can omit the parentheses, thanks to this convention; but if you want to write one of

$$(a - (b - c)) - d, \quad (a - b) - (c - d), \quad a - ((b - c) - d), \quad a - (b - (c - d))$$

then the convention does not help you and you must use parentheses.

The convention to evaluate the leftmost operation first (in the absence of parentheses) is universally accepted in the case of subtraction, but not for other operations. If you work with a non-associative operation other than subtraction, you should not assume that you can use that convention.

For instance, let $\wedge$ denote the operation of exponentiation on the set $A = \{1, 2, 3, \dots\}$ of positive integers (i.e., $2 \wedge 3 = 2^3 = 8$ and $3 \wedge 2 = 3^2 = 9$). Then $(A, \wedge)$ is a magma. If you enter the expression $2 \wedge 2 \wedge 3$ on your calculator, it will likely give you the answer 64; so the calculator evaluates the leftmost operation first:

$$2 \wedge 2 \wedge 3 = (2 \wedge 2) \wedge 3 = (2^2)^3 = 4^3 = 64.$$

However, if you ask a mathematician the same question, she will probably do this:

$$2 \wedge 2 \wedge 3 = 2^{2^3} = 2^8 = 256,$$

so the mathematician evaluates the rightmost operation first. For the same reason, a mathematician will always interpret e^{x^2} as meaning $e^{(x^2)}$, never as $(e^x)^2$.

Conclusion: Use parentheses whenever there is a possibility of confusion.

Exercises.

Exercise A.2.1. Compute the five possible interpretations of $16 \div 8 \div 4 \div 2$ (two of these interpretations give the same result but the others are all different, so you should get four different numbers).

A.3 Identity elements

Definition A.3.1 (Identity element)**.** Suppose that $(E, \star)$ is a magma. If e is an element of E satisfying

$$e \star x = x = x \star e \quad \text{for all } x \in E,$$

we call e an *identity element* of $(E, \star)$.

Examples A.3.2. (a) 0 is an identity element of $(\mathbb{R}, +)$ because

$$0 + x = x = x + 0 \quad \text{for all } x \in \mathbb{R}.$$

(b) 1 is an identity element of $(\mathbb{R}, \cdot)$ because

$$1 \cdot x = x = x \cdot 1 \quad \text{for all } x \in \mathbb{R}.$$

Note that in the above examples, the same set has different identities for different operations. So an identity element depends on the set *and* the operation.

Example A.3.3. Does $(\mathbb{R}, -)$ have an identity element? Suppose e were an identity element. Then we would have

$$x - e = x \quad \text{for all } x \in \mathbb{R}.$$

This is satisfied for $e = 0$ and so one might be tempted to think that 0 is an identity element. But we must also have

$$e - x = x \quad \text{for all } x \in \mathbb{R}.$$

For each *particular* x, the equation $e - x = x$ is only satisfied for $e = 2x$. But $2x$ has a different value for each x (and is only equal to zero when $x = 0$). Therefore, there is no $e \in \mathbb{R}$ that satisfies $e - x = x$ for *all* $x \in \mathbb{R}$. So $(\mathbb{R}, -)$ has no identity element.

We have seen that some magmas do not have an identity element, while others do. Is it possible for a magma to have more than one identity element?

Theorem A.3.4. *A given magma can have at most one identity element.*

Proof. Suppose that $(E, \star)$ is a magma and that e_1, e_2 are identity elements of $(E, \star)$. Then

$$\begin{aligned} e_1 &= e_1 \star e_2 && \text{(because } e_2 \text{ is an identity element)} \\ &= e_2 && \text{(because } e_1 \text{ is an identity element).} \end{aligned}$$

□

Exercises.

A.3.1. Verify that the magma $(E, \star)$ in Example A.1.5 does not have any identity element.

A.3.2. In the magma $(\mathbb{R}^*, \div)$, is the operation commutative? Associative? Does it have an identity element?

A.3.3. Consider the set $E = \{1, 2, 3, 4\}$ and the operation $\star$ defined by:

$\star$	1	2	3	4
1	1	2	1	4
2	2	3	2	4
3	1	2	3	4
4	4	4	4	4

Does the magma $(E, \star)$ have an identity element? If so, find all the invertible elements of $(E, \star)$.

A.3.4. Consider the pair $(\mathbb{R}, \star)$, where $x \star y = 2^x 2^y$. For example, $3 \star (-1) = 2^3 2^{-1} = 4$.

(a) Is the pair $(\mathbb{R}, \star)$ a magma? If so, does it have an identity element?

(b) If we restrict $\star$ to $\mathbb{Z}$, is the pair $(\mathbb{Z}, \star)$ a magma?

A.4 Invertible elements

Definition A.4.1 (Invertible, inverse)**.** Let $(E, \star)$ be a magma and suppose that $(E, \star)$ has an identity element $e \in E$.

An element $a \in E$ is *invertible* if there exists at least one $x \in E$ satisfying

$$a * x = e = x * a.$$

Any such x is then called *an inverse* of a.

It is important to note that it only makes sense to talk about invertibility and inverses if the magma has an identity element.

Examples A.4.2. (a) In $(\mathbb{R}, +)$, is -2 invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$(-2) + x = 0 = x + (-2)?$$

Yes, $x = 2$ satisfies this requirement. So -2 is invertible and 2 is an inverse of -2. In fact, in $(\mathbb{R}, +)$, all elements are invertible: the inverse of $x \in \mathbb{R}$ is $-x$.

(b) In $(\mathbb{R}, \cdot)$, is -2 invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$(-2) \cdot x = 1 = x \cdot (-2)?$$

Yes, $x = -1/2$ satisfies this requirement. So -2 is invertible and $-1/2$ is an inverse of -2.

Again in $(\mathbb{R}, \cdot)$, is 0 invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$0 \cdot x = 1 = x \cdot 0?$$

No, such an x does not exist, so 0 is not invertible. In fact, the set of invertible elements of $(\mathbb{R}, \cdot)$ is equal to $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$. The inverse of $x \in \mathbb{R}^\times$ is $1/x$.

(c) The element 2 is not an invertible element in the magma $(\mathbb{Z}, \cdot)$. In fact, the set of invertible elements of $(\mathbb{Z}, \cdot)$ is equal to $\{1, -1\}$.

(d) In $(\mathbb{R}, -)$, is 3 invertible? STOP! This question does not make sense, because $(\mathbb{R}, -)$ does not have an identity element.

When $(E, *)$ does not have an identity element, the concept of invertibility is not defined, so it is absurd to ask whether a given element is invertible. Read Definition A.4.1 again and pay attention to the role played by the identity element in that definition. Do you see that, if e does not exist, it makes no sense to speak of invertibility?

(e) Here is a disturbing example. Consider the set $E = \{1, 2, 3, 4\}$ and the operation $*$ defined by the table:

$*$	1	2	3	4
1	1	2	3	4
2	2	3	1	1
3	3	1	3	1
4	4	1	1	4

So $(E, *)$ is a magma. Observe that $1 * x = x = x * 1$ for all $x \in E$, so $(E, *)$ has an identity element (namely, 1) and consequently the concept of invertibility makes sense.

Is 2 invertible? Does there exist $x \in E$ satisfying

$$2 * x = 1 = x * 2?$$

Yes, there are even two such x: $x = 3$ and $x = 4$ satisfy this condition. So 2 is invertible and has two inverses: 3 and 4 are inverses of 2.

So it is possible for a given element to have more than one inverse! However, we will see below that this unpleasant phenomenon never happens when the operation is associative.

Exercises.

A.4.1. Prove that the magma of Example A.4.2(e) is not associative.

A.4.2. Let $(E, \star)$ be a magma, with identity element e. Show that e is invertible, and is its own inverse (more precisely, show that e is the unique inverse of e). Let $a, b \in E$; show that if b is an inverse of a then a is an inverse of b.

A.4.3. Let $\times$ denotes the vector product (or "cross product") on $\mathbb{R}^3$. Then $(\mathbb{R}^3, \times)$ is a magma. Is $(1, 1, 2)$ an invertible element?

A.5 Monoids

Definition A.5.1 (Monoid)**.** A *monoid* is a magma $(E, \star)$ such that

- $(E, \star)$ has an identity element, and
- the operation $\star$ is associative.

Examples A.5.2. (a) $(\mathbb{R}, +)$, $(\mathbb{R}, \cdot)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}, \cdot)$, $(\mathbb{N}, +)$, and $(\mathbb{N}, \cdot)$ are examples of monoids. Also: $(\mathbb{R}^2, +)$, $(\mathbb{R}^3, +)$, and $(\mathbb{R}^n, +)$ (for any $n \geq 1$) are monoids. All these examples are commutative monoids.

(b) Let $A = \{x \in \mathbb{Z} \mid x \geq 1\} = \{1, 2, 3, 4, \dots\}$ and let $+$ be ordinary addition. Then $(A, +)$ is a magma but is not a monoid: the operation is associative but there is no identity element.

(c) Consider $(E, *)$ in Example A.4.2(e). Then $(E, *)$ is a magma but is not a monoid: there is an identity element but the operation is not associative (see Exercise A.4.1).

(d) Consider the set $E = \{1, 2, 3, 4\}$ and the operation $*$ defined by the table:

$*$	1	2	3	4
1	1	2	3	4
2	2	1	4	3
3	3	3	3	3
4	4	4	4	4

So $(E, *)$ is a magma and $(E, *)$ has an identity element (namely, 1). One can also check that $*$ is associative (this takes a bit of work; just take it for granted). So $(E, *)$ is a monoid, and in fact it is a *non-commutative* monoid.

(e) In MAT 1341, you learned how to multiply matrices. Let E be the set of all 2×2 matrices with entries in $\mathbb{R}$, and let $\cdot$ denote the multiplication of matrices. Then $\cdot$ is an operation on the set E (because the product of any two 2×2 matrices with entries in $\mathbb{R}$ is again a 2×2 matrix with entries in $\mathbb{R}$), and consequently $(E, \cdot)$ is a magma. Moreover,

- the element $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ of the set E is an identity element of this magma;
- in $(E, \cdot)$, the operation is associative (matrix product is associative).

So $(E, \cdot)$ is a monoid; in fact it is a non-commutative monoid. What are the invertible elements in this monoid?

Note that, in any monoid, the concept of invertibility makes sense (because it makes sense in any magma which has an identity element).

Theorem A.5.3. *In a monoid, a given element can have at most one inverse.*

Proof. Let $(E, \star)$ be a monoid and let e be its identity element. Consider $a \in E$ and suppose that $x_1, x_2 \in E$ are inverses of a; we have to show that $x_1 = x_2$.

As x_1, x_2 are inverses of a,

$$a \star x_1 = e = x_1 \star a \quad \text{and} \quad a \star x_2 = e = x_2 \star a.$$

It follows that

$$\begin{aligned} x_1 &= x_1 \star e && \text{(since } e \text{ is an identity element)} \\ &= x_1 \star (a \star x_2) && \text{(since } x_2 \text{ is an inverse of } a\text{)} \\ &= (x_1 \star a) \star x_2 && \text{(since } * \text{ is associative)} \\ &= e \star x_2 && \text{(since } x_1 \text{ is an inverse of } a\text{)} \\ &= x_2, && \text{(since } e \text{ is an identity element),} \end{aligned}$$

which proves that $x_1 = x_2$ and hence completes the proof. □

Suppose that $(E, *)$ is a monoid. If a is an invertible element then we know, by Theorem A.5.3, that a has exactly one inverse; so it makes sense to speak of *the* inverse of a (as opposed to *an* inverse of a).

(a) Suppose that the operation in our monoid is an addition; for instance our monoid could be $(\mathbb{R}, +)$ or $(\mathbb{Z}, +)$ or $(\mathbb{R}^n, +)$, etc. Then we say that our monoid is an *additive monoid* and we often use special notation:

- we write $(E, +)$ instead of $(E, \star)$
- the identity element is *usually* denoted 0
- if $a \in E$ is an invertible element then the inverse of a is *usually* denoted $-a$. If this notation is used then $-a$ is, by definition, the unique element of E satisfying $a + (-a) = 0 = (-a) + a$.

(b) If the operation in our monoid $(E, \star)$ is *not* an addition then the inverse of an element a is often denoted a^{-1}. Then, by definition of inverse, a^{-1} is the unique element of E which satisfies

$$a \star a^{-1} = e = a^{-1} \star a,$$

where e denotes the identity element of $(E, \star)$.

If f and g are functions from $\mathbb{R}$ to $\mathbb{R}$, we define a new function "$f+g$" from $\mathbb{R}$ to $\mathbb{R}$ by:

$$(f+g)(x) = f(x) + g(x), \quad \text{for all } x \in \mathbb{R}.$$

For a concrete example of addition of functions, suppose that:

- f is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by $f(x) = \frac{x-1}{x^4+1}$
- g is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by $g(x) = \frac{x+1}{x^4+1}$.

Then the definition of $f+g$ says that $(f+g)(2) = f(2)+g(2) = \frac{1}{17}+\frac{3}{17} = \frac{4}{17}$. More generally, the definition of $f+g$ says that for each $x \in \mathbb{R}$ we have

$$(f+g)(x) = f(x) + g(x) = \frac{x-1}{x^4+1} + \frac{x+1}{x^4+1} = \frac{2x}{x^4+1}$$

so $f+g$ is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by

$$(f+g)(x) = \frac{2x}{x^4+1}, \quad \text{for all } x \in \mathbb{R}.$$

Let $\mathcal{F}(\mathbb{R})$ denote the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Then the addition of functions that we just defined is an operation on the set $\mathcal{F}(\mathbb{R})$, i.e., any $f, g \in \mathcal{F}(\mathbb{R})$ determine a well-defined element $f+g \in \mathcal{F}(\mathbb{R})$. So $(\mathcal{F}(\mathbb{R}), +)$ is a magma; let us argue that it is a commutative monoid. Before going into this argument, it is useful to recall what it means for two functions to be equal.

Definition A.5.4 (Equality of functions)**.** Suppose that f, g are functions from $\mathbb{R}$ to $\mathbb{R}$. We say that f and g are equal, and write $f = g$, if for all $x \in \mathbb{R}$, the *real numbers* $f(x)$ and $g(x)$ are equal.

Let us use this definition to show that addition of functions is commutative. Let f, g be functions from $\mathbb{R}$ to $\mathbb{R}$. Then $f+g$ and $g+f$ are functions from $\mathbb{R}$ to $\mathbb{R}$ and, in order to prove that they are equal, we have to show that for each $x \in \mathbb{R}$, the real numbers $(f+g)(x)$ and $(g+f)(x)$ are equal. Now, for any given $x \in \mathbb{R}$, we have

$$(f+g)(x) = f(x) + g(x) = g(x) + f(x) = (g+f)(x)$$

where the equality in the middle is true because addition *of real numbers* is commutative, and the other two equalities are true by definition of addition of functions. This proves that $f+g = g+f$, so addition of functions is a commutative operation on the set $\mathcal{F}(\mathbb{R})$. One can imitate the above argument to show that addition of function is associative (Exercise A.5.2).

Let $\mathbf{0}$ denote the function from $\mathbb{R}$ to $\mathbb{R}$ which is identically equal to zero:

$$\mathbf{0}(x) = 0, \quad \text{for all } x \in \mathbb{R}.$$

Then $\mathbf{0} \in \mathcal{F}(\mathbb{R})$. One can show (Exercise A.5.3) that $\mathbf{0}$ is an identity element of $\mathcal{F}(\mathbb{R})$. We conclude that $(\mathcal{F}(\mathbb{R}), +)$ is a commutative monoid.

Theorem A.5.5. *If a, b are invertible elements in a monoid $(E, \star)$ then $a \star b$ is invertible and $(a \star b)^{-1} = b^{-1} \star a^{-1}$.*

Proof. Let $(E, \star)$ be a monoid, with identity element e.

Suppose that a, b are invertible elements of $(E, \star)$. Let $v = b^{-1} \star a^{-1}$ and note that v exists and is an element of E. Let us compute $(a \star b) \star v$ and $v \star (a \star b)$; remember that $\star$ is associative, so parentheses can be moved at will, or even omitted:

$$\begin{aligned}(a \star b) \star v &= (a \star b) \star (b^{-1} \star a^{-1}) = a \star (b \star b^{-1}) \star a^{-1} = a \star e \star a^{-1} \\ &= (a \star e) \star a^{-1} = a \star a^{-1} = e, \\ v \star (a \star b) &= (b^{-1} \star a^{-1}) \star (a \star b) = b^{-1} \star (a^{-1} \star a) \star b = b^{-1} \star e \star b \\ &= (b^{-1} \star e) \star b = b^{-1} \star b = e.\end{aligned}$$

So $(a \star b) \star v = e = v \star (a \star b)$; this proves that $a \star b$ is invertible and that its inverse is v. □

Exercises.

A.5.1. Find all invertible elements of the monoid of Example A.5.2(d). Also find an inverse of each invertible element.

A.5.2. Prove that addition of functions is associative.

A.5.3. Show that $\mathbf{0} + f = f = f + \mathbf{0}$ for all $f \in \mathcal{F}(\mathbb{R})$. *Hint*: $\mathbf{0} + f$ and f are functions from $\mathbb{R}$ to $\mathbb{R}$; to prove that they are equal, use Definition A.5.4.

A.5.4. Show that, in the monoid $(\mathcal{F}(\mathbb{R}), +)$, each element is invertible. Following the conventions for additive monoids, the inverse of an element f is denoted $-f$. (Here we are talking about the additive inverse of f; this has nothing to do with the concept of inverse function.)

A.5.5. Consider the magma $(\mathbb{R}^2, \star)$ where the operation $\star$ is defined by

$$(x, y) \star (x', y') = (xx', yy') \quad \text{for all } (x, y), (x', y') \in \mathbb{R}^2.$$

(a) Is this magma a monoid?

(b) What are the invertible elements of $(\mathbb{R}^2, \star)$? Show that $(2, 3)$ is an invertible element of $(\mathbb{R}^2, \star)$ and find all the inverses of $(2, 3)$. Make sure you justify that you have found them all.

A.5.6. Define a new operation $\oplus$ on the set $\mathbb{R}$ by $x \oplus y = x + y + 3$, where the "+" in the right hand side is the ordinary addition of numbers. For instance, $7 \oplus 5 = 15$. Then $(\mathbb{R}, \oplus)$ is a magma.

(a) Check that the real number 0 is *not* an identity element of $(\mathbb{R}, \oplus)$.

(b) Find a real number e which is an identity element of $(\mathbb{R}, \oplus)$ (then, by Theorem A.3.4, e is the unique identity element of $(\mathbb{R}, \oplus)$). Note that, even though $\oplus$ looks like an addition, it would be a bad idea to denote the identity element of $(\mathbb{R}, \oplus)$ by the symbol 0, because then 0 would denote two different numbers (the real number zero and the identity of the monoid $(\mathbb{R}, \oplus)$). So let e denote the identity element of $(\mathbb{R}, \oplus)$.

(c) Let x, y, z be arbitrary real numbers. Compute the real number $(x \oplus y) \oplus z$; then compute the real number $x \oplus (y \oplus z)$; then check that you obtained the same answer in the two calculations. This argument shows that $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ for all $x, y, z \in \mathbb{R}$, so you have just proved that $\oplus$ is associative. In view of (b), you can conclude that $(\mathbb{R}, \oplus)$ is a monoid. Also prove that $\oplus$ is commutative (compute $x \oplus y$ and $y \oplus x$ separately, and see that you get the same number in the two cases) so $(\mathbb{R}, \oplus)$ is a commutative monoid.

(d) Show that 5 is an invertible element of $(\mathbb{R}, \oplus)$ and find its inverse. (Let us use the notation $\overline{5}$ for the inverse of 5 in $(\mathbb{R}, \oplus)$.)

(e) Show that in the monoid $(\mathbb{R}, \oplus)$, every element is invertible. Find the inverse $\overline{a}$ of each element a of the monoid. What is $\overline{(-3)}$? What is $\overline{0}$?

(f) If we think of $\oplus$ as being an addition, then we would like to define a new operation $\ominus$ on $\mathbb{R}$ which would act like a subtraction. The idea is that subtracting b should be equivalent to adding the inverse of b. So we define a new operation $\ominus$ on $\mathbb{R}$ by: given $a, b \in \mathbb{R}$, $a \ominus b = a \oplus \overline{b}$. Compute $0 \ominus 5$ and $5 \ominus 0$ (careful! this is confusing). Note that $(\mathbb{R}, \ominus)$ is a magma, but $\ominus$ is not commutative, not associative, and there is no identity element. Show that the solution x to the equation $x \oplus a = b$ is $x = b \ominus a$ (be careful; show that $x = b \ominus a$ is a solution to the given equation, and show that it is *the only* solution).

A.6 Fields

Definition A.6.1 (Field)**.** A *field* is a triple $(F, +, \cdot)$ where F is a set, $+$ and $\cdot$ are two binary operations on F, called *addition* and *multiplication*, respectively, and the following conditions are satisfied:

(A1) For all $a, b \in F$, we have $a + b = b + a$. (*commutativity of addition*)

(A2) For all $a, b, c \in F$, we have $(a + b) + c = a + (b + c)$. (*associativity of addition*)

(A3) There is an element $0 \in F$ such that, for all $a \in F$, $a + 0 = 0 + a = a$. The element 0 is unique and is called the *additive identity*.

(A4) For any $a \in F$, there exists an element $-a \in F$ such that $a + (-a) = 0$. The element $-a$ is uniquely determined by a and is called the *additive inverse* of a.

(M1) For all $a, b \in F$, we have $a \cdot b = b \cdot a$. (*commutativity of addition*)

(M2) For all $a, b, c \in F$, we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. (*associativity of multiplication*)

(M3) There is a *nonzero* element $1 \in F$ such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in F$. The element 1 is unique and is called the *multiplicative identity*.

(M4) For any $a \in F$, $a \neq 0$, there exists an element $a^{-1} \in F$ such that $aa^{-1} = 1$. The element a^{-1} is uniquely determined by a and is called the *multiplicative inverse* of a.

(AM1) For any $a, b, c \in F$, we have $(a + b) \cdot c = a \cdot c + b \cdot c$. (*distributivity*)

We often denote multiplication by juxtaposition. For example, if $a, b \in F$, then ab means $a \cdot b$. We also sometimes say "F is a field" (instead of "$(F, +, \cdot)$ is a field") when the two operations (addition and multiplication) are clear from the context.

Remark A.6.2. (a) Suppose $(F, +, \cdot)$ is a field. Definition A.6.1 implies that $(F, +)$ and $(F, \cdot)$ are both *commutative monoids*: sets with an associative operation and an identity element (see Section A.5). In any monoid, identity elements are unique and invertible elements have exactly one inverse (see Theorem A.5.3). This justifies the uniqueness claims made in Definition A.6.1.

(b) Our assumption in (M3) that $1 \neq 0$ is crucial, as we will see in Remark A.6.9.

Remark A.6.3. A set R with two operations $+$ and $\cdot$ satisfying all of the axioms of Definition A.6.1 except possibly (M1) and (M4) is called a *ring*. If it also satisfies (M1), it is a *commutative ring*. So a field is a commutative ring in which every nonzero element has a multiplicative inverse.

Note that distributivity and commutativity of multiplication imply

$$(a + b)c = ac + bc \quad \text{for all} \quad a, b, c \in F$$

since

$$(a + b)c = c(a + b) = ca + cb = ac + bc.$$

Thus, we have distributivity in both directions.

If F is a field, we can view any integer as an element of F. First, we view the integer zero as the $0 \in F$. Then, for $n > 0$, we have

$$n = \underbrace{1 + 1 + \cdots + 1}_{n \text{ summands}} \in F, \tag{A.1}$$

where 1 here is the multiplicative identity of F. For $n < 0$, we then define

$$n = -(-n) = -(\underbrace{1 + 1 + \cdots + 1}_{-n \text{ summands}}) \in F. \tag{A.2}$$

Then, for instance, if $a \in F$, we can write expressions like

$$5a = (1 + 1 + 1 + 1 + 1)a = a + a + a + a + a.$$

Much of the algebra that we have done in $\mathbb{R}$ or $\mathbb{C}$ can be done in an arbitrary field. However, when working in an arbitrary field, you should be careful that you are only using the properties guaranteed by Definition A.6.1. In particular, watch out for the following:

- In an arbitrary field, you should not use inequalities. The expression $a < b$ does not make sense for a, b in an arbitrary field F, even though it makes sense when $F = \mathbb{R}$. Even when $F = \mathbb{C}$, there is no good notion of order.
- Even if n and m are distinct integers, they can be equal when considered as elements of a field F as in (A.1) and (A.2). For example $5 = 0$ in $\mathbb{F}_5$.

Examples A.6.4. (a) We see that $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{Q}$ are fields, with the usual addition and multiplication.

(b) The integers $\mathbb{Z}$ (with the usual addition and multiplication) is *not* a field because it contains nonzero elements that do not have multiplicative inverses (for instance, 2).

(c) The set $M_n(\mathbb{R})$ of $n \times n$ matrices (with real number entries) with matrix addition and multiplication, is *not* a field for $n \geq 2$ because there are nonzero matrices (the zero matrix is the additive identity) that do not have multiplicative inverses. For example, the matrix

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$$

is not the zero matrix but is still not invertible since its determinant is zero. However $M_n(\mathbb{R})$ is a ring (see Remark A.6.3).

For the remainder of this section, F is a field.

Lemma A.6.5 (Cancellation laws)**.** *For all $a, b, c \in F$, the following statements hold:*

(a) If $a + b = a + c$, then $b = c$.

(b) If $ab = ac$ and $a \neq 0$, then $b = c$.

Proof. (a) Suppose $a, b, c \in F$. Then

$$\begin{aligned} & a + b = a + c \\ \implies & (-a) + a + b = (-a) + a + c \\ \implies & 0 + b = 0 + c && \text{(by (A4))} \\ \implies & b = c. && \text{(by (A3))} \end{aligned}$$

(b) The proof of this part is left as an exercise (Exercise A.6.3).

□

Example A.6.6. In the field $\mathbb{R}$ the equality $0 \cdot 2 = 0 \cdot 3$ holds, but $2 \neq 3$. This does not violate Lemma A.6.5(b) because of the requirement $a \neq 0$ there.

Example A.6.7. Let $M_2(\mathbb{R})$ be the set of 2×2 matrices with real number entries. We have operations of matrix addition and matrix multiplication on $M_2(\mathbb{R})$. Suppose

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad X' = \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then

$$AX = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = AX',$$

but $X \neq X'$. The problem here is that $M_2(\mathbb{R})$ is not a field (see Example A.6.4(c)). In particular, A does not have a multiplicative inverse.

Proposition A.6.8. *Let F be a field.*

(a) $0x = 0$ *for all* $x \in F$.

(b) $(-1)x = -x$ *for all* $x \in F$.

(c) $(-a)b = -(ab) = a(-b)$ *for all* $a, b \in F$.

Proof. (a) Let $x \in F$. Then

$$0x + 0x = (0 + 0)x = 0x = 0x + 0.$$

Then we have $0x = 0$ by Lemma A.6.5(a).

(b) Let $x \in F$. Then

$$x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0.$$

This implies that $(-1)x$ is the additive inverse of x, hence $(-1)x = -x$.

(c) Let $a, b \in F$. Then

$$(-a)b = ((-1)a)b = (-1)(ab) = -(ab)$$

and

$$a(-b) = a((-1)b) = ((-1)b)a = (-1)(ba) = (-1)(ab) = -(ab).$$

□

Remark A.6.9. We now see why the assumption $1 \neq 0$ in (M3) is so important. If $1 = 0$ then, for all $a \in F$, we have

$$a = 1a = 0a = 0.$$

Thus F has only the zero element, and thus is not very interesting!

Proposition A.6.10. *Suppose* x, y *are elements of a field* F. *If* $x \neq 0$ *and* $y \neq 0$, *then* $xy \neq 0$.

Proof. The proof of this proposition is left as an exercise (Exercise A.6.5). □

Definition A.6.11 (Subtraction)**.** Suppose F is a field. We define the operation $-$ of subtraction on F by

$$a - b = a + (-b), \quad \text{for all} \quad a, b \in F.$$

Exercises.

A.6.1. Verify directly that $\mathbb{F}_2$, as defined in Example 1.1.3 is a field, using Definition A.6.1.

A.6.2. Define an addition and multiplication on $\mathbb{R}^2$ by

$$(x, y) + (x', y') = (x + x', y + y') \quad \text{and} \quad (x, y)(x', y') = (xx', yy'), \quad \text{for } (x, y), (x', y') \in \mathbb{R}^2.$$

Is $(\mathbb{R}^2, +, \cdot)$ a field?

A.6.3. Prove Lemma A.6.5(b).

A.6.4. Show that the element 0 in a field F does not have a multiplicative inverse (*Hint:* Use Proposition A.6.8(a) and the fact that $1 \neq 0$, as guaranteed by (M3).) Since, by the definition of a field, all nonzero elements have a multiplicative inverse, this shows that the set of elements of a field F with a multiplicative inverse is *exactly* $F^\times$.

A.6.5. Prove Proposition A.6.10.

A.6.6. Suppose a, b, c are elements of a field F. Show that

(a) $a - a = 0$,

(b) $a(b - c) = ab - ac$, and

(c) $(a - b)c = ac - bc$.

A.6.7. Suppose that F is a field containing an element c such that $c \neq 0$ and $c + c = 0$. Show that $1 + 1 = 0$ in this field.

A.6.8. Show that if F is a field and $x, y \in F$ satisfy $x \neq 0$ and $y \neq 0$, then $xy \neq 0$.

A.6.9. Suppose $n \in \mathbb{N}_+ = \{1, 2, 3, \dots\}$ is not prime. Let $\mathbb{F}_n = \{0, 1, 2, \dots, n - 1\}$ and define addition and multiplication as follows: For $a, b \in \mathbb{F}_n$,

$$\begin{aligned} a + b &= \text{ remainder after dividing } a + b \text{ by } n, \\ a \cdot b &= \text{ remainder after dividing } ab \text{ by } n. \end{aligned}$$

Prove that $\mathbb{F}_n$ is not a field. *Hint*: Use Exercise A.6.8.

Appendix B

Quotient spaces and the First Isomorphism Theorem

In this optional appendix we discuss the idea of a quotient vector space. Quotient vector spaces are vectors spaces whose elements are equivalence classes under a certain equivalence relation. We recall the idea of an equivalence relation, define quotient vector spaces, and then prove the important First Isomorphism Theorem. This gives us alternative method for proving the Dimension Theorem (Theorem 3.5.1).

B.1 Equivalence relations and quotient sets

Definition B.1.1 (Equivalence relation)**.** Let X be a nonempty set. Suppose that for each ordered pair (x, y) of elements of X, we are given a statement $S(x, y)$ about x and y. We write $x \sim y$ if the statement $S(x, y)$ is true. We say that $\sim$ is an *equivalence relation* on X if the following three conditions hold:

(a) *reflexivity*: $x \sim x$ for all $x \in X$,

(b) *symmetry*: if $x \sim y$ then $y \sim x$,

(c) *transitivity*: if $x \sim y$ and $y \sim z$, then $x \sim z$.

If $x \sim y$, we say that x is *equivalent* to y (under the relation $\sim$).

Example B.1.2 (Congruence modulo n). Fix a positive integer n and consider the set $\mathbb{Z}$ of integers. For $x, y \in \mathbb{Z}$, let $S(x, y)$ be the statement

$$\text{“}(x - y) \text{ is an integral multiplie of } n.\text{”}$$

Then $x \sim y$ if and only if $x - y$ is an integrable multiple of n. In other words, $x \sim y$ if $(x - y) = kn$ for some $k \in \mathbb{Z}$. If we write $n\mathbb{Z} = \{kn \mid k \in \mathbb{Z}\}$ for the set of integral multiples of n, then we have

$$x \sim y \iff (x - y) \in n\mathbb{Z}.$$

Let's check that $\sim$ is an equivalence relation on the set $\mathbb{Z}$.

- *Symmetry*: For all $x \in \mathbb{Z}$, we have $x - x = 0 \in n\mathbb{Z}$, so $x \sim x$.

- *Reflexivity*: Suppose $x \sim y$ for some $x, y \in \mathbb{Z}$. Then $x - y \in n\mathbb{Z}$. But then $y - x = -(x - y) \in n\mathbb{Z}$ (since if $x - y = kz$ for some integer k, then $y - x = (-k)z$, which is also an integer multiple of n) and so $y \sim x$.
- *Transitivity*: Suppose $x \sim y$ and $y \sim z$ for some $x, y, z \in \mathbb{Z}$. Then $x - y = k_1 n$ for some $k_1 \in \mathbb{Z}$ and $y - z = k_2 n$ for some $k_2 \in \mathbb{Z}$. But then

$$x - z = (x - y) + (y - z) = k_1 n + k_2 n = (k_1 + k_2)n \in n\mathbb{Z},$$

since $k_1 + k_2 \in \mathbb{Z}$.

Thus $\sim$ is an equivalence relation on $\mathbb{Z}$. This equivalence relation has its own special notation. If $x \sim y$, we write

$$x \equiv y \mod n,$$

and say x is *congruent to y modulo n*.

Example B.1.3. Suppose A and B are sets and $f\colon A \to B$ is any function (map of sets). For $x, y \in A$, write $x \sim y$ if $f(x) = f(y)$. Let's check that $\sim$ is an equivalence relation.

(a) *Reflexive:* For all $x \in A$, we have $x \sim x$ since $f(x) = f(x)$.

(b) *Symmetric:* If $x, y \in A$ such that $x \sim y$, then $f(x) = f(y)$. Hence $f(y) = f(x)$ and so $y \sim x$.

(c) *Transitive:* Suppose $x, y, z \in A$ such that $x \sim y$ and $y \sim z$. Then $f(x) = f(y) = f(z)$ and so $x \sim z$.

Example B.1.4. Suppose M is a subspace of a vector space V. For $u, v \in V$, write $u \sim v$ if $u - v \in M$. Then $\sim$ is an equivalence relations on V. We check the three properties of an equivalence relation.

(a) *Reflexive:* For all $v \in V$, we have $v - v = \mathbf{0} \in M$, since M is a subspace.

(b) *Symmetric:* For all $u, v \in V$ such that $u \sim v$, we have $u - v \in M$. Then

$$v - u = -(u - v) = (-1)(u - v) \in M$$

since M is a subspace and hence closed under scalar multiplication.

(c) *Transitive:* Suppose $u, v, w \in V$ such that $u \sim v$ and $v \sim w$. Then $u - v \in M$ and $v - w \in M$. Then

$$u - w = (u - v) + (v - w) \in M$$

since M is a subspace and hence closed under vector addition.

Definition B.1.5 (Equivalence class)**.** Suppose $\sim$ is an equivalence relation on a set X. Then, for $x \in X$, we define

$$[x] = \{y \in X \mid x \sim y\}$$

to be the set of all elements of X that are equivalent to x. We call $[x]$ the *equivalence class* of x.

Example B.1.6. Consider the equivalence relation on $\mathbb{Z}$ given by congruence modulo n (Example B.1.2). Then, for $a \in \mathbb{Z}$, we have

$$[a] = \{b \in \mathbb{Z} \mid a - b \in n\mathbb{Z}\} = \{a + kn \mid k \in \mathbb{Z}\} := a + n\mathbb{Z}.$$

Example B.1.7. Suppose A, B are sets, $f\colon A \to B$ is any function, and $\sim$ is the equivalence relation of Example B.1.3. Then, for $x \in A$,

$$[x] = \{y \in A \mid f(y) = f(x)\} = f^{-1}(\{f(x)\}).$$

Example B.1.8. Suppose M is a subspace of a vector space V and $\sim$ is the equivalence relation of Example B.1.4. Then, for $v \in V$,

$$[v] = \{v + z \mid z \in M\}.$$

To prove this, suppose that $w \in [v]$. Then, by definition, $v \sim w$ and so $v - w \in M$. That is, $v - w = z$ for some $z \in M$. Thus $w = v + (-z)$. Therefore $[v] \subseteq \{v + z \mid z \in M\}$.

Now suppose $w = v + z$ for some $z \in M$. Then $v - w = -z = (-1)z \in M$, since M is a subspace. Hence $v \sim w$ and so $w \in [v]$. Hence, $\{v + z \mid z \in M\} \subseteq [v]$.

Definition B.1.9 (Coset)**.** For a subspace M of a vector space V and $v \in V$, we write

$$v + M = \{v + z \mid z \in M\}$$

and call this the *coset of v modulo M*.

Example B.1.10. Define a linear form $f\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = 2x_1 - x_2 \quad \forall\ (x_1, x_2) \in \mathbb{R}^2.$$

Let

$$M = \operatorname{Ker} f = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = 0\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = 0\}.$$

So M is a line through the origin. Note that the range of f is $\mathbb{R}$ since, for instance, for any $x \in \mathbb{R}$, we have $f(0, -x) = x$.

For $c \in \mathbb{R}$, we have

$$f^{-1}(\{c\}) = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = c\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = c\},$$

which is a line parallel to M (and equal to M if and only if $c = 0$).

Consider the equivalence relation of Example B.1.3. Under this equivalence relation, for $x, y \in \mathbb{R}^2$, we have

$$x \sim y \iff f(x) = f(y) \iff f(x) - f(y) = 0 \iff f(x - y) = 0 \iff x - y \in \operatorname{Ker} f = M,$$

where we have used the fact that f is linear. Thus we see that this equivalence relation agrees with the one of Example B.1.4.

For any $v \in \mathbb{R}^2$,

$$[v] = v + M = f^{-1}(\{f(v)\}) = \{x \in \mathbb{R}^2 \mid 2x_1 - x_2 = f(v)\}.$$

This is the line in $\mathbb{R}^2$ passing through v and parallel to M. For example, if $v = (3, 1)$, then $[v]$ is the plane with equation

$$2x_1 - x_2 = 2(3) - (1) = 5.$$

We see that the equivalence relation 'decomposes' the plane into a set of parallel lines. Each point of the plane $\mathbb{R}^2$ lies on exactly one of these lines.

Definition B.1.11 (Partition)**.** Suppose X is a nonempty set. A *partition* of X is a collection $\mathscr{A}$ of subsets of X satisfying the following properties:

(a) every $A \in \mathscr{A}$ is a *nonempty* subset of X,

(b) if $A, B \in \mathscr{A}$ and $A \neq B$, then $A \cap B = \varnothing$,

(c) every element of X belongs to one of the subsets in $\mathscr{A}$ (in other words, for each $x \in X$, there is some $A \in \mathscr{A}$ such that $x \in A$).

Remark B.1.12. The third property of a partition says that each element of X belongs to one of the subsets of the partition. The second property says that no element lies in more than one of the subsets of the partition. Therefore, combining these two facts, we see that the fundamental property of a partition is that every element of X belongs to *exactly one* of the subsets in the partition.

In turns out that equivalence relations and partitions are simply two different ways of thinking about the same thing. The precise relationship is given in the following theorem.

Theorem B.1.13. *Let X be a nonempty set.*

(a) If $\sim$ is an equivalence relation on X, then the set of equivalence classes is a partition of X.

(b) Suppose $\mathscr{A}$ is a partition of X. Write $x \sim y$ if x and y belong to the same element of $\mathscr{A}$ (that is, there exists an $A \in \mathscr{A}$ such that $x, y \in A$). Then $\sim$ is an equivalence relation on X.

Proof. (a) Suppose $\sim$ is an equivalence relation on X and let $\mathscr{A}$ be the set of equivalence classes. We want to show that $\mathscr{A}$ satisfies the conditions of Definition B.1.11.

- First of all, for any equivalence class $[x]$, we have $x \in [x]$ (since $x \sim x$ by the reflexivity of an equivalence relation). Therefore, all the equivalence classes are nonempty.
- Suppose $[x]$ and $[y]$ are two equivalence classes and $[x] \neq [y]$. We want to show that $[x] \cap [y] = \varnothing$. We prove this by contradiction. Suppose $[x] \cap [y] \neq \varnothing$. Then we can choose some element $z \in [x] \cap [y]$. We will show that this implies that $[x] = [y]$. Let $w \in [x]$. Then $x \sim w$. Since $z \in [x] \cap [y]$, we have $z \in [x]$ and $z \in [y]$. Thus $x \sim z$ and $y \sim z$. By symmetry, $z \sim x$. Thus

$$y \sim z \sim x \sim w.$$

By transitivity, $y \sim w$ and so $w \in [y]$. Hence $[x] \subseteq [y]$. Repeating the above argument but interchanging the roles of x and y shows that $[y] \subseteq [x]$. Hence $[x] = [y]$. This contradicts the assumption that $[x] \neq [y]$. Therefore, $[x] \cap [y] = \varnothing$.

- The last property is easy since for every $x \in X$, we have $x \in [x]$ and so every element of x belongs to some element of $\mathscr{A}$.

(b) Now assume that $\mathscr{A}$ is a partition of X and define $x \sim y$ if there is some $A \in \mathscr{A}$ such that $x, y \in A$. We wish to verify that $\sim$ is an equivalence relation on X.

- *Reflexivity*: For any $x \in X$, we have $x \in [x]$. Thus $x \sim x$.
- *Symmetry:* Suppose $x, y \in X$ and $x \sim y$. Thus, there is some $A \in \mathscr{A}$ such that $x, y \in A$. Thus $y \sim x$ as well.
- *Transitivity:* Suppose $x, y, z \in X$, $x \sim y$, and $y \sim z$. Then there are $A, B \in \mathscr{A}$ such that $x, y \in A$ and $y, z \in B$. This implies in particular that $y \in A \cap B$. Thus $A \cap B \neq \varnothing$. Hence, by the definition of a partition, we have $A = B$. Therefore, x and z belong to the same element $A = B$ of $\mathscr{A}$. Hence $x \sim z$.

□

From this theorem, we can deduce some properties of equivalence classes of relations.

Corollary B.1.14. *Suppose $\sim$ is an equivalence relation on a nonempty set X.*

(a) For any $x, y \in X$, we have either $[x] = [y]$ or $[x] \cap [y] = \varnothing$.

(b) For $x, y \in X$, we have $x \sim y$ if and only if $[x] = [y]$.

Proof. (a) This follows immediately from Theorem B.1.13, which says that the equivalence classes form a partition.

(b) If $x \sim y$, then $y \in [x]$. We also have $y \in [y]$. Thus $[x] \cap [y] \neq \varnothing$ and so $[x] = [y]$ by the first part of the corollary. Conversely, if $[x] = [y]$, then $y \in [x]$ (since $y \in [y]$). Hence $x \sim y$.

□

We can apply this corollary to the equivalence relation of Example B.1.4 to get the following result.

Corollary B.1.15. *If M is a subspace of a vector space V, then*

(a) for any $x, y \in V$, we have $x + M = y + M$ or $(x + M) \cap (y + M) = \varnothing$.

(b) $x - y \in M$ if and only if $x + M = y + M$, and

Definition B.1.16 (Quotient set)**.** Suppose $\sim$ is an equivalence relation on a set X. Then the set of equivalence classes

$$\{[x] \mid x \in X\}$$

is called the *quotient set* of X for the relation $\sim$, and is denoted $X/\sim$. In the special case where M is a subspace of V and the equivalence relation is the one of Example B.1.4, the quotient set is denoted V/M. Thus

$$V/M = \{x + M \mid x \in V\}.$$

Example B.1.17. Consider the situation of Example B.1.10, where $f\colon \mathbb{R}^2 \to \mathbb{R}$ is the linear form defined by

$$f(x_1, x_2) = 2x_1 - x_2 \quad \forall\ (x_1, x_2) \in \mathbb{R}^2,$$

and

$$M = \operatorname{Ker} f = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = 0\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = 0\}.$$

Then $\mathbb{R}^2/M$ is the set of lines parallel to M.

Definition B.1.18 (Quotient map)**.** Suppose $\sim$ is an equivalence relation on a set X. Then we have a surjective map

$$q\colon X \to X/\sim, \quad q(x) = [x],$$

called the *quotient map* of X onto $X/\sim$.

Remark B.1.19. Note that if $q\colon X \to X/\sim$ is the quotient map corresponding to some equivalence relation $\sim$ on a set X, then

$$q(x) = q(y) \iff [x] = [y] \iff x \sim y.$$

Thus, any equivalence relation is of the type seen in Example B.1.3.

Exercises.

B.1.1 ([Ber14, Ex. 2.5.2]). Let X be the set of all nonzero vectors in $\mathbb{R}^n$. In other words,

$$X = \{v \in \mathbb{R}^n \mid v \neq 0\}.$$

For $x, y \in X$, write $x \sim y$ if $x = cy$ for some scalar c. Prove that this is an equivalence relation on X.

B.2 Quotient vector spaces

In this course, the most important quotient set will be V/M for a subspace M of a vector space V. One reason for this is that, in this case, the quotient set is more than just a *set*: it is a vector space.

First, we need to define vector addition and scalar multiplication. Suppose V is a vector space over a field F and M is a subspace of V. For $x + M, y + M \in V/M$, we define their sum to be

$$(x + M) + (y + M) = (x + y) + M.$$

However, this definition should worry you. Why? Note that, in general, an element of V/M can be written in the form $x + M$ for more than one x. Since our definition of vector addition

refers explicitly to this x, our vector addition is not well-defined unless we get the result of our operation is independent of this choice of x. That is, if

$$x + M = x' + M \quad \text{and} \quad y + M = y' + M,$$

then it must be true that

$$(x + M) + (y + M) = (x' + M) + (y' + M).$$

By our definition of addition, this is true if and only if

$$(x + y) + M = (x' + y') + M.$$

Let's check this. Since $x+M = x'+M$, we have $x-x' \in M$. Similarly, since $y+M = y'+M$, we have $y - y' \in M$. Thus

$$(x + y) - (x' + y') = (x - x') + (y - y') \in M,$$

since M is a subspace of V. Hence $(x + y) + M = (x' + y') + M$ as desired.

We now define scalar multiplication on V/M by the formula

$$c(x + M) = cx + M, \quad c \in F,\ x \in V.$$

Again we need to check that this is well-defined. In particular, we need to check that if $x + M = x' + M$, then $cx + M = cx' + M$. To see this, note that $x + M = x' + M$ implies $x - x' \in M$. Thus $c(x - x') \in M$ since M is a subspace (and hence closed under scalar multiplication). Thus $cx + M = cx' + M$ as desired.

Now that we've defined an addition and scalar multiplication, we can prove the following theorem.

Theorem B.2.1. *Suppose V is a vector space over a field F and M is a subspace of V. Then, under the operations defined above, V/M is a vector space over F and the quotient map*

$$Q\colon V \to V/M, \quad Q(x) = x + M,$$

is a linear map with $\operatorname{Ker} Q = M$*. The vector space V/M is called the* quotient vector space *of V by the subspace M.*

Proof. We must check that our operations satisfy the axioms of Definition 1.2.1. For $x + M, y + M, z + M \in V/M$, we have

$$\begin{gathered}((x + M) + (y + M)) + (z + M) = ((x + y) + M) + (z + M) \\ = (x + y + z) + M = (x + M) + ((y + z) + M) = (x + M) + ((y + M) + (z + M)),\end{gathered}$$

so the addition is associative. Also

$$\begin{gathered}(x + M) + (y + M) = (x + y) + M = (y + x) + M = (y + M) + (x + M), \text{ and} \\ (x + M) + (\mathbf{0} + M) = (x + \mathbf{0}) + M = x + M.\end{gathered}$$

Thus the addition is commutative and $\mathbf{0} + M = M$ is an additive identity.

For all $x + M \in V/M$, we have

$$(x + M) + (-x + M) = (x - x) + M = \mathbf{0} + M,$$

and so every element of V/M has an additive inverse. We leave it as an exercise (Exercise B.2.1) to check the remaining axioms in the definition of a vector space.

To check that the quotient map Q is linear, suppose $u, v \in V$ and $c \in F$. Then

$$\begin{aligned} Q(u + v) &= (u + v) + M = (u + M) + (v + M) = Q(u) + Q(v), \\ Q(cv) &= cv + M = c(v + M). \end{aligned}$$

□

Exercises.

B.2.1. Complete the proof of Theorem B.2.1 by showing that V/M satisfies the remaining axioms of a vector space.

B.2.2 ([Ber14, Ex. 2.6.1]). Prove that if $V = M \oplus N$, then $V/M \cong N$. *Hint:* Restrict the quotient mapping $V \to V/M$ to N and find the kernel and image of the restricted mapping.

B.3 The First Isomorphism Theorem

Lemma B.3.1. *If $T\colon V \to W$ is a linear map, then*

$$T^{-1}(\{Tv\}) = v + \operatorname{Ker} T, \quad \textit{for all } v \in V.$$

Proof. Suppose $u \in T^{-1}(\{Tv\})$. Then

$$Tu = Tv \implies Tu - Tv = \mathbf{0} \implies T(u - v) = \mathbf{0} \implies u - v \in \operatorname{Ker} T \implies u \in v + \operatorname{Ker} T.$$

Now suppose $u \in v + \operatorname{Ker} T$. Then

$$u - v \in \operatorname{Ker} T \implies T(u - v) = \mathbf{0} \implies Tu - Tv = \{\mathbf{0}\} \implies Tu = Tv \implies u \in T^{-1}(\{Tv\}).$$

□

Theorem B.3.2 (First Isomorphism Theorem)**.** *Suppose $T\colon V \to W$ is a linear map and $N = \operatorname{Ker} T$. Then*

(a) N is a subspace of V,

(b) $T(V)$ is a subspace of W, and

(c) The map $V/N \to T(V)$ given by $v + N \mapsto Tv$ is an isomorphism. Thus, $V/N \cong T(V)$.

In other words,

$$V/\operatorname{Ker} T \cong \operatorname{Im} T$$

for every linear map T with domain V.

Proof. We've already proven (a) and (b) (see Corollary 2.2.4). So it remains to prove (c).

We need to find a bijective linear map $S\colon V/N \to T(V)$. Suppose $x \in V/N$. Then x is a subset of V and is equal to $v + N$ for some $v \in V$. By Lemma B.3.1, T is constant on the subset x. Thus we can define a map

$$S\colon V/N \to T(V), \quad S(v+N) = Tv.$$

Note that the map S is only well-defined by the comments above. We first show that S is linear. Suppose $x, y \in V/N$ and c is a scalar. Then $x = u + N$ and $y = v + N$ for some $u, v \in V$. Thus,

$$S(x+y) = S(u+v+N) = T(u+v) = Tu + Tv = S(u+N) + S(v+N) = S(x) + S(y),$$
$$S(cx) = S(cu+N) = T(cu) = cTu = cS(u+N) = cS(x).$$

It remains to show that S is bijective. It is clearly surjective since any element of $T(V)$ is of the form Tv for some $v \in V$ and we have $S(v+N) = Tv$. Since S is linear, we can show it is injective by proving that $\operatorname{Ker} S = \{\mathbf{0}\}$. Now, if $x = u + N \in \operatorname{Ker} S$, then

$$S(u+N) = \mathbf{0} \implies Tu = \mathbf{0} \implies u \in \operatorname{Ker} T \implies u + N = N,$$

which is the zero vector of V/N. □

Remark B.3.3. One of the main uses of the First Isomorphism Theorem is the following. If we want to prove that $V/U \cong W$ (where V and W are vector spaces and U is a subspace of V), then we can do this by finding a surjective linear map from V to W with kernel U.

Example B.3.4. Let

$$U = \{(x_1, x_2, x_3, x_4) \mid 2x_1 - x_2 = 0 \text{ and } x_1 + 3x_2 - 4x_3 - x_4 = 0\}.$$

Prove that $\mathbb{R}^4/U \cong \mathbb{R}^2$.

We need to find a surjective map $T\colon \mathbb{R}^4 \to \mathbb{R}^2$ such that $U = \operatorname{Ker} T$. Let

$$T(x_1, x_2, x_3, x_4) = (2x_1 - x_2, x_1 + 3x_2 - 4x_3 - x_4).$$

This is a linear map since it corresponds to multiplication by the matrix

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ 1 & 3 & -4 & -1 \end{bmatrix}.$$

It is surjective since for any $(a, b) \in \mathbb{R}^2$, we have

$$T\left(\frac{a}{2}, 0, 0, \frac{a}{2} - b\right) = (a, b).$$

(Recalling MAT 1341, showing that T is surjective is equivalent to showing that the above matrix has rank 2). It is clear that $\operatorname{Ker} T = U$. Thus $\mathbb{R}^4/U \cong \mathbb{R}^2$ by the First Isomorphism Theorem.

Example B.3.5. Let X be a set and Y be a subset of X. Let F be a field and define

$$W = \{f \in \mathcal{F}(X, F) \mid f(x) = 0 \text{ for all } x \in Y\}.$$

Let's prove that $\mathcal{F}(X, F)/W \cong \mathcal{F}(Y, F)$.

First of all, it only makes sense to write $\mathcal{F}(X, F)/W$ if W is a subspace of $\mathcal{F}(X, F)$. We leave it as an exercise (Exercise 1.5.9) to show that this is indeed the case.

Next, we want to find a surjective map $T\colon \mathcal{F}(X, F) \to \mathcal{F}(Y, F)$ such that $\operatorname{Ker} T = W$. Define T by

$$T(f) = f|_Y.$$

So T restricts a function to Y. This map is linear (Exercise B.3.1). It is clear that $\operatorname{Ker} T = W$. Also, T is surjective since, given any F-valued function f on Y, we can extend it to a function g on all of X by given it any values we want on points of $X \setminus Y$. Then $T(g) = f$. Thus, by the First Isomorphism Theorem, we have $\mathcal{F}(X, F)/W \cong \mathcal{F}(Y, F)$.

Exercises.

B.3.1. Prove that the map T of Example B.3.5 is linear.

B.3.2. (a) Let $S\colon U \to V$ and $T\colon V \to W$ be linear maps. Show that $\operatorname{Ker}(TS) = S^{-1}(\operatorname{Ker} T)$.

(b) Let $S\colon V \to W$ be a surjective linear map and M a subspace of W. Show that $V/S^{-1}(M) \cong W/M$. *Hint:* Apply part (a) to $S\colon V \to W$ and $Q\colon W \to W/M$.

B.3.3 ([Ber14, Ex. 2.7.2]). Let M, N be subspaces of the vector spaces V, W (respectively), and let

$$T\colon V \times W \to (V/M) \times (W/N)$$

be the linear map defined by $T(x, y) = (x + M, y + N)$. Find the kernel of T and prove that

$$(V \times W)/(M \times N) \cong (V/M) \times (W/N).$$

B.3.4 ([Ber14, Ex. 2.7.4]). Let V be a vector space, $V \times V$ the product vector space, and

$$\Delta = \{(v, v) \mid v \in V\} \subseteq V \times V.$$

(The set Δ is called the *diagonal* of $V \times V$.) Prove that Δ is a subspace of $V \times V$ and that $(V \times V)/\Delta \cong V$. *Hint:* See Exercise 2.2.8.

B.4 Another proof of the Dimension Theorem

We now use quotient spaces to give an alternative proof of the Dimension Theorem (Theorem 3.5.1).

Theorem B.4.1. *Suppose V is a vector space and $v_1, \dots, v_n \in V$. Let $1 \leq r < n$, let $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, and let $Q\colon V \to V/M$ be the quotient map. Then the following conditions are equivalent:*

(a) $v_1, \dots, v_n$ are linearly independent in V,

(b) $Qv_1, \dots, Qv_r$ are independent in V/M and $v_{r+1}, \dots, v_n$ are linearly independent in V.

Proof. (a) $\implies$ (b): $v_{r+1}, \dots, v_n$ are clearly independent since $v_1, \dots, v_n$ are (any subset of an independent set is independent). So it remains to show that $Qv_1, \dots, Qv_n$ are independent. Suppose

$$c_1(Qv_1) + \cdots + c_r(Qv_r) = Q\mathbf{0}$$

(Recall that $Q\mathbf{0} = M$ is the zero vector of V/M.) We need to show that $c_1 = \cdots = c_r = 0$. Since Q is linear, we have

$$Q(c_1v_1 + \cdots + c_nv_n) = Q\mathbf{0},$$

and so

$$c_1v_1 + \cdots + c_nv_n \in \operatorname{Ker} Q = M.$$

Since $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, we thus have

$$c_1v_1 + \cdots + c_nv_n = c_{r+1}v_{r+1} + \cdots + c_nv_n,$$

for some scalars $c_{r+1}, \dots, c_n$. Then

$$c_1v_1 + \cdots + c_rv_r + (-c_{r+1})v_{r+1} + \cdots + (-c_n)v_n = \mathbf{0}.$$

Since $v_1, \dots, v_n$ are independent, we have that all the coefficients are zero.

(b) $\implies$ (a): Suppose

$$c_1v_1 + \cdots + c_nv_n = \mathbf{0}. \tag{B.1}$$

We want to show $c_1 = \cdots = c_n = 0$. Let

$$z = c_{r+1}v_{r+1} + \cdots + c_nv_n.$$

Then $z \in M = \operatorname{Ker} Q$. Therefore,

$$Q\mathbf{0} = c_1(Qv_1) + \cdots + c_r(Qv_r) + Qz = c_1(Qv_1) + \cdots + c_r(Qv_r).$$

Since $Qv_1, \dots, Qv_r$ are independent, we have $c_1 = \cdots = c_r = 0$. But then B.1 becomes

$$c_{r+1}v_{r+1} + \cdots + c_nv_n = \mathbf{0}.$$

Since $v_{r+1}, \dots, v_n$ are linearly independence, we have $c_{r+1} = \cdots = c_n = 0$. Hence all the c_i are zero. □

Theorem B.4.2. *Suppose V is a vector space and $v_1, \dots, v_n \in V$. Let $1 \leq r < n$, let $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, and let $Q\colon V \to V/M$ be the quotient map. The following conditions are equivalent.*

(a) $v_1, \dots, v_n$ generate V,

(b) $Qv_1, \dots, Qv_r$ generate V/M.

Proof. (a) $\implies$ (b): Suppose $u \in V/M$. We want to show that we can write u as a linear combination of $Qv_1, \dots, Qv_n$. Since Q is surjective, there is a $v \in V$ such that $u = Qv$. Since $v_1, \dots, v_n$ generate V, we have

$$v = c_1 v_1 + \cdots + c_n v_n$$

for some scalars $c_1, \dots, c_n$. We apply the linear map Q to both sides and use the fact that $Qv_{r+1} = \cdots = Qv_n = Q\mathbf{0}$ (since $M = \operatorname{Ker} Q$) to obtain

$$u = Qv = c_1 Qv_1 + \cdots + c_r Qv_r.$$

(b) $\implies$ (a): Suppose $v \in V$. We want to show that we can write v as a linear combination of $v_1, \dots, v_n$. Since $Qv_1, \dots, Qv_r$ generated V/M, we have

$$Qv = c_1(Qv_1) + \cdots + c_r(Qv_r)$$

for some scalars $c_1, \dots, c_r$. Thus

$$v - (c_1 v_1 + \cdots + c_r v_r) \in \operatorname{Ker} Q = M,$$

and so

$$v - (c_1 v_1 + \cdots + c_r v_r) = c_{r+1} v_{r+1} + \cdots + c_n v_n$$

for some scalars $c_{r+1}, \dots, c_n$. Therefore, $v = c_1 v_1 + \cdots + c_n v_n$. □

Theorem B.4.3. *If V is a vector space and M is a subspace of V, then the following conditions are equivalent:*

(a) V is finite dimensional,

(b) M and V/M are finite dimensional.

When these conditions hold, we have

$$\dim V = \dim(V/M) + \dim M. \tag{B.2}$$

Proof. (a) $\Rightarrow$ (b): Suppose V is finite dimensional. Then M is finite-dimensional by Theorem 3.4.21 and V/M is finite dimensional by applying Theorem 3.3.7 to the quotient map $Q\colon V \to V/M$.

(b) $\Rightarrow$ (a): If $M = V$, the implication is trivial. Furthermore, $V/M = V/V$ consists of the single coset $\mathbf{0} + V = V$ and so V/V is the zero vector space. Thus B.2 becomes

$$\dim V = 0 + \dim V,$$

which is obviously true.

If $M = \{\mathbf{0}\}$, then the quotient map $V \to V/M$ is a linear surjection with zero kernel. Thus it is an isomorphism and so $V \cong V/M$. Thus V is finite dimensional because V/M is. Furthermore, B.2 becomes

$$\dim V = \dim(V/\{\mathbf{0}\}) + 0 = \dim V + 0,$$

which is clearly true.

Now assume $M \neq \{\mathbf{0}\}$ and $M \neq V$. Choose a basis $x_1, \ldots, x_m$ of M and $u_1, \ldots, u_r$ of V/M (we can do this since, by assumption, M and V/M are finite dimensional). Choose $y_k \in V$ such that $u_k = y_k + M$, $1 \leq k \leq r$. Then, by Theorem B.4.2, the list

$$x_1, \ldots, x_m, y_1, \ldots, y_r$$

generates V. Furthermore, this list is independent by Theorem B.4.1. Therefore, it is a basis for V. It follows that V is finite dimensional and

$$\dim V = m + r = \dim M + \dim V/M. \qquad \square$$

We can now give an alternative proof of the Dimension Theorem.

Alternative proof of Theorem 3.5.1. Since V is finite dimensional, so is its image $T(V)$ (T is a surjection onto its image, and we call apply Theorem 3.3.7) and the subspace $\operatorname{Ker} T$ (by Theorem 3.4.21). By the First Isomorphism Theorem, $T(V) \cong V/\operatorname{Ker} T$. Thus

$$\dim T(V) = \dim(V/\operatorname{Ker} T).$$

Now, by Theorem B.4.3, we have $\dim T(V) = \dim V - \dim(\operatorname{Ker} T)$. The result follows. $\square$

Index

Bibliography

[Ber14] Sterling K. Berberian. *Linear algebra*. Dover Publications, Inc., Mineola, NY, 2014. Reprint of the 1992 original, with a new errata and comments.

[Sav] Alistair Savage. Mathematical reasoning & proofs. Notes for MAT 1362. Available at https://alistairsavage.ca/mat1362/notes/MAT1362-Mathematical_reasoning_and_proofs.pdf.

[Tre] Sergei Treil. *Linear algbra done wrong*. Available at http://www.math.brown.edu/~treil/papers/LADW/LADW.html.