Linear Algebra

David B. Leep
Department of Mathematics
University of Kentucky
Lexington KY 40506-0027

Contents

Introduction **2**

1 Vector Spaces **4**
1.1 Basics of Vector Spaces 4
1.2 Systems of Linear Equations 16
1.3 Appendix 22
1.4 Proof of Theorems 1.20 and 1.21 23

2 Linear Transformations **27**
2.1 Basic Results 27
2.2 The vector space of linear transformations $f : V \to W$ 32
2.3 Quotient Spaces 34
2.4 Applications 37

3 Matrix Representations of a Linear Transformation **44**
3.1 Basic Results 44
3.2 The Transpose of a Matrix 50
3.3 The Row Space, Column Space, and Null Space of a Matrix . 51
3.4 Elementary Matrices 54
3.5 Permutations and Permutation Matrices 56
3.6 The Rank of a Matrix 59
3.7 Systems of Linear Equations 63
3.8 Bases of Subspaces 65

4 The Dual Space **70**
4.1 Basic Results 70
4.2 Exact sequences and more on the rank of a matrix 72
4.3 The double dual of a vector space 77

4.4 Annihilators of subspaces 78
4.5 Subspaces of V and V^* 80
4.6 A calculation 82

5 Inner Product Spaces 84
5.1 The Basics 84
5.2 Orthogonality 87
5.3 The Matrix of an Inner Product on V 91
5.4 The Adjoint of a Linear Transformation 94
5.5 Normal Transformations 96

6 Determinants 107
6.1 The Existence of n-Alternating Forms 107
6.2 Further Results and Applications 115
6.3 The Adjoint of a matrix 117
6.4 The vector space of alternating forms 121
6.5 The adjoint as a matrix of a linear transformation 122
6.6 The Vandermonde Determinant and Applications 125

7 Canonical Forms of a Linear Transformation 133
7.1 Preliminaries 133
7.2 The Cayley-Hamilton Theorem 134
7.3 Eigenvectors, eigenvalues, diagonalizability 137
7.4 More results on T-Invariant Subspaces 140
7.5 Primary Decomposition 141
7.6 Further Decomposition 147

Introduction

These notes are for a graduate course in linear algebra. It is assumed that the reader has already studied matrix algebra or linear algebra, however, these notes are completely self-contained. The material is developed completely from scratch, but at a faster pace than a beginning linear algebra course. Less motivation is presented than in a beginning course. For example, it is assumed that the reader knows that vector spaces are important to study, even though many details about vector spaces might have been forgotten from an earlier course.

I have tried to present material efficiently without sacrificing any of the details. Examples are often given after stating definitions or proving theorems, instead of beforehand. There are many occasions when a second method or approach to a result is useful to see. I have often given a second approach in the text or in the exercises. I try to emphasize a basis-free approach to results in this text. Mathematically this is often the best approach, but pedagogically this is not always the best route to follow. For this reason, I have often indicated a second approach, either in the text or in the exercises, that uses a basis in order to make results more clear.

Linear algebra is most conveniently developed over an arbitrary field k. For readers not comfortable with such generality, very little is lost if one always thinks of k as the field of real numbers $\mathbb{R}$, or the field of complex numbers $\mathbb{C}$. It will be clearly pointed out in the text if particular properties of a field are used or assumed.

Chapter 1

Vector Spaces

1.1 Basics of Vector Spaces

We begin by giving the definition of a vector space and deriving the most basic properties. They are fundamental to all that follows. The main result of this chapter is that all finitely generated vector spaces have a basis and that any two bases of a vector space have the same cardinality. On the way to proving this result, we introduce the concept of subspaces, linear combinations of vectors, and linearly independent vectors. These results lead to the concept of the dimension of a vector space. We close the chapter with a brief discussion of direct sums of vector spaces.

Let k denote an arbitrary field. We begin with the definition of a vector space. Example 1 (just after Proposition 1.2) gives the most important example of a vector space.

Definition 1.1. *A* vector space V *over a field* k *is a nonempty set* V *together with two binary operations, called addition and scalar multiplication, which satisfy the following ten axioms.*

1. *Addition is given by a function*

$$\begin{array}{rcl} V \times V & \to & V \\ (v, w) & \longmapsto & v + w. \end{array}$$

2. $v + w = w + v$ *for all* $v, w \in V$.

3. $(v + w) + z = v + (w + z)$ *for all* $v, w, z \in V$.

4. *There exists an element* $0 \in V$ *such that* $v + 0 = v$ *for all* $v \in V$.

5. *Given* $v \in V$, *there exists* $w \in V$ *such that* $v + w = 0$.

6. *Scalar multiplication is given by a function*

$$\begin{array}{rcl} k \times V & \to & V \\ (a, v) & \longmapsto & av. \end{array}$$

7. $a(v + w) = av + aw$ *for all* $a \in k,\ v, w, \in V$.

8. $(a + b)v = av + bv$ *for all* $a, b \in k,\ v \in V$.

9. $a(bv) = (ab)v$ *for all* $a, b \in k,\ v \in V$.

10. $1v = v$ *for all* $v \in V$, *where* 1 *is the multiplicative identity of* k.

Axioms (1)-(5) state that V is an abelian group under the operation of addition. The following result gives some simple consequences of axioms (1)-(5).

Proposition 1.2. *The following statements hold in a vector space* V.

1. *The element* 0 *in axiom (4) is uniquely determined.*

2. *(Cancellation property) If* $v, y, z \in V$ *and* $y + v = z + v$, *then* $y = z$.

3. *The element* w *in axiom (5) is uniquely determined.*

Proof. 1. Suppose 0_1 and 0_2 are elements in V that satisfy axiom (4). Then axioms (4) and (2) give $0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2$.

2. Assume $v + w = 0$. Then $y = y + 0 = y + (v + w) = (y + v) + w = (z + v) + w = z + (v + w) = z + 0 = z$.

3. If $v + w_1 = 0 = v + w_2$, then $w_1 = w_2$ by (2). □

The notation "$-v$" will stand for the unique element w in axiom (5). Thus, $v + (-v) = 0$. The notation "$v + (-w)$" is usually shortened to "$v - w$".

Example 1. This is the most basic and most important example of a vector space. Let $n \geq 1$, n an integer. Let

$$k^{(n)} = \{(a_1, \dots, a_n) | a_i \in k\},$$

the set of all n-tuples of elements of k. Two elements $(a_1, \dots, a_n)$ and $(b_1, \dots b_n)$ are equal if and only if $a_1 = b_1, \dots, a_n = b_n$. Define addition and scalar multiplication by

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$$

and

$$c(a_1, \dots, a_n) = (ca_1, \dots, ca_n).$$

Then $k^{(n)}$ is a vector space under these two operations. The zero element of $k^{(n)}$ is $(0, \dots, 0)$ and $-(a_1, \dots, a_n) = (-a_1, \dots, -a_n)$. Check that axioms (1)-(10) are valid and conclude that $k^{(n)}$ is a vector space over k (Exercise 1).

We will show in Proposition 2.8 that every finitely generated vector space over k is isomorphic to $k^{(n)}$ for some n. This is why Example 1 is so important.

Proposition 1.3. *The following computations are valid in a vector space V over k.*

1. $0v = 0$ *for all* $v \in V$. *(Note that the first "0" lies in k and the second "0" lies in V.)*
2. $a0 = 0$ *for all* $a \in k$.
3. $a(-v) = -(av)$ *for all* $a \in k, v \in V$.
4. $-v = (-1)v$, *for all* $v \in V$.
5. *If* $a \in k, a \neq 0, v \in V$, *and* $av = 0$, *then* $v = 0$.
6. *If* $a_1v_1 + \cdots + a_nv_n = 0, a_1 \neq 0$, *then* $v_1 = (-a_2/a_1)v_2 + \cdots + (-a_n/a_1)v_n$.

Proof. 1. $0v + 0 = 0v = (0 + 0)v = 0v + 0v$, so $0 = 0v$ by the cancellation property (Proposition 1.2(2)).

2. $a0 + 0 = a0 = a(0 + 0) = a0 + a0$, so $0 = a0$ by the cancellation property.

3. $a(-v) + av = a(-v + v) = a0 = 0$, by (2). Since $-(av) + av = 0$ by axiom (5), we have $a(-v) = -(av)$ by Proposition 1.2(3).

4. $(-1)v + v = (-1)v + 1v = (-1 + 1)v = 0v = 0$, and so $(-1)v = -v$ by Proposition 1.2(3).

5. $v = 1v = ((1/a) \cdot a)v = (1/a)(av) = (1/a)(0) = 0$, by (2).

6. $a_1v_1 = -(a_2v_2 + \cdots a_nv_n) = (-1)(a_2v_2 + \cdots a_nv_n) = -a_2v_2 - \cdots - a_nv_n$, which gives
$$\begin{aligned} v_1 = 1 \cdot v_1 = ((1/a_1) \cdot a_1)v_1 = (1/a_1)(a_1v_1) = (1/a_1)(-a_2v_2 - \cdots - a_nv_n) \\ = (-a_2/a_1)v_2 + \cdots + (-a_n/a_1)v_n. \end{aligned}$$ □

Definition 1.4. *A subset W of a vector space V is a* subspace *of V if $W \subseteq V$ and W is a vector space over k with respect to the operations of V. W is a* proper *subspace of V if W is a subspace of V and $W \subsetneq V$.*

Example 2. $\{0\}$ and V are subspaces of V.

We will use the following convention. The subspace of V consisting of only the vector 0 will be written (0). A set consisting of only the zero vector will be written $\{0\}$. Although the difference is small, it is useful to distinguish the two notions.

Proposition 1.5. *Let W be a nonempty subset of a vector space V over k. Then W is a subspace of V if and only if the following two statements hold.*

1. *If $v, w \in W$, then $v + w \in W$.*

2. *If $a \in k, v \in W$, then $av \in W$.*

Proof. If W is a subspace of V, then (1) and (2) hold because of axioms (1) and (6). Now suppose that (1) and (2) hold. Let $v \in W$ (W is nonempty). Then $0 = 0v \in W$ and $-v = (-1)v \in W$ by (2). Now it is easy to check that axioms (1)-(10) hold in W. (Most of the axioms are obvious since they are already valid in V.) □

Definition 1.6. *Let S be a nonempty subset of V. A* linear combination *of elements of S is an expression $\sum_{v \in S} a_v v$, $a_v \in k$, where only finitely many of the a_v's are nonzero. (It is often convenient to think of a linear combination as a finite sum $\sum_{i=1}^{n} a_i v_i$ where $v_1, \ldots, v_n$ are distinct elements of S). A* nontrivial linear combination *of elements of S is a linear combination as above in which at least one of the coefficients a_v is nonzero.*

Let W be the set of all linear combinations of elements of S. Then W is a subspace of V by Proposition 1.5, since if $\sum_{v\in S} a_v v$, $\sum_{v\in S} b_v v \in W$, and $c \in k$, then

$$\sum_{v\in S} a_v v + \sum_{v\in S} b_v v = \sum_{v\in S}(a_v + b_v)v \in W$$

and

$$c\sum_{v\in S} a_v v = \sum_{v\in S}(ca_v)v \in W.$$

The set W is called the subspace generated by S, or spanned by S, and we write $W = \langle S\rangle$.

Definition 1.7.

1. *A vector space V is* finitely generated *if $V = \langle S\rangle$ for some finite subset S of V.*

2. *A subset S of V is* linearly independent *over k if every equation $\sum_{i=1}^{n} a_i v_i = 0$, where $a_i \in k$ and $\{v_1, \ldots, v_n\}$ are distinct elements of S, implies that $a_i = 0, 1 \le i \le n$. In this situation, one also says that the elements of S are linearly independent over k.*

3. *A subset S of V is* linearly dependent *over k if S is not linearly independent over k. That is, there exist distinct elements $v_1, \ldots, v_n$ in S, elements $a_1, \ldots, a_n \in k$, and an equation $\sum_{i=1}^{n} a_i v_i = 0$ where some $a_i \neq 0$.*

4. *A subset S of V is a* basis *of V if S is a linearly independent subset of V and $V = \langle S\rangle$.*

The following convention is used. If $S = \emptyset$, the empty set, then $\langle S\rangle = (0)$. The empty set is a linearly independent set.

Example 3. Let $k^{(n)}$ be the vector space defined in Example 1. Let $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $e_n = (0, \ldots, 0, 1)$. Thus e_i is the n-tuple with zero in each coordinate except for a one in the i^{th} coordinate. Then $\{e_1, \ldots, e_n\}$ is a basis of $k^{(n)}$ (Exercise 2). This basis is called the *standard basis* of $k^{(n)}$.

Here are two more examples of vector spaces over a field k that are useful to remember.

Example 4. Let $k^{(\infty)}$ denote the set of all sequences $(a_1, a_2, a_3, \ldots)$ with each $a_i \in k$. Then $k^{(\infty)}$ is a vector space where addition is performed coordinatewise and $c(a_1, a_2, a_3, \ldots) = (ca_1, ca_2, ca_3, \ldots)$.

Example 5. Let $\overline{k^{(\infty)}}$ denote the set of all sequences $(a_1, a_2, a_3, \ldots)$ with each $a_i \in k$ such that only finitely many of the a_i's are nonzero. Then $\overline{k^{(\infty)}}$ is a subspace of $k^{(\infty)}$.

Here is a basis of $\overline{k^{(\infty)}}$ that is easy to describe. Let $e_i = (0, \ldots, 0, 1, 0, \ldots)$, where the i^{th} coordinate is 1 and every other coordinate is zero. Then $\epsilon_\infty = \{e_1, e_2, e_3, \ldots\}$ is a basis of $\overline{k^{(\infty)}}$ called the standard basis of $\overline{k^{(\infty)}}$. (The vector space $k^{(\infty)}$ also has a basis but it is not so easy to describe.)

Proposition 1.8. *The following statements are equivalent for a nonempty set S.*

1. *S is a basis of V.*

2. *Every element of V can be written uniquely as a linear combination $\sum_{v \in S} a_v v$, $a_v \in k$.*

Proof. (1) $\Rightarrow$ (2). Since $V = \langle S \rangle$, each element $v \in V$ can be written as a linear combination as in (2) and the uniqueness follows from Exercise 10.

(2) $\Rightarrow$ (1). The hypothesis in (2) implies that $V = \langle S \rangle$. Suppose that $\sum_{v \in S} a_v v = 0$. Since $0 = \sum_{v \in S} 0v$, the uniqueness part of (2) implies $a_v = 0$ for all $v \in S$. This shows S is a linearly independent subset and so S is a basis of V. □

Theorem 1.9. *Let V be a vector space over k and let $V = \langle T \rangle$ for some nonempty finite subset T of V. Let S be a subset of T that is linearly independent. Then there exists a basis B of V such that $S \subseteq B \subseteq T$.*

Proof. Since T is finite, there is a maximal subset B of T containing S that is linearly independent over k. Let $W = \langle B \rangle$. If $W \neq V$, then there exists some $x \in T$ such that $x \notin W$ (since if $T \subseteq W$, then $V = \langle T \rangle \subseteq W$). Then $B \subsetneq B \cup \{x\}$ and so the maximality of B implies that $B \cup \{x\}$ is a linearly dependent set. It follows that $x \in \langle B \rangle = W$, a contradiction. (See Exercise 13). Therefore, $W = V$ and B is a basis of V. □

Theorem 1.9 also holds if T is an infinite set. The proof is similar but requires the use of Zorn's Lemma.

Corollary 1.10. *Let V be a finitely generated vector space.*

1. *V has a basis.*
2. *Every finite linearly independent subset of V can be extended to a basis of V.*
3. *Every finite generating set of V contains a basis of V.*

Proof. Suppose that $V = \langle T \rangle$ for some finite set T. Let S equal the empty set and apply Theorem 1.9 to conclude that there exists a basis B of V such that $S \subseteq B \subseteq T$. This proves (1) and (3).

To prove (2), let S be a finite linearly independent subset of V. Then $S \cup T$ is a finite set, $S \subseteq S \cup T$, and $V = \langle S \cup T \rangle$. Theorem 1.9. implies that V has a basis B such that $S \subseteq B \subseteq S \cup T$. Thus S has been extended to a basis of V. This proves (2). □

Corollary 1.10 also holds for vector spaces that are not finitely generated. The proof uses the general case of Theorem 1.9.

Lemma 1.11 (Replacement Lemma). *Suppose $\{v_1, \dots, v_m\}$ is a basis of V. Let $w \in V$, $c_1, \dots, c_m \in k$, and suppose that $w = c_1v_1 + \cdots + c_mv_m$ with $c_1 \neq 0$. Then $\{w, v_2, \dots, v_m\}$ is a basis of V.*

Proof. Since $v_1 = c_1^{-1}(w - c_2v_2 - \cdots - c_mv_m) \in \langle w, v_2, \dots, v_m \rangle$, it follows that $\langle w, v_2, \dots, v_m \rangle = V$. (See exercise 11.)

Now suppose that $a_1w + a_2v_2 + \cdots + a_mv_m = 0$, with each $a_i \in k$. Then $a_1c_1v_1 + \sum_{i=2}^{m}(a_1c_i + a_i)v_i = 0$. Since $\{v_1, \dots, v_m\}$ is a linearly independent set, we have that $a_1c_1 = 0$. Thus $a_1 = 0$ because $c_1 \neq 0$. This gives $a_2v_2 + \cdots + a_mv_m = 0$. We conclude from the linear independence of $\{v_2, \dots, v_m\}$ that $a_i = 0$ for $2 \leq i \leq m$. Therefore $\{w, v_2, \dots, v_m\}$ is a linearly independent set, and so $\{w, v_2, \dots, v_m\}$ is a basis of V. □

Theorem 1.12. *Any two bases of a vector space V have the same cardinality.*

Proof. We shall assume that V is finitely generated. The case when V is not finitely generated will not be needed.

Since V is finitely generated, V contains a finite basis $\{v_1, \dots, v_m\}$ by Theorem 1.9. Thus, $V = \langle v_1, \dots, v_m \rangle$. Let $w_1, \dots, w_n$ be any elements of V that are linearly independent over k. We will show that $n \leq m$. Assuming this has been done, we may conclude that any other basis of V has at most

m elements. If a second basis of V has l elements, $l \leq m$, then the symmetry of this argument lets us conclude that $m \leq l$ also. Thus, any basis of V has exactly m elements.

It remains to show that $n \leq m$. Suppose that $n > m$. Since $V = \langle v_1, \ldots, v_m \rangle$, we have $w_1 = c_1v_1 + \cdots + c_mv_m$, $c_i \in k$. As $w_1 \neq 0$ (see Exercise 9), we may relabel the v_i's to assume that $c_1 \neq 0$. Then $\{w_1, v_2, \ldots, v_m\}$ is a basis of V by Lemma 1.11. Suppose by induction, and after relabeling, that $\{w_1, \ldots, w_r, v_{r+1}, \ldots, v_m\}$ is a basis of V, $1 \leq r < m$. Then

$$w_{r+1} = c_1w_1 + \cdots + c_rw_r + c_{r+1}v_{r+1} + \cdots + c_mv_m, \ c_i \in k.$$

Now $c_{r+1}, \ldots, c_m$ cannot all equal zero since $w_1, \ldots, w_{r+1}$ are linearly independent over k. Thus, we may relabel to assume that $c_{r+1} \neq 0$. Lemma 1.11 implies that $\{w_1, \ldots, w_{r+1}, v_{r+2}, \ldots, v_m\}$ is a basis of V. Since $n > m$, we may continue this process and conclude that $\{w_1, \ldots, w_m\}$ is a basis of V. Then $\{w_1, \ldots, w_{m+1}\}$ would be a linearly dependent set, since w_{m+1} would be contained in $V = \langle w_1, \ldots, w_m \rangle$. This is impossible since $\{w_1, \ldots, w_n\}$ is a linearly independent set. Therefore $n \leq m$ and the proof is complete. □

Definition 1.13. *The* dimension *of a finitely generated vector space V over k is the number m of elements in any, and hence all, bases of V. This is written* $\dim_k V = m$.

If $V = (0)$, then $\dim_k V = 0$ since the cardinality of the empty set equals zero. If there is no confusion, we will usually write $\dim V$ instead of $\dim_k V$. It follows from Exercise 2 that $\dim(k^{(n)}) = n$.

Proposition 1.14. *Let $S = \{v_1, \ldots, v_n\}$ and assume that* $\dim V = n$. *The following statements are equivalent.*

1. *S is a basis of V.*

2. *S is a linearly independent set.*

3. *$V = \langle S \rangle$. That is, V is spanned by S.*

Proof. (1) ⇒ (2) is obvious.

(2) ⇒ (3). Extend S to a basis T of V using Corollary 1.10. Since $|T| = n$ by Theorem 1.12, it follows $S = T$ and $V = \langle S \rangle$.

(3) ⇒ (1). Use Corollary 1.10 to choose $R \subseteq S$ such that R is a basis of V. Again, $|R| = n$ by Theorem 1.12 so $R = S$ and it follows that S is a basis of V. □

The intersection of two subspaces of V is another subspace of V. (See Exercise 4.) The sum of two subspaces of a vector space V is defined in Exercise 6 and is shown there to be a subspace of V . The next result gives information on the dimensions of these subspaces in terms of the original two subspaces.

Proposition 1.15. *Let U, W be subspaces of a finitely generated vector space V. Then*

$$\dim U + \dim W = \dim(U + W) + \dim(U \cap W).$$

Proof. Let $R = \{y_1, \dots, y_l\}$ be a basis of $U \cap W$. Use Corollary 1.10 to choose bases $S = \{y_1, \dots, y_l, u_{l+1}, \dots, u_m\}$ of U and $T = \{y_1, \dots, y_l, w_{l+1}, \dots, w_n\}$ of W. Our goal is to show that

$$S \cup T = \{y_1, \dots, y_l, u_{l+1}, \dots, u_m, w_{l+1}, \dots, w_n\}$$

is a basis of $U + W$. Assuming that this has been shown, then

$$\dim(U+W)+\dim(U\cap W) = [l+(m-l)+(n-l)]+l = m+n = \dim U+\dim W.$$

Let $v \in U + W$. Then $v = v_1 + v_2$ where $v_1 \in U$, and $v_2 \in W$. Since $v_1 \in \langle S\rangle$ and $v_2 \in \langle T\rangle$, it follows that $v \in \langle S \cup T\rangle$ and so $U + W = \langle S \cup T\rangle$.

Suppose that $\sum_{i=1}^{l} a_i y_i+\sum_{i=l+1}^{m} b_i u_i+\sum_{i=l+1}^{n} c_i w_i = 0$ where $a_i, b_i, c_i \in k$. Then

$$\sum_{i=l+1}^{n} c_i w_i = -\sum_{i=1}^{l} a_i y_i - \sum_{i=l+1}^{m} b_i u_i \in U \cap W.$$

Thus, $\sum_{i=l+1}^{n} c_i w_i = \sum_{i=1}^{l} d_i y_i$, since R is a basis of $U \cap W$. Since T is a basis of W, it follows that $c_i = 0$, $l + 1 \leq i \leq n$, and $d_i = 0, 1 \leq i \leq l$. Now we have $\sum_{i=1}^{l} a_i y_i + \sum_{i=l+1}^{m} b_i u_i = 0$. Since S is a basis of U, it follows that $a_i = 0$, $1 \leq i \leq l$, and $b_i = 0$, $l + 1 \leq i \leq m$. This shows $S \cup T$ is a linearly independent set and so $S \cup T$ is a basis of $U + W$. □

The next result will be used in the definition of a direct sum of vector spaces.

Proposition 1.16. *Let $W_1, \dots, W_m$ be subspaces of V. The following statements are equivalent.*

1. $V = W_1 + \cdots + W_m$ *and*

$$(W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_m) \cap W_i = (0),\ 1 \leq i \leq m.$$

2. *Each $v \in V$ has a unique expression $v = w_1 + \cdots + w_m$ where $w_i \in W_i$, $1 \leq i \leq m$.*

Proof. (1) $\Rightarrow$ (2). Statement (1) implies that each $v \in V$ has an expression as in (2). Suppose that $w_1 + \cdots + w_m = y_1 + \cdots + y_m$, where $w_i, y_i \in W_i$. Then

$$(w_1 - y_1) + \cdots + (w_{i-1} - y_{i-1}) + (w_{i+1} - y_{i+1}) + \cdots + (w_m - y_m) = \\ y_i - w_i \in (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_m) \cap W_i = (0), 1 \leq i \leq m.$$

Therefore, $y_i - w_i = 0$, so $w_i = y_i, 1 \leq i \leq m$.

(2) $\Rightarrow$ (1). Let $v \in (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_m) \cap W_i$. Then we have $v = w_1 + \cdots + w_{i-1} + w_{i+1} + \cdots + w_m = w_i$, where $w_j \in W_j, 1 \leq j \leq m$. The uniqueness part of (2) implies that each $w_j = 0$, so it follows that $v = 0$. The remaining part of (1) is obvious from (2). □

Definition 1.17. *If $W_1, \ldots, W_m$ are subspaces of a vector space V that satisfy the statements of Proposition 1.16, then we say that V is the (internal)* direct sum *of $W_1, \ldots, W_m$ and write $V = W_1 \bigoplus \cdots \bigoplus W_m$.*

There is a slightly different notion of direct sum that is given in Exercise 7. These are distinguished by referring to the direct sum in Definition 1.17 as the internal direct sum and the direct sum in Exercise 7 as the external direct sum. The distinction is not large since we will show in Chapter 2 that the two direct sums are essentially the same.

Notes to Section 1.1

We have proved that all finitely generated vector spaces V have a basis, and that all such bases have the same cardinality. There remains the question of finding or describing all bases of V. In addition, we need computational methods to find bases of vector spaces and/or calculate the dimensions of vector spaces. The following is a specific problem that is important to solve. Let $v_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \in k^{(n)}$, $1 \leq i \leq m$. Let $W = \langle v_1, \ldots, v_m \rangle$, the subspace of V spanned by $\{v_1, \ldots, v_m\}$. What is $\dim W$ and how does one find a basis of W? Some methods will be described in Chapter 3.

There are other approaches to this material. See, for example, the text by Hoffman and Kunze.

An exposition of Zorn's Lemma can be found in Hungerford's Algebra book.

Exercises

1. Check that $k^{(n)}$, as defined in Example 1, satisfies the ten axioms of a vector space.

2. Check that $\{e_1, \ldots, e_n\}$ defined in Example 3 is a basis of $k^{(n)}$.

3. Check the details of Examples 4 and 5.

4. Let $W_1, \ldots, W_n$ be subspaces of a vector space V. Then $W_1 \cap \cdots \cap W_n$ is also a subspace of V.

5. Suppose S is a subset of a vector space V and $W = \langle S \rangle$. Then W is the smallest subspace of V that contains S, and W is the intersection of all subspaces of V that contain S.

6. Let $W_1, \ldots, W_n$ be subspaces of a vector space V. Define $W_1 + \cdots + W_n$ to be $\{w_1 + \cdots + w_n | w_i \in W_i, 1 \leq i \leq n\}$. Then $W_1 + \cdots + W_n$ is also a subspace of V. Verify that it is the smallest subspace of V that contains $W_1, \ldots, W_n$. That is, $W_1 + \cdots + W_n = \langle W_1 \cup \cdots \cup W_n \rangle$.

7. Let $V_1, \ldots, V_n$ be vector spaces over k. Define $V_1 \bigoplus \cdots \bigoplus V_n$ to be $\{(v_1, \ldots, v_n) | v_i \in V_i, 1 \leq i \leq n\}$. Define addition and scalar multiplication on this set according to the following rules.

$$(v_1, \ldots, v_n) + (y_1, \ldots, y_n) = ((v_1 + y_1), \ldots, (v_n + y_n))$$

$$c(v_1, \ldots, v_n) = (cv_1, \ldots, cv_n)$$

   Then $V_1 \bigoplus \cdots \bigoplus V_n$ is a vector space over k. This vector space is called the (external) direct sum of $V_1, \ldots, V_n$.

8. (a) Let W_1, W_2 be two subspaces of a vector space V. Suppose $V = W_1 \cup W_2$. Then either $V = W_1$ or $V = W_2$. That is, V is not the union of two proper subspaces.

   (b) (harder) If k is an infinite field, show that V is not a finite union of proper subspaces. What can be said if k is a finite field?

9. Every element of a linearly independent subset is nonzero. The set containing only the zero vector is a linearly dependent set.

10. If S is a linearly independent subset of V and $\sum_{v\in S} a_v v = \sum_{v\in S} b_v v$, $a_v, b_v \in k$, then $a_v = b_v$ for all $v \in S$.

11. Let $\{v_1, \dots, v_n\}$ be a generating set of V and let $w_1, \dots, w_m \in V$. If $v_1 \dots, v_n \in \langle w_1, \dots, w_m\rangle$, then $V = \langle w_1, \dots w_m\rangle$.

12. Let S be a linearly independent subset of a vector space V and let $x \in V$. Then $S \cup \{x\}$ is a linearly independent subset if and only if $x \notin \langle S\rangle$.

13. Let V be a finitely generated vector space and let W be a subspace of V. Then

    (a) W is finitely generated.

    (b) $\dim W \le \dim V$.

    (c) If $\dim W = \dim V$, then $W = V$.

14. A subset S of V containing only nonzero vectors is linearly dependent if and only if some vector v in S can be expressed as a nontrivial linear combination of vectors in S distinct from v. This statement is false if the phrase "some vector v" is replaced by "each vector v". The statement is also false if S is allowed to contain the zero vector.

15. Let $S \subseteq T$ be subsets of a vector space V. If T is a linearly independent set, then so is S. If S is a linearly dependent set, then so is T.

16. Assume $\dim V = n$. Any set $S \subseteq V$ that contains more than n elements must be a linearly dependent set. Any set $S \subseteq V$ that contains fewer than n elements does not generate V.

17. Let $W_1, \dots, W_n$ be subspaces of V. Suppose that $V = W_1 \bigoplus \cdots \bigoplus W_n$ (as in Definition 1.17) and that V is finitely generated. Then $\dim V = \sum_{i=1}^n \dim W_i$. In fact if S_i is a basis for W_i, then $\bigcup S_i$ is a basis of V.

18. Let $V_1, \dots, V_n$ be finitely generated vector spaces over k. Then

$$\dim(V_1 \bigoplus \cdots \bigoplus V_n) = \sum_{i=1}^n \dim(V_i).$$

    Describe a basis of $V_1 \bigoplus \cdots \bigoplus V_n$ in terms of bases of $V_1, \dots, V_n$. (Note that this is the external direct sum as defined in Exercise 7.)

19. Let W be a subspace of V, with $\dim V$ finite. Then there exists a subspace Y of V such that $V = W \bigoplus Y$.

20. Show that $\{(2,3),(3,4)\}$ is a basis of $\mathbb{R}^{(2)}$.

21. Let $a,b,c,d \in \mathbb{R}$. Show that $\{(a,b),(c,d)\}$ is a basis of $\mathbb{R}^{(2)}$ if and only if $ad - bc \neq 0$.

1.2 Systems of Linear Equations

We introduce in this section the main subspaces associated with a system of linear equations. Studying systems of linear equations gives a lot of motivation for most of the concepts introduced in later chapters. Many of the topics introduced here will be introduced in more detail later in this text.

Consider the following system of m linear equations in n variables with coefficients in a field k.

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m$$

We have $a_{ij}, c_i \in k$ for all i and j.

A system of linear equations is called a system of homogeneous linear equations if $c_1 = \cdots = c_m = 0$.

The matrix of coefficients of a system of linear equations is the $m \times n$ matrix A given by

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

We let $\mathcal{M}_{m\times n}(k)$ denote the set of all $m \times n$ matrices with entries in k.

The rows of A are the m vectors

$$(a_{11}, a_{12}, \ldots, a_{1n})$$

$$(a_{21}, a_{22}, \ldots, a_{2n})$$

$$\vdots$$

$$(a_{m1}, a_{m2}, \ldots, a_{mn})$$

in $k^{(n)}$, $1 \le i \le m$.

The columns of A are the n vectors

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \in k^{(m)}, \quad 1 \le j \le n.$$

We will write such vectors vertically or horizontally, whichever is more convenient.

The null set of A is the set of vectors in $k^{(n)}$ that are solutions to the homogeneous system of linear equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0.$$

It is straightforward to verify that the null set of A is a subspace of $k^{(n)}$. (See Exercise 1.)

Definition 1.18. *Let $A \in \mathcal{M}_{m\times n}(k)$.*

1. *The* row space of A*, denoted $\mathcal{R}(A)$, is the subspace of $k^{(n)}$ spanned by the m rows of A.*

2. *The* column space of A*, denoted $\mathcal{C}(A)$, is the subspace of $k^{(m)}$ spanned by the n columns of A.*

3. *The* null space of A*, denoted $\mathcal{N}(A)$, is the subspace of vectors in $k^{(n)}$ that are solutions to the above homogeneous system of linear equations.*

The product of a matrix $A \in \mathcal{M}_{m \times n}(k)$ and a vector $b \in k^{(n)}$ is a particular vector $c \in k^{(m)}$ defined below. We follow the traditional notation and write b and c as column vectors. Thus we are defining $Ab = c$ and writing the following expression.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}.$$

The vector c is defined by setting

$$c_i = \sum_{j=1}^{n} a_{ij} b_j, \text{ for } 1 \leq i \leq m.$$

We will see in Chapter 3 a conceptual reason for this seemingly arbitrary definition.

It is convenient to introduce additional notation to deal with sums of this type. Given two vectors $v, w \in k^{(n)}$, we define the dot product of v and w as follows. Let $v = (a_1, a_2, \ldots, a_n)$ and $w = (b_1, b_2, \ldots, b_n)$. The dot product of v and w is defined by

$$v \cdot w = \sum_{j=1}^{n} a_j b_j.$$

The definition of Ab above can be expressed in terms of dot products of vectors. Let $v_1, \ldots, v_m$ denote the m row vectors of A in $k^{(n)}$. Then $Ab = c$ where

$$c = \begin{pmatrix} v_1 \cdot b \\ v_2 \cdot b \\ \vdots \\ v_m \cdot b \end{pmatrix}.$$

There is a second way to express $Ab = c$. Let $w_1, \ldots w_n$ denote the n column vectors of A in $k^{(m)}$. We continue to write

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ and } c = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}.$$

Then it is straightforward to verify that $Ab = c$ where

$$c = b_1w_1 + b_2w_2 + \cdots b_nw_n.$$

(See Exercise 9.)

The dot product satisfies the following properties. Let $v, w, y \in k^{(n)}$ be arbitrary vectors and let $a, b \in k$.

1. $v \cdot w = w \cdot v$
2. $v \cdot (w + y) = v \cdot w + v \cdot y$
3. $v \cdot aw = a(v \cdot w)$

Thus $v \cdot (aw + by) = a(v \cdot w) + b(v \cdot y)$. (See Exercise 2.)

Motivated by results on dot products of vectors in vector spaces over the real numbers, we say that two vectors $v, w \in k^{(n)}$ are orthogonal if $v \cdot w = 0$. We write $v \perp w$ to denote that $v \cdot w = 0$.

Let W be an arbitrary subspace of $k^{(n)}$. We define the orthogonal complement of W, denoted $W^\perp$, by

$$W^\perp = \{v \in k^{(n)} \mid v \cdot w = 0 \text{ for all } w \in W\}.$$

Using the properties of dot products, it is straightforward to verify that $W^\perp$ is a subspace of $k^{(n)}$. (See Exercise 3.)

Putting together all of our definitions and notation, it follows that

$$\mathcal{N}(A) = (\mathcal{R}(A))^\perp.$$

A system of linear equations can be expressed by the single matrix equation

$$Ax = c, \text{ where } x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

For a matrix $A \in \mathcal{M}_{m\times n}(k)$, define the function

$$L_A : k^{(n)} \to k^{(m)} \text{ by } L_A(b) = Ab.$$

The function L_A is an example of a linear transformation. Linear transformations will be studied in detail in Chapter 2. In particular, for all $v, v_1, v_2 \in k^{(n)}$ and $a \in k$, the following properties hold.

1. $L_A(v_1 + v_2) = L_A(v_1) + L_A(v_2)$
2. $L_A(av) = aL_A(v)$

We see that $\mathrm{im}(L_A) = \{Ab \mid b \in k^{(n)}\} = \mathcal{C}(A)$. Note that the system of linear equations $Ax = c$ has a solution if and only if $c \in \mathcal{C}(A)$.

Suppose we know that b is a solution of the matrix equation $Ax = c$, and so $L_A(b) = c$. Then all solutions of the equation $Ax = c$ are given by

$$b + \mathcal{N}(A) = \{b + y | y \in \mathcal{N}(A)\}.$$

(See Exercise 4.) For the special case that $c = 0$, we may certainly take $b = 0$, and then this result simply reduces to the original definition of $\mathcal{N}(A)$.

Thus the existence of solutions of a system of linear equations and the characterization of all such solutions can be described in terms of the function L_A and the subspaces $\mathcal{N}(A)$ and $\mathcal{C}(A) = \mathrm{im}(L_A)$.

Theorem 1.19. *Using the notations from above, we have*

$$\dim(\mathcal{N}(A)) + \dim(\mathcal{C}(A)) = n.$$

Proof. Let $\{w_1, \ldots, w_s\}$ be a basis of $\mathcal{N}(A)$. As $\mathcal{N}(A) \subseteq k^{(n)}$, we can extend $\{w_1, \ldots, w_s\}$ to a basis $\{w_1, \ldots, w_s, v_1, \ldots v_{n-s}\}$ of $k^{(n)}$. We have $Aw_i = 0$, $1 \le i \le s$.

We will now show that $\{Av_1, \ldots Av_{n-s}\}$ is a basis of $\mathcal{C}(A)$. This will let us conclude that $\dim(\mathcal{C}(A)) = n - s$, which will finish the proof. Since $\{w_1, \ldots, w_s, v_1, \ldots v_{n-s}\}$ is a basis of $k^{(n)}$, we have that $\mathcal{C}(A) = \mathrm{im}(L_A)$ is spanned by $\{Aw_1, \ldots, Aw_s, Av_1, \ldots Av_{n-s}\}$. Since $Aw_i = 0$ for $1 \le i \le s$, it follows that $\mathcal{C}(A)$ is spanned by $\{Av_1, \ldots Av_{n-s}\}$.

To show that $\{Av_1, \ldots Av_{n-s}\}$ is a linearly independent set, suppose that $c_1Av_1 + \cdots + c_{n-s}Av_{n-s} = 0$, where each $c_i \in k$. Then

$$A(c_1v_1 + \cdots + c_{n-s}v_{n-s}) = 0.$$

Thus $c_1v_1 + \cdots + c_{n-s}v_{n-s} \in \mathcal{N}(A) = \mathrm{Span}\{w_1, \ldots, w_s\}$. It follows that $c_1 = \cdots = c_{n-s} = 0$ because $\{v_1, \ldots, v_{n-s}, w_1, \ldots, w_s\}$ is a basis of $k^{(n)}$. Thus $\{Av_1, \ldots Av_{n-s}\}$ is a linearly independent set, and so is also a basis of $\mathcal{C}(A)$. □

The equation $\dim(\mathcal{N}(A)) + \dim(\mathcal{C}(A)) = n$ allows us to recover many of the standard results about systems of linear equations. We need a method to compute a basis of $\mathcal{N}(A)$ and a basis of $\mathcal{C}(A)$ in order to completely describe the set of solutions to a system of linear equations. This is one of the main applications of matrix methods developed in Chapter 3.

There are three important results about $\dim(\mathcal{C}(A))$, $\dim(\mathcal{R}(A))$, and $\dim(\mathcal{N}(A))$. Theorem (1.19) is the first result. We also have the following two theorems.

Theorem 1.20. *For any subspace $V \subseteq k^{(n)}$, we have*

$$\dim(V) + \dim(V^{\perp}) = n.$$

In particular, since $\mathcal{N}(A) = \mathcal{R}(A)^{\perp}$, we have

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n.$$

Theorem 1.21. $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$.

For any subspace $V \subseteq k^{(n)}$, there exists a matrix $A \in \mathcal{M}_{m\times n}(k)$ with $m \geq \dim(V)$ such that $V = \mathcal{R}(A)$. (See Exercise 5.) From this it follows that any two of Theorems (1.19), (1.20), and (1.21) easily implies the third. (See Exercise 6.) There are several proofs of Theorems (1.20) and (1.21) in these notes that use different methods. Each of the proofs is trickier than our previous proofs.

Exercises

1. For $A \in \mathcal{M}_{m\times n}(k)$, prove that the null set of A is a subspace of $k^{(n)}$.
2. Verify the properties of the dot product of two vectors as given in the text.
3. For a subspace $W \subseteq k^{(n)}$, verify that $W^{\perp}$ is a subspace of $k^{(n)}$.
4. If b is a solution of the system $Ax = c$, prove that all solutions are given by $b + \mathcal{N}(A) = \{b + y | y \in \mathcal{N}(A)\}$.
5. For any subspace $V \subseteq k^{(n)}$, prove that for $m \geq \dim(V)$ there exists a matrix $A \in \mathcal{M}_{m\times n}(k)$ such that $V = \mathcal{R}(A)$.

6. Using the previous exercise, prove that any two of Theorems 1.19, 1.20, 1.21 implies the third.

7. For any $v \in k^{(n)}$, let $\langle v \rangle$ denote the subspace of $k^{(n)}$ generated by v.

   (a) For any nonzero $a \in k$, prove that $\langle av \rangle^{\perp} = \langle v \rangle^{\perp}$.

   (b) Prove that $\dim(\langle v \rangle^{\perp}) = \begin{cases} n-1 & \text{if } v \neq 0 \\ n & \text{if } v = 0. \end{cases}$

   Note that this exercise gives a proof of the special case of Theorem 1.20 when $\dim(V) = 0$ or 1.

8. Consider a system of homogeneous linear equations where $m < n$. Use Theorem 1.19 to prove that there exists a nonzero solution to the system.

9. Let $e_1, \ldots, e_n$ be the standard basis of $k^{(n)}$.

   (a) Show that Ae_j is the j^{th} column of A.

   (b) Suppose that $b = b_1e_1 + \cdots + b_ne_n$ where each $b_j \in k$. Show that $Ab = \sum_{j=1}^{n} b_j(\text{column } j \text{ of } A)$.

10. Let $V, W \subseteq k^{(n)}$ be subspaces. Prove the following statements.

    (a) If $W \subseteq V$, then $V^{\perp} \subseteq W^{\perp}$.

    (b) $(V + W)^{\perp} = V^{\perp} \cap W^{\perp}$

    (c) $(V \cap W)^{\perp} = V^{\perp} + W^{\perp}$

    In (c), the inclusion $(V \cap W)^{\perp} \subseteq V^{\perp} + W^{\perp}$ is difficult to prove at this point. One strategy is to prove that $(V \cap W)^{\perp} \supseteq V^{\perp} + W^{\perp}$ and to use Theorem 1.20 (which has not yet been proved) to prove that $\dim((V \cap W)^{\perp}) = \dim(V^{\perp} + W^{\perp})$. Later we will be able to give a complete proof of this inclusion.

1.3 Appendix

Let V be a vector space defined over a field k. In this appendix, we show that the axiom stating that addition is commutative is actually redundant.

That is, the other axioms for a vector space already imply that addition is commutative.

(The following argument is actually valid for an arbitrary R-module M defined over a commutative ring R with a multiplicative identity.)

We assume that V is a group, not necessarily commutative, and we assume the following three facts concerning scalar multiplication.

1. $r \cdot (\vec{v} + \vec{w}) = r \cdot \vec{v} + r \cdot \vec{w}$ for all $\vec{v}, \vec{w} \in V$ and for all $r \in k$.

2. $(r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$ for all $v \in V$ and for all $r, s \in k$.

3. $1 \cdot \vec{v} = \vec{v}$ for all $v \in V$.

On the one hand, using (1) and then (2) we have

$$(r + s) \cdot (\vec{v} + \vec{w}) = (r + s) \cdot \vec{v} + (r + s) \cdot \vec{w} = r \cdot \vec{v} + s \cdot \vec{v} + r \cdot \vec{w} + s \cdot \vec{w}.$$

On the other hand, using (2) and then (1) we have

$$(r + s) \cdot (\vec{v} + \vec{w}) = r \cdot (\vec{v} + \vec{w}) + s \cdot (\vec{v} + \vec{w}) = r \cdot \vec{v} + r \cdot \vec{w} + s \cdot \vec{v} + s \cdot \vec{w}.$$

Since V is a group, we can cancel on both sides to obtain

$$s \cdot \vec{v} + r \cdot \vec{w} = r \cdot \vec{w} + s \cdot \vec{v}.$$

We let $r = s = 1$ and use (3) to conclude that $\vec{v} + \vec{w} = \vec{w} + \vec{v}$. Therefore, addition in V is commutative.

1.4 Proof of Theorems 1.20 and 1.21

We let $\mathcal{M}_{m \times n}(k)$ denote the set of $m \times n$ matrices with entries in a field k, and we let $\{e_1, \ldots, e_n\}$ denote the standard basis of $k^{(n)}$.

Lemma 1.22. *Let $A \in \mathcal{M}_{n \times n}(k)$. Then A is an invertible matrix if and only if $\{Ae_1, \ldots, Ae_n\}$ is a basis of $k^{(n)}$.*

Proof. Assume that A is an invertible matrix. We show that $\{Ae_1, \ldots, Ae_n\}$ is a linearly independent spanning set of $k^{(n)}$. Suppose that $c_1Ae_1 + \cdots + c_nAe_n = 0$, where each $c_i \in k$. Then $A(c_1e_1 + \cdots + c_ne_n) = 0$. Since A is invertible, it follows that $c_1e_1 + \cdots + c_ne_n = 0$, and so $c_1 = \cdots = c_n = 0$ because $\{e_1, \ldots, e_n\}$ is a linearly independent set. Thus $\{Ae_1, \ldots, Ae_n\}$

is a linearly independent set in $k^{(n)}$. To show that $\{Ae_1, \dots, Ae_n\}$ spans $k^{(n)}$, consider $v \in k^{(n)}$. Since $\{e_1, \dots, e_n\}$ is a basis of $k^{(n)}$, we may write $A^{-1}v = c_1e_1 + \cdots + c_ne_n$ where each $c_i \in k$. Applying A to both sides gives $v = A(c_1e_1 + \cdots + c_ne_n) = c_1Ae_1 + \cdots + c_nAe_n$. Therefore, $\{Ae_1, \dots, Ae_n\}$ is a basis of $k^{(n)}$.

Now assume that $\{Ae_1, \dots, Ae_n\}$ is a basis of $k^{(n)}$. There exist $b_{ij} \in k$, $1 \leq i \leq n$ and $1 \leq j \leq n$, such that $\sum_{i=1}^n b_{ij}(Ae_i) = e_j$. Let $B = (b_{ij})_{n\times n}$, an $n \times n$ matrix. Using the fact that

$$\sum_{i=1}^{n} b_{ij}(\text{column } i \text{ of } A) = \sum_{i=1}^{n} b_{ij}(Ae_i) = e_j,$$

we see that $AB = I_n$, where I_n denotes the $n \times n$ identity matrix. Therefore, A is in an invertible matrix. □

Proposition 1.23. *Let $\{v_1, \dots, v_n\}$ be a basis of $k^{(n)}$. Then there is a unique basis $\{w_1, \dots, w_n\}$ of $k^{(n)}$ such that $v_i \cdot w_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$*

Proof. Let $\{e_1, \dots, e_n\}$ be the standard basis of $k^{(n)}$. Let $A \in \mathcal{M}_{n\times n}(k)$ be the (unique) matrix such that $Ae_i = v_i$, $1 \leq i \leq n$. (The columns of A are the vectors $v_1, \dots, v_n$.) Then A is invertible by Lemma 1.22. Let $w_i = (A^t)^{-1}e_i$, $1 \leq i \leq n$. Then $\{w_1, \dots, w_n\}$ is a basis of $k^{(n)}$ by Lemma 1.22 because $(A^t)^{-1}$ is invertible. We have

$$\begin{aligned} v_i \cdot w_j &= v_i^t w_j = (Ae_i)^t (A^t)^{-1} e_j = e_i^t A^t (A^t)^{-1} e_j \\ &= e_i^t e_j = e_i \cdot e_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \end{aligned}$$

Suppose that $\{w_1', \dots, w_n'\}$ is another basis of $k^{(n)}$ such that $v_i \cdot w_j' = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$ Then $v_i \cdot (w_j - w_j') = v_i \cdot w_j - v_i \cdot w_j' = 0$ for all i and j. Then $w_j - w_j' \in (k^{(n)})^\perp$ for all j. Since $(k^{(n)})^\perp = (0)$ (see Exercise 1), it follows that $w_j = w_j'$ for all j. This proves the uniqueness statement. □

Lemma 1.24. *Suppose that $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_n\}$ are bases of $k^{(n)}$ such that $v_i \cdot w_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$ Let $V = \operatorname{Span}\{v_1, \dots, v_l\}$, where $1 \leq l \leq n$. Then $V^\perp = \operatorname{Span}\{w_{l+1}, \dots, w_n\}$.*

Proof. We have $\mathrm{Span}\{w_{l+1}, \ldots, w_n\} \subseteq V^\perp$ because if $1 \le i \le l$ and $l+1 \le j \le n$, then $i \ne j$ and so $v_i \cdot w_j = 0$.

Take $w \in V^\perp$. We may write $w = a_1 w_1 + \cdots + a_n w_n$. Since $v_i \cdot w = 0$ for $1 \le i \le l$, it follows that $a_i = 0$ for $1 \le i \le l$. Then $w = a_{l+1} w_{l+1} + \cdots + a_n w_n$ and thus $w \in \mathrm{Span}\{w_{l+1}, \ldots, w_n\}$. Therefore $V^\perp = \mathrm{Span}\{w_{l+1}, \ldots, w_n\}$. □

The next result recovers Theorem 1.20.

Proposition 1.25. *Any subspace* $V \subseteq k^{(n)}$ *satisfies* $\dim(V) + \dim(V^\perp) = n$.

Proof. Extend a basis $\{v_1, \ldots, v_l\}$ of V to a basis $\{v_1, \ldots, v_n\}$ of $k^{(n)}$. Let $\{w_1, \ldots, w_n\}$ be a basis of $k^{(n)}$ such that $v_i \cdot w_j = \begin{cases} 0 & \text{if } i \ne j \\ 1 & \text{if } i = j. \end{cases}$ Then $\mathrm{Span}\{w_{l+1}, \ldots, w_n\}$ is a basis of $V^\perp$. Thus $\dim(V) + \dim(V^\perp) = l + (n-l) = n$. □

We can now prove Theorem 1.21.

Theorem 1.26. *Let* $A \in \mathcal{M}_{m\times n}(k)$. *Then* $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$.

Proof. Since $\mathcal{N}(A) = \mathcal{R}(A)^\perp$, Theorems 1.19 and 1.20 imply that

$$\begin{aligned}\dim(\mathcal{N}(A)) + \dim(\mathcal{C}(A)) = n &= \dim(\mathcal{R}(A)) + \dim(\mathcal{R}(A)^\perp) \\ &= \dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)).\end{aligned}$$

Therefore $\dim(\mathcal{C}(A)) = \dim(\mathcal{R}(A))$. □

The next proposition develops a bit further the ideas in the proof of Proposition 1.23 and Lemma 1.24.

Proposition 1.27. *Assume that* $A \in \mathcal{M}_{n\times n}(k)$ *is an invertible matrix. Let* W *be a subspace of* $k^{(n)}$. *Then* $(A^t)^{-1}W^\perp = (AW)^\perp$.

Proof. Take $v \in AW$ and $w \in (A^t)^{-1}W^\perp$. Then $v = Av_1$ where $v_1 \in W$ and $w = (A^t)^{-1}w_1$ where $w_1 \in W^\perp$. Then

$$v \cdot w = v^t w = (v_1^t A^t)(A^t)^{-1} w_1 = v_1^t w_1 = v_1 \cdot w_1 = 0,$$

because $v_1 \in W$ and $w_1 \in W^\perp$. Since v, w are arbitrary, it follows that $(A^t)^{-1}W^\perp \subseteq (AW)^\perp$.

Now take $w \in (AW)^{\perp}$. Let $w_1 = A^t w$. We will show that $w_1 \in W^{\perp}$. Let $v \in W$. Then

$$v \cdot w_1 = v^t w_1 = v^t A^t w = (Av)^t w = Av \cdot w = 0,$$

because $w \in (AW)^{\perp}$. Thus $w_1 \in W^{\perp}$, and so $w = (A^t)^{-1} w_1 \in (A^t)^{-1} W^{\perp}$. Therefore, $(AW)^{\perp} \subseteq (A^t)^{-1} W^{\perp}$, and so $(A^t)^{-1} W^{\perp} = (AW)^{\perp}$. □

Exercises

1. Prove that $(k^{(n)})^{\perp} = (0)$.

2. Assume that $A \in \mathcal{M}_{n \times n}(k)$ is an invertible matrix. Show for any basis $\{v_1, \ldots, v_n\}$ of $k^{(n)}$ that $\{Av_1, \ldots, Av_n\}$ is a basis of $k^{(n)}$.

3. Show the following.

   (a) Show that $V \subseteq (V^{\perp})^{\perp}$.

   (b) Compute $\dim((V^{\perp})^{\perp})$ and conclude that $V = (V^{\perp})^{\perp}$.

Chapter 2

Linear Transformations

2.1 Basic Results

The most important functions between vector spaces are called linear transformations. In this chapter we will study the main properties of these functions.

Let V, W be vector spaces over a field k. Let $0_V, 0_W$ denote the zero vectors of V, W respectively. We will write 0 instead, if the meaning is clear from context.

Definition 2.1. *A function $f : V \to W$ is a* linear transformation *if*

1. $f(v_1 + v_2) = f(v_1) + f(v_2)$*, for all* $v_1, v_2 \in V$ *and*
2. $f(av) = af(v)$*, for all* $a \in k$*,* $v \in V$*.*

For the rest of section 2.1, let $f : V \to W$ be a fixed linear transformation.

Lemma 2.2. $f(0_V) = 0_W$ *and* $f(-v) = -f(v)$*.*

Proof. Using Proposition 1.1, parts (1) and (4), we see that $f(0_V) = f(0 \cdot 0_V) = 0f(0_V) = 0_W$, and $f(-v) = f((-1)v) = (-1)f(v) = -f(v)$. □

For any vector space V and any $v, y \in V$, recall from Chapter 1 that $v - y$ means $v + (-y)$. For any linear transformation $f : V \to W$, and $v, y \in V$, Lemma 2.2 lets us write $f(v - y) = f(v) - f(y)$ because

$$f(v - y) = f(v + (-y)) = f(v) + f(-y) = f(v) + (-f(y)) = f(v) - f(y).$$

Examples Here are some examples of linear transformations $f : V \to W$.

1. If $f(v) = 0$ for all $v \in V$, then f is a linear transformation called the zero transformation.

2. If $V = W$ and $f(v) = v$ for all $v \in V$, then f is a linear transformation called the identity transformation. We denote the identity transformation on V by $1_V : V \to V$.

3. If U is a subspace of V, then the inclusion mapping $\iota : U \to V$ is a linear transformation, where for $u \in U$ we have $\iota(u) = u \in V$.

4. Let $V = W \bigoplus Y$. Each $v \in V$ can be written uniquely as $v = w + y$ where $w \in W$ and $y \in Y$. Let $f(v) = w$. Then f is a linear transformation from V to W called the projection from V to W.

5. Let $V = W = k[x]$ where $k[x]$ denotes the ring of polynomials with coefficients in k. Thus each element of V can be written $a_0+a_1x+a_2x^2+\cdots+a_mx^m$ where each $a_i \in k$ and $m \geq 0$. Then V is a vector space over k where addition in V is the usual addition of polynomials and scalar multiplication is the usual multiplication in $k[x]$ of an element in k by a polynomial in $k[x]$. Let $f : V \to V$ where $f(a_0 + a_1x + a_2x^2 + \cdots + a_mx^m) = a_1 + a_2x + a_3x^2 + \cdots a_mx^{m-1}$. Let $g : V \to V$ where $g(a_0 + a_1x + a_2x^2 + \cdots + a_mx^m) = a_0x + a_1x^2 + a_2x^3 + \cdots + a_mx^{m+1}$. Then f and g are both linear transformations.

6. Let $V = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is } C_\infty\}$. That is, V is the set of infinitely differentiable real-valued functions defined on the set of real numbers. Using standard facts about differentiable functions, one can show that V is a vector space over the field of real numbers. Let $d : V \to V$ be defined by $d(f) = f'$, where f' is the derivative of f. Then d is a linear transformation. Note that the hypothesis of infinite differentiability is needed since a function f might be differentiable at each real number but its derivative f' might not be differentiable at each real number. For example, let $f(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ -x^2 & \text{if } x < 0. \end{cases}$ Then $f'(x) = 2|x|$, but $f'(x)$ is not differentiable at $x = 0$.

Definition 2.3.

1. *The* kernel *of f, written* $\ker(f)$*, is defined as* $\{v \in V | f(v) = 0\}$.

2. *The* image *of f, written* $\mathrm{im}(f)$, *is defined as* $\{w \in W | w = f(v)$ *for some* $v \in V\}$.

3. f *is* injective *if* $f(v_1) = f(v_2)$ *implies* $v_1 = v_2$, *where* $v_1, v_2 \in V$.

4. f *is* surjective *if* $\mathrm{im}(f) = W$.

5. f *is* bijective *if* f *is injective and surjective.*

Proposition 2.4.

1. $\ker(f)$ *is a subspace of* V.

2. $\mathrm{im}(f)$ *is a subspace of* W.

3. f *is injective if and only if* $\ker(f) = (0)$.

Proof. (1) If $v_1, v_2 \in \ker(f)$, and $a \in k$, then $f(v_1 + v_2) = f(v_1) + f(v_2) = 0 + 0 = 0$, and $f(av_1) = af(v_1) = a \cdot 0 = 0$. Since $\ker(f)$ is nonempty ($0 \in \ker(f)$), the result follows from Proposition 1.5.

(2) Let $w_1, w_2 \in \mathrm{im}(f)$ and $a \in k$. Then $w_1 = f(v_1)$ and $w_2 = f(v_2)$ for some $v_1, v_2 \in V$. Then $w_1 + w_2 = f(v_1) + f(v_2) = f(v_1 + v_2) \in \mathrm{im}(f)$, and $aw_1 = af(v_1) = f(av_1) \in \mathrm{im}(f)$. Since $\mathrm{im}(f)$ is nonempty, the result follows from Proposition 1.5.

(3) Suppose that f is injective and let $v \in \ker(f)$. Since $f(v) = 0 = f(0)$, it follows that $v = 0$ and thus, $\ker(f) = 0$. Conversely, suppose that $\ker(f) = 0$ and let $f(v_1) = f(v_2)$. Then $f(v_1 - v_2) = f(v_1) + f(-v_2) = f(v_1) - f(v_2) = 0$. Thus, $v_1 - v_2 \in \ker(f) = (0)$ and so $v_1 = v_2$. Therefore f is injective. □

For any function $f : V \to W$, if $w \in W$, recall that $f^{-1}(w)$ denotes the inverse image (or preimage) of w in V. That is, $f^{-1}(w)$ is the set of elements $x \in V$ such that $f(x) = w$.

Proposition 2.5. *Let* $f : V \to W$ *be a linear transformation and let* $f(v) = w$ *where* $v \in V$ *and* $w \in W$. *Then* $f^{-1}(w) = v + \ker(f)$.

Proof. Let $y \in \ker(f)$. Then $f(v + y) = f(v) + f(y) = f(v) = w$. Thus, $v + y \in f^{-1}(w)$, and so $v + \ker(f) \subseteq f^{-1}(w)$.

Now let $u \in f^{-1}(w)$. Then $f(u) = w = f(v)$. This gives $f(u - v) = f(u) - f(v) = 0$, so $u - v \in \ker(f)$. Then $u = v + (u - v) \in v + \ker(f)$, so $f^{-1}(w) \subseteq v + \ker(f)$. □

Proposition 2.6. *Let f be as above and assume that* $\dim V$ *is finite. Then*

$$\dim V = \dim(\ker(f)) + \dim(\operatorname{im}(f)).$$

Proof. Let $\{v_1, \dots, v_n\}$ be a basis of V. Then $\{f(v_1), \dots, f(v_n)\}$ spans $\operatorname{im}(f)$. (See Exercise 3(a).) It follows from Corollary 1.10(3) that $\operatorname{im}(f)$ has a finite basis. Since $\ker(f)$ is a subspace of V, Exercise 13 of Chapter 1 implies that $\ker(f)$ has a finite basis.

Let $\{w_1, \dots, w_m\}$ be a basis of $\operatorname{im}(f)$ and let $\{u_1 \dots, u_l\}$ be a basis of $\ker(f)$. Choose $y_i \in V$, $1 \leq i \leq m$, such that $f(y_i) = w_i$. We will show that $\{u_1, \dots, u_l, y_1, \dots, y_m\}$ is a basis of V. This will show that $\dim V = l + m = \dim(\ker(f)) + \dim(\operatorname{im}(f))$.

Let $v \in V$ and let $f(v) = \sum_{j=1}^m c_j w_j$ where each $c_j \in k$. Then $v - \sum_{j=1}^m c_j y_j \in \ker(f)$ because

$$\begin{aligned} f(v - \sum_{j=1}^m c_j y_j) &= f(v) - f(\sum_{j=1}^m c_j y_j) \\ &= f(v) - \sum_{j=1}^m c_j f(y_j) = f(v) - \sum_{j=1}^m c_j w_j = 0. \end{aligned}$$

Thus, $v - \sum_{j=1}^m c_j y_j = \sum_{i=1}^l a_i u_i$ where each $a_i \in k$, and so $v = \sum_{i=1}^l a_i u_i + \sum_{j=1}^m c_j y_j$. Therefore V is spanned by $\{u_1, \dots, u_l, y_1, \dots, y_m\}$.

To show linear independence, suppose that $\sum_{i=1}^l a_i u_i + \sum_{j=1}^m c_j y_j = 0$. Since $u_i \in \ker(f)$, we have

$$\begin{aligned} 0_W = f(0_V) = f(\sum_{i=1}^l a_i u_i + \sum_{j=1}^m c_j y_j) = \sum_{i=1}^l a_i f(u_i) + \sum_{j=1}^m c_j f(y_j) \\ = 0_V + \sum_{j=1}^m c_j f(y_j) = \sum_{j=1}^m c_j w_j. \end{aligned}$$

Since $\{w_1, \dots, w_m\}$ is a basis of $\operatorname{im}(f)$, we see that $c_j = 0$, $1 \leq j \leq m$. We now have that $\sum_{i=1}^l a_i u_i = 0$. Then $a_i = 0$, $1 \leq i \leq l$ because $\{u_1 \dots, u_l\}$ is a basis of $\ker(f)$. Therefore, $\{u_1, \dots, u_l, y_1, \dots, y_m\}$ is a linearly independent set and forms a basis of V. □

Definition 2.7. *A linear transformation $f : V \to W$ is an* isomorphism *if f is injective and surjective. We say that V is* isomorphic *to W, written $V \cong W$, if there exists an isomorphism $f : V \to W$.*

Proposition 2.8. *Let* $\dim V = n$, $n \geq 1$, *and let* $\beta = \{v_1, \ldots, v_n\}$ *be a basis of* V*. Then there exists an isomorphism* $\phi_\beta : V \to k^{(n)}$, *where* $k^{(n)}$ *is defined in Example 1 in Chapter 1.*

Proof. Define the function $\phi_\beta : V \to k^{(n)}$ by $\phi_\beta(v) = (a_1, \ldots, a_n)$, where $v = \sum_{i=1}^n a_i v_i$. Then ϕ_β is well defined by Proposition 1.8.

To show that ϕ_β is a linear transformation, let $v = \sum_{i=1}^n a_i v_i$ and $w = \sum_{i=1}^n b_i v_i$. Since $v + w = \sum_{i=1}^n (a_i + b_i) v_i$, we have

$$\begin{aligned}\phi_\beta(v + w) &= (a_1 + b_1, \ldots, a_n + b_n) = (a_1, \ldots, a_n) + (b_1, \ldots, b_n) \\ &= \phi_\beta(v) + \phi_\beta(w).\end{aligned}$$

If $c \in k$, then $cv = \sum_{i=1}^n ca_i v_i$ and hence

$$\phi_\beta(cv) = (ca_1, \ldots, ca_n) = c(a_1, \ldots, a_n) = c\phi_\beta(v).$$

If $v \in \ker(\phi_\beta)$, then $\phi_\beta(v) = (0, \ldots, 0)$. Then $v = \sum_{i=1}^n 0v_i = 0$ and therefore ϕ_β is injective by Proposition 2.4(3). If $(a_1, \ldots, a_n) \in k^{(n)}$ is given, then for $v = \sum_{i=1}^n a_i v_i$ we have $\phi_\beta(v) = (a_1, \ldots, a_n)$. Thus ϕ_β is surjective and therefore, ϕ_β is an isomorphism. □

The isomorphism ϕ_β constructed in Proposition 2.8 depends on the choice of basis β of V. This observation will be used in Chapter 3.

Proposition 2.9. *Let* $f : V \to W$ *be a linear transformation and assume that* $\dim V = \dim W$ *is finite. Then the following statements are equivalent.*

1. f *is an isomorphism.*

2. f *is injective.*

3. f *is surjective.*

Proof. We have $\dim V = \dim(\ker(f)) + \dim(\mathrm{im}(f))$ by Proposition 2.6. Thus, f is injective $\iff \ker(f) = (0) \iff \dim(\ker(f)) = 0 \iff \dim(\mathrm{im}(f)) = \dim V = \dim W \iff \mathrm{im}(f) = W \iff f$ is surjective. (See Exercise 13.) Thus (2),(3) are equivalent. If (1) is true, then both (2) and (3) are true by the definition of isomorphism. If either (2) or (3) is true then both are true, and then (1) is true. This completes the proof. □

Proposition 2.9 fails to hold if $\dim V$ is not finite. For example, consider Example 5 above. Using the notation of that example, it is easy to check that f is surjective, but f is not injective. Also, g is injective, but not surjective. (See Exercise 19.)

2.2 The vector space of linear transformations $f : V \to W$

Notation. Let $\mathcal{L}(V, W)$ be the set of linear transformations from V to W.

We will define addition and scalar multiplication in $\mathcal{L}(V, W)$ so as to make $\mathcal{L}(V, W)$ a vector space over k with respect to these operations.

Let $f, g \in \mathcal{L}(V, W)$. Define $f+g$ to be the function $f+g : V \to W$, where $(f+g)(v) = f(v) + g(v)$, $v \in V$. For $a \in k$, define af to be the function $af : V \to W$, where $(af)(v) = af(v)$.

Proposition 2.10. *$\mathcal{L}(V, W)$ is a vector space with respect to the operations defined above.*

Proof. Let $f, g \in \mathcal{L}(V, W)$, and let $v, w \in V$, $a, b \in k$. Then

$$\begin{aligned}(f+g)(v+w) &= f(v+w) + g(v+w) = f(v) + f(w) + g(v) + g(w)\\ &= f(v) + g(v) + f(w) + g(w) = (f+g)(v) + (f+g)(w),\end{aligned}$$

and

$$\begin{aligned}(f+g)(av) &= f(av) + g(av) = af(v) + ag(v) = a(f(v) + g(v))\\ &= a(f+g)(v).\end{aligned}$$

Therefore, $f + g \in \mathcal{L}(V, W)$.

As for af, we have

$$\begin{aligned}(af)(v+w) &= af(v+w) = a(f(v) + f(w)) = af(v) + af(w)\\ &= (af)(v) + (af)(w),\end{aligned}$$

and

$$(af)(bv) = af(bv) = abf(v) = baf(v) = b(af)(v).$$

Therefore, $af \in \mathcal{L}(V, W)$.

We have now verified axioms (1) and (6) in the definition of a vector space. The 0 element of $\mathcal{L}(V, W)$ is the zero transformation f where $f(v) = 0$ for all $v \in V$. If $f \in \mathcal{L}(V, W)$, then $-f$ is the element $(-1)f$. Now it is easy to verify that $\mathcal{L}(V, W)$ satisfies the ten axioms of a vector space over k. (See Exercise 8.) □

If $g : U \to V$ and $f : V \to W$, then fg (or $f \circ g$) denotes the composite function $fg : U \to W$.

Proposition 2.11. *Let U, V, W be vector spaces over k. Let $g \in \mathcal{L}(U, V)$ and let $f \in \mathcal{L}(V, W)$. Then the composite function $fg \in \mathcal{L}(U, W)$. (If $u \in U$, $(fg)(u)$ is defined to be $f(g(u))$.)*

Proof. Let $u_1, u_2, u \in U$ and let $a \in k$. Then

$$\begin{aligned}(fg)(u_1 + u_2) &= f(g(u_1 + u_2)) = f(g(u_1) + g(u_2)) \\ &= f(g(u_1)) + f(g(u_2)) = (fg)(u_1) + (fg)(u_2),\end{aligned}$$

and $(fg)(au) = f(g(au)) = f(ag(u)) = af(g(u)) = a(fg)(u)$. □

Proposition 2.12. *Let U, V, W, Y be vector spaces over k. Let $f, f_1, f_2 \in \mathcal{L}(V, W)$, let $g, g_1, g_2 \in \mathcal{L}(U, V)$, let $h \in \mathcal{L}(W, Y)$, and let $a \in k$. Then the following properties hold.*

1. $(f_1 + f_2)g = f_1g + f_2g$
2. $f(g_1 + g_2) = fg_1 + fg_2$
3. $a(fg) = (af)g = f(ag)$
4. $(hf)(g) = (h)(fg)$

Proof. See Exercise 11. □

Suppose now that $U = V = W = Y$. Then the composition of linear transformations in $\mathcal{L}(V, V)$ acts like a type of multiplication in $\mathcal{L}(V, V)$. The identity linear transformation 1_V has the property that $1_V f = f 1_V = f$ for all $f \in \mathcal{L}(V, V)$. The properties in Proposition 2.12 let us regard $\mathcal{L}(V, V)$ as an associative algebra. That is, $\mathcal{L}(V, V)$ is not only a vector space but also a ring whose multiplication is compatible with scalar multiplication.

Let $f \in \mathcal{L}(V, W)$ and assume that f is bijective. Then the inverse function $f^{-1} : W \to V$ is defined by $f^{-1}(w) = v \iff f(v) = w$. The function f^{-1} is also bijective.

Proposition 2.13. *If $f \in \mathcal{L}(V, W)$ and f is bijective, then $f^{-1} \in \mathcal{L}(W, V)$. Thus, if f is an isomorphism, then f^{-1} is an isomorphism.*

Proof. Let $w_1, w_2 \in W$, and let $f^{-1}(w_1) = v_1$ and $f^{-1}(w_2) = v_2$. Then $f(v_1) = w_1$ and $f(v_2) = w_2$. Since $f(v_1 + v_2) = f(v_1) + f(v_2) = w_1 + w_2$, it follows that $f^{-1}(w_1 + w_2) = v_1 + v_2 = f^{-1}(w_1) + f^{-1}(w_2)$.

If $a \in k$, then $f(av_1) = af(v_1) = aw_1$. Thus, $f^{-1}(aw_1) = av_1 = af^{-1}(w_1)$. □

2.3 Quotient Spaces

Let V be a vector space over a field k and let W be a subspace of V. There is an important vector space formed from V and W called the quotient space V mod W, written V/W.

We begin the construction by defining the following equivalence relation on V. Let $v, w \in V$. We say that $v \equiv w \bmod W$ if $v - w \in W$. Then the following three properties are easily verified.

1. $v \equiv v \bmod W$ for all $v \in V$. (reflexive)

2. If $v \equiv w \bmod W$, then $w \equiv v \bmod W$. (symmetric)

3. If $v \equiv w \bmod W$ and $w \equiv y \bmod W$, then $v \equiv y \bmod W$. (transitive)

To see that the transitive property holds, observe that if $v \equiv w \bmod W$ and $w \equiv y \bmod W$, then $v - w \in W$ and $w - y \in W$. Then $v - y = (v - w) + (w - y) \in W$, and so $v \equiv y \bmod W$.

These three properties show that $\equiv \bmod W$ is an equivalence relation.

Recall that $v + W$ denotes the set $\{v + w | w \in W\}$. Such a set is called a coset of W in V (or more precisely, a left coset of W in V).

Lemma 2.14. *Let $v \in V$. Then $\{y \in V | y \equiv v \bmod W\} = v + W$. That is, the equivalence classes of $\equiv \operatorname{mod} W$ are the cosets of W in V.*

Proof. $\subseteq$: If $y \equiv v \bmod W$, then $y - v = w \in W$ and so $y = v + w \in v + W$.

$\supseteq$: Let $y = v + w \in v + W$, where $w \in W$. Then $y \equiv v \bmod W$ because $y - v = w \in W$. □

Lemma 2.15. *Let $v, y \in V$. The following four statements are equivalent.*

1. $v \equiv y \bmod W$.

2. $v - y \in W$.

3. $v + W = y + W$.

4. $(v + W) \cap (y + W) \neq \emptyset$.

In particular, two cosets of W are either equal or disjoint.

Proof. The equivalence of statements 1 and 2 comes from the definition of $\equiv \bmod W$.

$2 \Rightarrow 3$: If $v - y \in W$, then $v = y + w$ where $w \in W$. Then $v + W = (y + w) + W \subseteq y + W$. Similarly, $y + W \subseteq v + W$ because $y = v - w$. Thus $v + W = y + W$.

$3 \Rightarrow 4$: This is obvious.

$4 \Rightarrow 2$: Let $z \in (v + W) \cap (y + W)$. Then $z = v + w_1$ and $z = y + w_2$ where $w_1, w_2 \in W$. Then $v - y = w_2 - w_1 \in W$.

The equivalence of statements 3 and 4 implies that two cosets of W are either equal or disjoint. □

The equivalence classes of $\equiv \bmod W$ partition V into disjoint sets. Thus, the cosets of W in V partition V into disjoint sets.

Let V/W denote the set of equivalence classes of $\equiv \bmod W$ (or the set of cosets of W in V). The set V/W can be given the structure of a vector space over k. We must define addition and scalar multiplication and then check that the ten axioms of a vector space hold.

Addition: $(v + W) + (y + W) = (v + y) + W$, for all $v, y \in V$.

Scalar multiplication: $a(v + W) = (av) + W$, for all $a \in k$, $v \in V$.

It is necessary to check that these definitions are well defined. That is, we must show that the two operations do not depend on the choice of v or y to represent the equivalence class. Thus, it must be shown that if $v_1 + W = v_2 + W$ and $y_1 + W = y_2 + W$, then $(v_1 + y_1) + W = (v_2 + y_2) + W$. Similarly, it must be shown that $(av_1) + W = (av_2) + W$. (See Exercise 13.)

Proposition 2.16. *The set V/W under the two operations defined above is a vector space over k.*

Proof. The zero element $0_{V/W}$ of V/W is $0 + W$ and $-(v + W) = (-v) + W$. Now it is straightforward to check the ten axioms of a vector space. Note that checking the axioms in V/W involves using the corresponding axioms in V. (See Exercise 13.) □

Proposition 2.17. *The function $\pi : V \to V/W$ defined by $\pi(v) = v + W$ is linear transformation with $\ker(\pi) = W$ and $\operatorname{im}(\pi) = V/W$.*

Proof. $\pi(v + y) = (v + y) + W = (v + W) + (y + W) = \pi(v) + \pi(y)$.
$\pi(av) = (av) + W = a(v + W) = a\pi(v)$, $a \in k$.
Thus π is a linear transformation. The image of π is clearly all of V/W since $v + W = \pi(v)$. For the kernel of π, $v \in \ker(\pi) \iff v + W = 0 + W \iff v \in W$. □

Corollary 2.18. *Assume that* $\dim V$ *is finite and let* W *be a subspace of* V*. Then* $\dim(V/W) = \dim V - \dim W$*.*

Proof. Using the notation from Proposition 2.17, we have $\dim(V/W) = \dim(\operatorname{im}(\pi)) = \dim V - \dim(\ker(\pi)) = \dim V - \dim W$, by Proposition 2.6. □

Theorem 2.19 (First Isomorphism Theorem)**.** *Let* $f : V \to Z$ *be a linear transformation of vector spaces over* k*. Then* $V/\ker(f) \cong \operatorname{im}(f)$*. There exists a unique isomorphism* $\overline{f} : V/\ker(f) \to \operatorname{im}(f)$ *such that* $\overline{f} \circ \pi = f$*, where* $\pi : V \to V/\ker(f)$*.*

Proof. Let $Y = \operatorname{im}(f)$ and let $W = \ker(f)$. Define $\overline{f} : V/W \to Y$ by $\overline{f}(v + W) = f(v)$. We have to check five properties of $\overline{f}$.

$\overline{f}$ is well defined: Suppose $v + W = y + W$. Then $v - y \in W = \ker(f)$ and so $f(v) = f(y)$ because $f(v) - f(y) = f(v) + f(-y) = f(v - y) = 0$. Now, $\overline{f}(v + W) = f(v) = f(y) = \overline{f}(y + W)$.

$\overline{f}$ is a linear transformation: $\overline{f}((v + W) + (y + W)) = \overline{f}((v + y) + W) = f(v + y) = f(v) + f(y) = \overline{f}(v + W) + \overline{f}(y + W)$.
$\overline{f}(a(v + W)) = \overline{f}(av + W) = f(av) = af(v) = a\overline{f}(v + W)$.

$\overline{f}$ is injective: Let $\overline{f}(v + W) = 0$. Then $f(v) = 0$ and so $v \in \ker(f) = W$. Therefore $v + W = 0 + W$ and therefore $\ker(\overline{f}) = (0 + W)$.

$\overline{f}$ is surjective: Let $y \in Y$. Then $y = f(v)$ for some $v \in V$ and $\overline{f}(v + W) = f(v) = y$.

This shows that $\overline{f} : V/\ker(f) \to \operatorname{im}(f)$ is an isomorphism.

$\overline{f} \circ \pi = f$: This is obvious.

Finally, $\overline{f}$ is unique since $\overline{f}(v + W)$ must equal $f(v)$. □

Theorem 2.20 (Second Isomorphism Theorem)**.** *Let* W, Y *be subspaces of* V*. Then* $(W + Y)/W \cong Y/(W \cap Y)$*.*

Proof. Let $f : Y \to W + Y$ be the linear transformation defined by $f(y) = y$ for all $y \in Y$, and let $g : W + Y \to (W + Y)/W$ be the linear transformation defined by $g(w + y) = (w + y) + W$ for all $w \in W$ and $y \in Y$. Then $g \circ f : Y \to (W + Y)/W$ is a linear transformation. We now compute $\ker(g \circ f)$ and $\operatorname{im}(g \circ f)$. We have

$$y \in \ker(g \circ f) \iff y \in Y \text{ and } y + W = 0 + W \iff y \in W \cap Y.$$

Thus $\ker(g \circ f) = W \cap Y$. Next we have that $g \circ f$ is surjective because if $(w+y)+W \in (W+Y)/W$, then $(g\circ f)(y) = g(y) = y+W = (w+y)+W$. Thus $\operatorname{im}(g \circ f) = (W + Y)/W$. Then the First Isomorphism Theorem (Theorem 2.19) implies that $W/\ker(g \circ f) \cong \operatorname{im}(g \circ f)$. Therefore, $W/(W \cap Y) \cong (W + Y)/W$. □

2.4 Applications

Direct Sums.

We will now show that the internal direct sum defined in Definition 1.17 and the external direct sum defined in Exercise 7 of Chapter 1 are essentially the same.

Let $V = V_1 \bigoplus \cdots \bigoplus V_n$ be an external direct sum of the vector spaces $V_1, \ldots, V_n$, and let

$$V_i^{'} = \{(0, \ldots, 0, v_i, 0, \ldots, 0) | v_i \in V_i\}.$$

Then $V_i^{'}$ is a subspace of V and it is easy to check that $V_i^{'}$ is isomorphic to V_i. It follows from Definition 1.17 and Proposition 1.16 that V equals the internal direct sum $V_1^{'} \bigoplus \cdots \bigoplus V_n^{'}$.

Now suppose that V is the internal direct sum of subspaces $W_1, \ldots, W_m$ as in Definition 1.17. Let $V^{'}$ equal the external direct sum $W_1 \bigoplus \cdots \bigoplus W_m$. Consider the function $f : V \to V^{'}$ defined as follows. If $v \in V$, Proposition 1.16 implies there is a unique expression $v = w_1 + \cdots + w_m$ where each $w_i \in W_i$. Set $f(v) = (w_1, \ldots, w_m)$. It is easy to check that f is a bijective linear transformation and so f is an isomorphism. We will no longer distinguish between the internal and external direct sum.

Exact Sequences.

Definition 2.21. *Let $U \xrightarrow{f} V \xrightarrow{g} W$ be a sequence of linear transformations. (This means $f \in \mathcal{L}(U, V)$ and $g \in \mathcal{L}(V, W)$.) We say $U \xrightarrow{f} V \xrightarrow{g} W$ is an* exact sequence *at V if* $\operatorname{im}(f) = \ker(g)$. *A sequence*

$$V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-2}} V_{n-1} \xrightarrow{f_{n-1}} V_n$$

is an exact sequence *if* $\operatorname{im}(f_{i-1}) = \ker(f_i)$, $2 \leq i \leq n-1$, *that is, if the sequence is exact at $V_2, \ldots, V_{n-1}$.*

When writing sequences of linear transformations, it is standard to write 0 for the vector space (0). Note that the linear transformations $0 \to V$ and $V \to 0$ are uniquely defined and hence we will never give a name to these functions. The following lemma gives some of the basic results.

Lemma 2.22.

1. *The sequence* $0 \to V \to 0$ *is exact if and only if* $V = (0)$.
2. *The sequence* $0 \to U \xrightarrow{f} V$ *is exact if and only if* f *is injective.*
3. *The sequence* $V \xrightarrow{g} W \to 0$ *is exact if and only if* g *is surjective.*
4. *The sequence* $0 \to V \xrightarrow{f} W \to 0$ *is exact if and only if* f *is an isomorphism.*
5. *The sequence* $0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$ *is exact if and only if* f *is injective,* g *is surjective, and* $\operatorname{im}(f) = \ker(g)$.

Proof. The proofs are immediate from the definitions upon noting that

$$\operatorname{im}(0 \to V) = 0, \text{ and } \ker(V \to 0) = V.$$

□

Proposition 2.23. *Assume that* $0 \to V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-1}} V_n \to 0$ *is an exact sequence of finite dimensional vector spaces,* $n \geq 1$. *Then*

$$\sum_{i=1}^{n} (-1)^i \dim V_i = 0.$$

Proof. We prove the result by induction on n. If $n = 1$, then $V_1 = (0)$ and so $\dim V_1 = 0$. If $n = 2$, then $V_1 \cong V_2$ by Lemma 2.22(4) and so $\dim V_1 = \dim V_2$ (see Exercise 7). Now assume that $n = 3$. In this case, $0 \to V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3 \to 0$ is an exact sequence. Then

$$\begin{aligned} \dim V_2 &= \dim(\ker(f_2)) + \dim(\operatorname{im}(f_2)) = \dim(\operatorname{im}(f_1)) + \dim V_3 \\ &= \dim V_1 + \dim V_3. \end{aligned}$$

(Note that $\ker(f_1) = (0)$ since f_1 is injective and so $\dim(\ker(f_1)) = 0$. Therefore $\dim V_1 = \dim(\operatorname{im}(f_1))$ by Proposition 2.6.) Thus, $-\dim V_1 + \dim V_2 - \dim V_3 = 0$ and the result holds for $n = 3$.

For the general case, suppose that $n \geq 4$ and that we have proved the result for smaller values of n. Then the given exact sequence is equivalent to the following two exact sequences because $\operatorname{im}(f_{n-2}) = \ker(f_{n-1})$.

$$0 \to V_1 \xrightarrow{f_1} \cdots \xrightarrow{f_{n-3}} V_{n-2} \xrightarrow{f_{n-2}} \operatorname{im}(f_{n-2}) \to 0,$$

$$0 \to \operatorname{im}(f_{n-2}) \xrightarrow{i} V_{n-1} \xrightarrow{f_{n-1}} V_n \to 0.$$

By induction, we have $\sum_{i=1}^{n-2}(-1)^i \dim V_i + (-1)^{n-1}\dim(\operatorname{im}(f_{n-2})) = 0$ and $(-1)^{n-2}\dim(\operatorname{im}(f_{n-2})) + (-1)^{n-1}\dim V_{n-1} + (-1)^n \dim V_n = 0$. Adding these two equations gives us the result. □

Systems of Linear Equations.

Our last application is studying systems of linear equations. This motivates much of the material in Chapter 3.

Consider the following system of m linear equations in n variables with coefficients in k.

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m$$

Now consider the following vectors in $k^{(m)}$. We will write such vectors vertically or horizontally, whichever is more convenient.

$$u_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, u_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ldots, u_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Define the function $f : k^{(n)} \to k^{(m)}$ by $f(b_1, \ldots, b_n) = b_1u_1 + \cdots + b_nu_n$. Then f is a linear transformation. Note that $f(b_1, \ldots, b_n) = (c_1, \ldots, c_m)$ if and only if $(b_1, \ldots, b_n)$ is a solution for $(x_1, \ldots, x_n)$ in the system of linear equations above. That is, $(c_1, \ldots, c_m) \in \operatorname{im}(f)$ if and only if the system of linear equations above has a solution. Let us abbreviate the notation by letting $b = (b_1, \ldots, b_n)$ and $c = (c_1, \ldots, c_m)$. Then $f(b) = c$ means that b is a solution of the system of linear equations.

Suppose we know that b is a solution of the system and so $f(b) = c$. Then all solutions are given by $b + \ker(f) = \{b + y | y \in \ker(f)\}$ by Proposition 2.5.

Thus the existence of solutions of a system of linear equations and the characterization of all such solutions can be described in terms of a linear transformation f and the subspaces $\ker(f)$ and $\operatorname{im}(f)$.

The equations $n = \dim(k^{(n)}) = \dim(\ker(f)) + \dim(\operatorname{im}(f))$ allow us to recover many of the usual results about systems of linear equations. We need a method to compute a basis of $\ker(f)$ and a basis of $\operatorname{im}(f)$ in order to completely describe the set of solutions to a system of linear equations. This is one of the main applications of matrix methods developed in Chapter 3.

Notes to Chapter 2. The proof of Proposition 2.10 contains a step where we need to know that $ab = ba$ in a field. This is an interesting point since much of linear algebra can be developed over skew fields. (A skew field is a ring where each nonzero element is invertible, but multiplication need not be commutative.) One defines left and right vector spaces over a skew field D and other related concepts. A good reference for this is the book by E. Artin. In this setting, Proposition 2.10 does not carry over completely.

A common problem that students have is knowing when it is necessary to show that a function is well defined. If the domain of a function consists of equivalence classes, and if a function is defined in terms of an element of the equivalence class, then it is necessary to show that the definition of the function does not depend on the chosen element of the equivalence class.

Exercises

1. Give a second proof of Lemma 2.2 using the equations $0 + 0 = 0$ and $-v + v = 0$.

2. Verify that Examples 1-5 following Lemma 2.2 are linear transformations.

3. Let $f : V \to W$ be a linear transformation. Let $v_1, \ldots, v_n \in V$ and let $w_i = f(v_i)$, $1 \leq i \leq n$.

   (a) If $\{v_1, \ldots, v_n\}$ spans V, then $\{w_1, \ldots, w_n\}$ spans $\operatorname{im}(f)$. That is, $\langle\{w_1, \ldots, w_n\}\rangle = \operatorname{im}(f)$.

   (b) If $\{v_1, \ldots, v_n\}$ is a linearly independent set in V and f is injective, then $\{w_1, \ldots, w_n\}$ is a linearly independent set in W.

(c) If $\{w_1, \ldots, w_n\}$ is a linearly independent set in W, then $\{v_1, \ldots, v_n\}$ is a linearly independent set in V.

(d) If $\{w_1, \ldots, w_n\}$ is a linearly independent set in W and $\{v_1, \ldots, v_n\}$ is a basis of V, then f is injective.

(e) Assume that $\{v_1, \ldots, v_n\}$ is a basis of V. Then f is an isomorphism if and only if $\{w_1, \ldots, w_n\}$ is a basis of W.

4. Let W be a subspace of V, $\dim V$ finite. Define the *codimension* of W in V, written $\mathrm{codim}_k W$ (or $\mathrm{codim} W$ if k is understood from context), by the equation $\mathrm{codim} W = \dim V - \dim W$. More generally, we can use Corollary 2.18 as motivation to define $\mathrm{codim} W$ by $\mathrm{codim} W = \dim(V/W)$.

   (a) If W_1, W_2 are subspaces of V with $\mathrm{codim} W_1 = \mathrm{codim} W_2 = 1$, then $\mathrm{codim}(W_1 \cap W_2) \leq 2$. If $W_1 \neq W_2$, then $\mathrm{codim}(W_1 \cap W_2) = 2$.

   (b) Let W_1, W_2 be subspaces of V. If each $\mathrm{codim} W_i$ is finite, then
   $$\mathrm{codim}(W_1 \cap W_2) \leq \mathrm{codim} W_1 + \mathrm{codim} W_2.$$
   We have $\mathrm{codim}(W_1 \cap W_2) = \mathrm{codim} W_1 + \mathrm{codim} W_2$ if and only if $W_1 + W_2 = V$.

   (c) Let $W_1, \ldots, W_n$ be subspaces of V. If each $\mathrm{codim} W_i$ is finite, then
   $$\mathrm{codim}(W_1 \cap \cdots \cap W_n) \leq \sum_{i=1}^{n} \mathrm{codim} W_i.$$

5. Isomorphism is an equivalence relation on vector spaces.

6. Let $f : V \to W$ be an isomorphism of vector spaces. Let Y be a subspace of V. Then $Y \cong f(Y)$.

7. Two finitely generated vector spaces, V, W, over k are isomorphic if and only if $\dim V = \dim W$. (One can use Proposition 2.8, and Exercise 5 for one implication, and Exercise 3 (e) for the other implication.)

8. Finish the proof of Proposition 2.10 that $\mathcal{L}(V, W)$ is a vector space over k.

9. Let U, V, W be arbitrary sets and let $f : U \to V$ and $g : V \to W$ be arbitrary functions.

   (a) If f and g are injective, then $g \circ f$ is injective.

   (b) If $g \circ f$ is injective, then f is injective, but g may not be injective. Give such an example.

   (c) If f and g are surjective, then $g \circ f$ is surjective.

   (d) If $g \circ f$ is surjective, then g is surjective, but f may not be surjective. Give such an example.

10. Let V, W be vector spaces, and let $f \in \mathcal{L}(V, W)$ and $g \in \mathcal{L}(W, V)$.

    (a) If $fg = 1_W$, then g is injective and f is surjective. Similarly, if $gf = 1_V$, then f is injective and g is surjective. (This part depends only on sets and functions and has nothing to do with Linear Algebra.)

    (b) Assume that $\dim V = \dim W$ is finite. Then $fg = 1_W$ if and only if $gf = 1_V$.

    (c) If $\dim V \neq \dim W$, give an example where $gf = 1_V$ but $fg \neq 1_W$.

11. Prove Proposition 2.12.

12. Let $V = V_1 \bigoplus \cdots \bigoplus V_n$ and $W = W_1 \bigoplus \cdots \bigoplus W_m$. Then

    (a) $\mathcal{L}(V, W) \cong \bigoplus_{i=1}^{n} \mathcal{L}(V_i, W)$

    (b) $\mathcal{L}(V, W) \cong \bigoplus_{j=1}^{m} \mathcal{L}(V, W_j)$

    (c) $\mathcal{L}(V, W) \cong \bigoplus_{i=1}^{n} \bigoplus_{j=1}^{m} \mathcal{L}(V_i, W_j)$

    (d) If $\dim V = \dim W = 1$, then $\mathcal{L}(V, W) \cong k$.

13. Show that the definitions of addition and scalar multiplication on V/W are well defined. Finish the proof of Proposition 2.16 that V/W is a vector space.

14. Let W be a subspace of V and assume $\dim V$ is finite. Let $\{v_1, \ldots, v_l\}$ be a basis of W. Extend this basis to a basis $\{v_1, \ldots, v_l, v_{l+1}, \ldots, v_n\}$ of V. Then $\{v_{l+1} + W, \ldots, v_n + W\}$ is a basis of V/W.

15. Use the previous exercise to give a different proof of Corollary 2.18. Now use The First Isomorphism Theorem, Exercise 7, and Corollary 2.18 to give a different proof of Proposition 2.6.

16. Show that Proposition 1.15 is a consequence of the Second Isomorphism Theorem (Theorem 2.20), Exercise 7, and Corollary 2.18.

17. (Strong form of the First Isomorphism Theorem) Let $f : V \to Z$ be a linear transformation of vector spaces over k. Suppose $W \subseteq \ker(f)$. Then show there exists a unique linear transformation $\overline{f} : V/W \to \operatorname{im}(f)$ such that $\overline{f} \circ \pi = f$, where $\pi : V \to V/W$. Show $\overline{f}$ is surjective and $\ker(\overline{f}) \cong \ker(f)/W$. Conclude that $(V/W)/(\ker(f)/W) \cong \operatorname{im}(f)$.

18. (Third Isomorphism Theorem) Let V be a vector space and let $W \subseteq Y \subseteq V$ be subspaces. Then

$$(V/W)/(Y/W) \cong V/Y.$$

(Hint: Consider the linear transformation $f : V \to V/Y$ and apply the strong form of the First Isomorphism Theorem from Exercise 17.)

19. Check the assertion following Proposition 2.9 that in Example 5, the linear transformation f is surjective, but not injective, while g is injective, but not surjective.

20. Let V be a vector space and let $W \subseteq Y \subseteq V$ be subspaces. Assume that $\dim(V/W)$ is finite. Then $\dim(V/W) = \dim(V/Y) + \dim(Y/W)$. (Hint: We have $(V/W)/(Y/W) \cong V/Y$ from Problem 18. Then apply Corollary 2.18.)

21. Suppose that $0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$ is an exact sequence. Prove that $V \cong U \bigoplus W$.

22. Recall $\overline{k^\infty}$ from Example 5 above Chapter 1, Proposition 1.8, and recall $k[x]$ from Example 5 above Definition 2.3. Prove that $k[x]$ and $\overline{k^\infty}$ are isomorphic vector spaces over k.

Chapter 3

Matrix Representations of a Linear Transformation

3.1 Basic Results

Let V, W be vector spaces over a field k and assume that $\dim V = n$ and $\dim W = m$. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V and let $\gamma = \{w_1, \ldots, w_m\}$ be a basis of W. Let $f \in \mathcal{L}(V, W)$. We will construct an $m \times n$ matrix that is associated with the linear transformation f and the two bases β and γ.

Proposition 3.1. *Assume the notations above.*

1. *f is completely determined by $\{f(v_1), \ldots, f(v_n)\}$.*

2. *If $y_1, \ldots, y_n \in W$ are chosen arbitrarily, then there exists a unique linear transformation $f \in \mathcal{L}(V, W)$ such that $f(v_i) = y_i$, $1 \leq i \leq n$.*

Proof. 1. Let $v \in V$. Then $v = \sum_{i=1}^n c_i v_i$. Each $c_i \in k$ is uniquely determined because β is a basis of V. We have $f(v) = \sum_{i=1}^n c_i f(v_i)$, because f is a linear transformation.

2. If f exists, then it is certainly unique by (1). As for the existence of f, let $v \in V$, $v = \sum_{i=1}^n c_i v_i$. Define $f : V \to W$ by $f(v) = \sum_{i=1}^n c_i y_i$. Note that f is well defined because β is a basis of V and so v determines $c_1, \ldots, c_n$ uniquely. Clearly $f(v_i) = y_i$. Now we show that $f \in \mathcal{L}(V, W)$.

Let $v' = \sum_{i=1}^{n} d_i v_i$ and let $a \in k$. Then

$$\begin{aligned} f(v+v') &= f\left(\sum_{i=1}^{n}(c_i+d_i)v_i\right) = \sum_{i=1}^{n}(c_i+d_i)y_i = \sum_{i=1}^{n} c_i y_i + \sum_{i=1}^{n} d_i y_i \\ &= f(v) + f(v'), \end{aligned}$$

and

$$f(av) = f\left(a\sum_{i=1}^{n} c_i v_i\right) = \sum_{i=1}^{n} a c_i y_i = a\sum_{i=1}^{n} c_i y_i = af(v).$$

□

Definition 3.2. *An* ordered basis *of a finite dimensional vector space V is a basis $\{v_1, \dots, v_n\}$ where the order of the basis vectors $v_1, \dots, v_n$ is fixed.*

Let $\beta = \{v_1, \dots, v_n\}$ be an ordered basis of V, let $\gamma = \{w_1, \dots, w_m\}$ be an ordered basis of W, and let $f \in \mathcal{L}(V, W)$. Assume that $f(v_j) = \sum_{i=1}^{m} a_{ij} w_i$, $1 \le j \le n$. Thus, the linear transformation f gives rise to an $m \times n$ matrix $(a_{ij})_{m\times n}$ whose (i,j)-entry is given by a_{ij}. Note that the j^{th} column of this matrix gives the expression of $f(v_j)$ in terms of the basis $\{w_1, \dots, w_m\}$. We shall denote this matrix by $[f]_\beta^\gamma$ and refer to it as the matrix of f with respect to the ordered bases β, γ. The notation shows the dependence of this matrix on the linear transformation f and on the ordered bases β and γ. The notation

$$[f]_\beta^\gamma = (a_{ij})_{m\times n}$$

means that $f(v_j) = \sum_{i=1}^{m} a_{ij} w_i$, $1 \le j \le n$.

Let $\mathcal{M}_{m\times n}(k)$ denote the set of $m \times n$ matrices with entries in the field k. Let $A = (a_{ij})$, $B = (b_{ij})$ be elements in $\mathcal{M}_{m\times n}(k)$ and let $c \in k$. We define addition in $\mathcal{M}_{m\times n}(k)$ by setting $A + B$ to be the matrix $C = (c_{ij})$, where $c_{ij} = a_{ij} + b_{ij}$. We define scalar multiplication in $\mathcal{M}_{m\times n}(k)$ by setting cA to be the matrix $C = (c_{ij})$, where $c_{ij} = ca_{ij}$.

Proposition 3.3. *Under the operations defined above, the set $\mathcal{M}_{m\times n}(k)$ is a vector space over k that is isomorphic to $k^{(mn)}$. The dimension of $\mathcal{M}_{m\times n}(k)$ is mn.*

Proof. See Exercise 1. □

Proposition 3.4. *Let β, γ be ordered bases of V, W, as above. The function $\phi : \mathcal{L}(V,W) \to \mathcal{M}_{m\times n}(k)$, defined by $\phi(f) = [f]_\beta^\gamma$ is an isomorphism of vector spaces. In particular, $\mathcal{L}(V,W) \cong \mathcal{M}_{m\times n}(k) \cong k^{(mn)}$ and $\dim \mathcal{L}(V,W) = mn$.*

Proof. If $\phi(f) = \phi(g)$, then $f(v_j) = g(v_j)$, $1 \le j \le n$. Thus $f = g$ by Proposition 3.1(1). Therefore ϕ is injective. Now let $(a_{ij}) \in \mathcal{M}_{m\times n}(k)$ be given. Let $y_j = \sum_{i=1}^m a_{ij}w_i$, $1 \le j \le n$. There exists $f \in \mathcal{L}(V,W)$ such that $f(v_j) = y_j$, $1 \le j \le n$, by Proposition 3.1(2). Then $\phi(f) = [f]_\beta^\gamma = (a_{ij})$. Therefore ϕ is surjective and it follows that ϕ is a bijection.

Let $f, g \in \mathcal{L}(V,W)$ and let $[f]_\beta^\gamma = (a_{ij})$, $[g]_\beta^\gamma = (b_{ij})$. Then $f(v_j) = \sum_{i=1}^m a_{ij}w_i$, $g(v_j) = \sum_{i=1}^m b_{ij}w_i$, and $(cf)(v_j) = cf(v_j) = \sum_{i=1}^m ca_{ij}w_i$. Since

$$(f+g)(v_j) = f(v_j) + g(v_j) = \sum_{i=1}^m (a_{ij} + b_{ij})w_i,$$

it follows that

$$\phi(f+g) = [f+g]_\beta^\gamma = (a_{ij} + b_{ij}) = (a_{ij}) + (b_{ij}) = [f]_\beta^\gamma + [g]_\beta^\gamma = \phi(f) + \phi(g).$$

Similarly,

$$\phi(cf) = [cf]_\beta^\gamma = (ca_{ij}) = c(a_{ij}) = c[f]_\beta^\gamma = c\phi(f).$$

Thus ϕ is a linear transformation and so ϕ is an isomorphism.

The rest follows easily from Proposition 3.3. □

Note that the isomorphism ϕ depends on the choice of basis for V and W, although the notation doesn't reflect this.

Another proof that $\dim \mathcal{L}(V,W) = mn$ can be deduced from the result in Exercise 12 of Chapter 2 and the special case that $\dim \mathcal{L}(V,W) = 1$ when $\dim V = \dim W = 1$.

We have already defined addition and scalar multiplication of matrices in $\mathcal{M}_{m\times n}(k)$. Next we define multiplication of matrices. We will see in Proposition 3.5 below that the product of two matrices corresponds to the matrix of a composition of two linear transformations.

Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be an $n \times p$ matrix with entries in k. The product AB is defined to be the $m \times p$ matrix $C = (c_{ij})$ whose (i,j)-entry is given by $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$. A way to think of c_{ij} is to note that c_{ij} is the usual "dot product" of the i^{th} row of A with the j^{th} column of B. The following notation summarizes this.

$$A_{m\times n}B_{n\times p} = C_{m\times p}, \text{ where } c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

An important observation about matrix multiplication occurs in the special case that $p = 1$. In that case B is a "column matrix". Suppose the entries of B are $(b_1, \ldots, b_n)$ Then a short calculation shows that $AB = \sum_{j=1}^{n} b_j(j^{th}$ column of $A)$.

Note that the definition of matrix multiplication puts restrictions on the sizes of the matrices. An easy way to remember this restriction is to remember $(m \times n)(n \times p) = (m \times p)$.

Proposition 3.5. *Let U, V, W be finite dimensional vector spaces over k. Let $\alpha = \{u_1, \ldots, u_p\}$, $\beta = \{v_1, \ldots, v_n\}$, $\gamma = \{w_1, \ldots, w_m\}$ be ordered bases of U, V, W, respectively. Let $g \in \mathcal{L}(U, V)$ and $f \in \mathcal{L}(V, W)$ (and so $fg \in \mathcal{L}(U, W)$ by Proposition 2.11). Then $[fg]_\alpha^\gamma = [f]_\beta^\gamma [g]_\alpha^\beta$.*

Proof. Let $[f]_\beta^\gamma = (a_{ij})_{m\times n}$, $[g]_\alpha^\beta = (b_{ij})_{n\times p}$, and $[fg]_\alpha^\gamma = (c_{ij})_{m\times p}$. Then

$$\begin{aligned} (fg)(u_j) &= f(g(u_j)) = f\left(\sum_{k=1}^{n} b_{kj} v_k\right) = \sum_{k=1}^{n} b_{kj} f(v_k) \\ &= \sum_{k=1}^{n} b_{kj}\left(\sum_{i=1}^{m} a_{ik} w_i\right) = \sum_{i=1}^{m}\left(\sum_{k=1}^{n} a_{ik} b_{kj}\right) w_i. \end{aligned}$$

This shows that $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$, which is precisely the (i,j)-entry of $[f]_\beta^\gamma [g]_\alpha^\beta$. □

Proposition 3.6. *Let $A \in \mathcal{M}_{m\times n}(k)$, $B \in \mathcal{M}_{n\times p}(k)$, and $C \in \mathcal{M}_{p\times q}(k)$. Then $(AB)C = A(BC)$. In other words, multiplication of matrices satisfies the associative law for multiplication, as long as the matrices have compatible sizes.*

Proof. See Exercise 2. □

Exercise 19 gives some properties of matrices that are analogous to Proposition 2.12.

Let I_n denote the $n \times n$ matrix (a_{ij}) where $a_{ij} = 0$ if $i \neq j$ and $a_{ij} = 1$ if $i = j$. For any $A \in \mathcal{M}_{m\times n}(k)$, it is easy to check that $I_m A = A I_n = A$. If $m = n$, then $I_n A = A I_n = A$ For this reason, I_n is called the $n \times n$ identity matrix.

Definition 3.7. *A matrix $A \in \mathcal{M}_{n\times n}(k)$ is* invertible *if there exists a matrix $B \in \mathcal{M}_{n\times n}(k)$ such that $AB = I_n$ and $BA = I_n$ The matrix B is called the inverse of A and we write $B = A^{-1}$.*

Proposition 3.8.

1. *The inverse of an invertible matrix in $\mathcal{M}_{n\times n}(k)$ is uniquely determined and thus the notation A^{-1} is well defined.*

2. *Let A, B be invertible matrices in $\mathcal{M}_{n\times n}(k)$. Then $(AB)^{-1} = B^{-1}A^{-1}$.*

3. *Let $A_1, \ldots, A_l$ be invertible matrices in $\mathcal{M}_{n\times n}(k)$. Then $(A_1 \cdots A_l)^{-1} = A_l^{-1} \cdots A_1^{-1}$. In particular, the product of invertible matrices is invertible.*

4. $(A^{-1})^{-1} = A$.

Proof. Suppose $AB = BA = I_n$ and $AC = CA = I_n$. Then $B = BI_n = B(AC) = (BA)C = I_nC = C$. This proves (1). The other statements are easy to check. □

Proposition 3.9. *Let β be an ordered basis of V and let γ be an ordered basis of W. Assume that $\dim V = \dim W = n$. Suppose that $f : V \to W$ is an isomorphism and let $f^{-1} : W \to V$ denote the inverse isomorphism. (See Proposition 2.13.) Then $([f]_\beta^\gamma)^{-1} = [f^{-1}]_\gamma^\beta$.*

Proof. $[f]_\beta^\gamma[f^{-1}]_\gamma^\beta = [ff^{-1}]_\gamma^\gamma = [1_W]_\gamma^\gamma = I_n$. Similarly, $[f^{-1}]_\gamma^\beta[f]_\beta^\gamma = [f^{-1}f]_\beta^\beta = [1_V]_\beta^\beta = I_n$. The result follows from this and the definition of the inverse of a matrix. □

If $U = V = W$ and $\alpha = \beta = \gamma$ in Proposition 3.5, then the result in Proposition 3.5 implies that the isomorphism in Proposition 3.4 is also a ring isomorphism. (See the remark preceding Proposition 2.13.) Next we investigate how different choices of bases of V, W in Proposition 3.4 affect the matrix $[f]_\beta^\gamma$.

Proposition 3.10. *Let β, β' be ordered bases of V and let γ, γ' be ordered bases of W. Let $f \in \mathcal{L}(V, W)$, and let $1_V : V \to V$ and $1_W : W \to W$ be the identity maps on V, W, respectively. Then $[f]_{\beta'}^{\gamma'} = [1_W]_\gamma^{\gamma'}[f]_\beta^\gamma[1_V]_{\beta'}^\beta$.*

Proof. Since $f = 1_W \circ f \circ 1_V$, the result follows immediately from Proposition 3.5. □

An important special case of Proposition 3.10 occurs when $V = W$, $\dim V = n$, $\beta = \gamma$, and $\beta' = \gamma'$. Then $[f]_{\beta'}^{\beta'} = [1_V]_{\beta}^{\beta'}[f]_{\beta}^{\beta}[1_V]_{\beta'}^{\beta}$. Let $A = [1_V]_{\beta'}^{\beta}$ and $B = [1_V]_{\beta}^{\beta'}$. Then $AB = [1_V]_{\beta'}^{\beta}[1_V]_{\beta}^{\beta'} = [1_V]_{\beta}^{\beta} = I_n$ and similarly $BA = I_n$. Thus $B = A^{-1}$.

We can summarize this with more compact notation. Let $M = [f]_{\beta}^{\beta}$ and let $M' = [f]_{\beta'}^{\beta'}$. Then $M' = A^{-1}MA$.

In Proposition 2.7 we proved that if V has dimension n, then V is isomorphic to $k^{(n)}$. We will now extend this idea to help with computations involving linear transformations.

Let V be a vector space over k with basis $\beta = \{v_1, \dots, v_n\}$ and let W be a vector space over k with basis $\gamma = \{w_1, \dots, w_m\}$. Let ϵ_n be the standard basis of $k^{(n)}$. That is, $\epsilon_n = \{e_1, \dots, e_n\}$ where e_i is the vector in $k^{(n)}$ with coordinates equal to zero everywhere except in the i^{th} position where the coordinate equals 1. Similarly, let ϵ_m be the standard basis of $k^{(m)}$. Let $\phi_\beta : V \to k^{(n)}$ be the isomorphism (as in Proposition 2.7) that takes $v = \sum_{i=1}^{n} b_i v_i$ to $(b_1, \dots b_n)$. We will write $[v]_\beta$ for $\phi_\beta(v)$. Thus $[v]_\beta$ denotes the coordinates of v with respect to the basis β. Similarly, let $\phi_\gamma : W \to k^{(m)}$ be the isomorphism that takes w to $[w]_\gamma$.

Let $A = (a_{ij})_{m\times n} \in \mathcal{M}_{m\times n}(k)$. We let $L_A : k^{(n)} \to k^{(m)}$ denote the function that takes a vector b (considered as an $n \times 1$ column matrix) to the vector $Ab \in k^{(m)}$, where Ab is given by usual matrix multiplication. It is straightforward to check that $L_A \in \mathcal{L}(k^{(n)}, k^{(m)})$.

Lemma 3.11. *Using the notation from above, we have $[L_A]_{\epsilon_n}^{\epsilon_m} = A$, $[\phi_\beta]_{\beta}^{\epsilon_n} = I_n$, and $[\phi_\gamma]_{\gamma}^{\epsilon_m} = I_m$.*

Proof. $L_A(e_j) = Ae_j = \sum_{i=1}^{m} a_{ij} e_i$, since Ae_j is the j^{th} column of A. The result for L_A now follows from the definition of $[L_A]_{\epsilon_n}^{\epsilon_m}$.

The other two results follow easily upon noting that $\phi_\beta(v_j) = e_j \in k^{(n)}$ and $\phi_\gamma(v_j) = e_j \in k^{(m)}$. □

Proposition 3.12. *Let $f \in \mathcal{L}(V, W)$ and let $A = [f]_{\beta}^{\gamma}$. Then the following diagram commutes. That is, $\phi_\gamma \circ f = L_A \circ \phi_\beta$. In particular, $[f(v)]_\gamma = [f]_{\beta}^{\gamma}[v]_\beta$.*

$$\begin{tikzcd}
V \arrow[r, "f"] \arrow[d, "\phi_\beta"'] & W \arrow[d, "\phi_\gamma"] \\
k^{(n)} \arrow[r, "L_A"] & k^{(m)}
\end{tikzcd}$$

Proof. The second statement follows from the first because

$$[f(v)]_\gamma = (\phi_\gamma \circ f)(v) = (L_A \circ \phi_\beta)(v) = A[v]_\beta = [f]_\beta^\gamma[v]_\beta.$$

For the first statement, we have

$$[\phi_\gamma \circ f]_\beta^{\epsilon_m} = [\phi_\gamma]_\gamma^{\epsilon_m}[f]_\beta^\gamma = I_mA = A = AI_n = [L_A]_{\epsilon_n}^{\epsilon_m}[\phi_\beta]_\beta^{\epsilon_n} = [L_A \circ \phi_\beta]_\beta^{\epsilon_m}.$$

Therefore, $\phi_\gamma \circ f = L_A \circ \phi_\beta$ by Proposition 3.4. □

Here is another proof of the second statement of Proposition 3.12 that is a more concrete calculation.

Let $v \in V$, $v = \sum_{j=1}^n b_jv_j$. Then $f(v_j) = \sum_{i=1}^m a_{ij}w_i$, where $A = (a_{ij}) = [f]_\beta^\gamma$. We have

$$\begin{aligned} f(v) = f(\sum_{j=1}^n b_jv_j) &= \sum_{j=1}^n b_jf(v_j) = \sum_{j=1}^n b_j(\sum_{i=1}^m a_{ij}w_i) \\ &= \sum_{i=1}^m(\sum_{j=1}^n a_{ij}b_j)w_i. \end{aligned}$$

Therefore,

$$[f(v)]_\gamma = \begin{pmatrix} \sum_{j=1}^n a_{1j}b_j \\ \vdots \\ \sum_{j=1}^n a_{mj}b_j \end{pmatrix} = A\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = A[v]_\beta = [f]_\beta^\gamma[v]_\beta.$$

Proposition 3.13. *Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of a finite dimensional vector space V. If $f \in \mathcal{L}(V, V)$ is an isomorphism, define $\tau(f)$ to be the ordered basis $f(\beta)$, where $f(\beta) = \{f(v_1), \ldots, f(v_n)\}$. Then τ is a bijection between the set of isomorphisms $f \in \mathcal{L}(V, V)$ and the set of ordered bases of V.*

Proof. See Exercise 4. □

3.2 The Transpose of a Matrix

Definition 3.14. *Let $A \in \mathcal{M}_{m\times n}(k)$. The* transpose of A*, written A^t, is the $n \times m$ matrix in $\mathcal{M}_{n\times m}(k)$ whose (i, j)-entry is the (j, i)-entry of A. That is, if $A = (a_{ij})_{m\times n}$ and if $A^t = (b_{ij})_{n\times m}$, then $b_{ij} = a_{ji}$.*

It follows that the i^{th} row of A is the i^{th} column of A^t, $1 \leq i \leq m$, and the j^{th} column of A is the j^{th} row of A^t, $1 \leq j \leq n$.

Proposition 3.15. *1. Let $A, B \in \mathcal{M}_{m\times n}(k)$ and let $a \in k$. Then $(A+B)^t = A^t + B^t$, $(aA)^t = aA^t$, and $(A^t)^t = A$.*

2. The map $g : \mathcal{M}_{m\times n}(k) \to \mathcal{M}_{n\times m}(k)$ given by $A \mapsto A^t$ is an isomorphism of vector spaces.

3. Let $A \in \mathcal{M}_{m\times n}(k)$, $B \in \mathcal{M}_{n\times p}(k)$. Then $(AB)^t = B^tA^t$.

Proof. For (1) and (2), see Exercise 7.

(3) Let $C = AB = (c_{ij})_{m\times p}$. Let $A^t = (a'_{ij})_{n\times m}$ and $B^t = (b'_{ij})_{p\times n}$. Then

$$\begin{aligned}(i,j)\text{-entry of } B^tA^t &= \sum_{l=1}^{n} b'_{il}a'_{lj} = \sum_{l=1}^{n} b_{li}a_{jl} = \sum_{l=1}^{n} a_{jl}b_{li} = c_{ji} \\ &= (j,i)\text{-entry of } AB \\ &= (i,j)\text{-entry of } (AB)^t \ .\end{aligned}$$

Therefore, $(AB)^t = B^tA^t$. □

Proposition 3.16. *Let $A \in \mathcal{M}_{n\times n}(k)$ and assume that A is invertible. Then A^t is invertible and $(A^t)^{-1} = (A^{-1})^t$.*

Proof. Let $B = A^{-1}$. Then $A^tB^t = (BA)^t = I_n^t = I_n$ and $B^tA^t = (AB)^t = I_n^t = I_n$. Therefore the inverse of A^t is $B^t = (A^{-1})^t$. □

A more conceptual description of the transpose of a matrix and the results in Propositions 3.15 and 3.16 will be given in Chapter 4.

3.3 The Row Space, Column Space, and Null Space of a Matrix

Definition 3.17. *Let $A \in \mathcal{M}_{m\times n}(k)$.*

1. *The* row space of A *is the subspace of $k^{(n)}$ spanned by the rows of A.*

2. *The* column space of A *is the subspace of $k^{(m)}$ spanned by the columns of A.*

3. *The* null space of A *is the set of vectors* $b \in k^{(n)}$ *such that* $Ab = 0$.

Let $A \in \mathcal{M}_{m \times n}(k)$ and let

$$\begin{aligned}\mathcal{R}(A) &= \text{row space of } A,\\ \mathcal{C}(A) &= \text{column space of } A,\\ \mathcal{N}(A) &= \text{null space of } A.\end{aligned}$$

Let $L_A : k^{(n)} \to k^{(m)}$ be the linear transformation that takes $b \in k^{(n)}$ to $Ab \in k^{(m)}$.

The definition of matrix multiplication shows that $\text{im}(L_A) = \mathcal{C}(A)$ and $\ker(L_A) = \mathcal{N}(A)$. Therefore $\mathcal{N}(A)$ is a subspace of $k^{(n)}$. Also, $\mathcal{R}(A) = \mathcal{C}(A^t)$. The subspace $\mathcal{N}(A^t)$ is called the *left null space* of A. This is because if $c \in k^{(m)}$ is considered as a $1 \times m$ matrix, then $cA = 0 \iff (cA)^t = 0 \iff A^t c^t = 0$. The dimension formula from Proposition 2.6 implies that $n = \dim(\ker(L_A)) + \dim(\text{im}(L_A)) = \dim(\mathcal{N}(A)) + \dim(\mathcal{C}(A))$.

Proposition 3.18. *Let* $A \in \mathcal{M}_{m \times n}(k)$*,* $D \in \mathcal{M}_{p \times m}(k)$*, and* $E \in \mathcal{M}_{n \times p}(k)$*. Then*

1. $\mathcal{R}(DA) \subseteq \mathcal{R}(A)$.

2. $\mathcal{C}(AE) \subseteq \mathcal{C}(A)$.

3. $\mathcal{N}(DA) \supseteq \mathcal{N}(A)$.

If $p = m$ *and* D *is invertible, then equality occurs in (1) and (3). If* $p = n$ *and* E *is invertible, then equality occurs in (2).*

Proof. Let $D = (d_{ij})$. The i^{th} row of DA is given by

$$d_{i1}(\text{first row of } A) + \cdots + d_{in}(n^{th} \text{ row of } A),$$

which is contained in the row space of A. This shows (1).

Let $E = (e_{ij})$. The j^{th} column of AE is given by

$$e_{1j}(\text{first column of } A) + \cdots + e_{nj}(n^{th} \text{ column of } A),$$

which is contained in the column space of A. This shows (2).

If $Ax = 0$, then $DAx = 0$ and thus (3) holds.

If $p = m$ and D is invertible, then apply (1) to $A = D^{-1}(DA)$. If $p = n$ and E is invertible, then apply (2) to $A = (AE)E^{-1}$. If D is invertible, then $DAx = 0$ implies $D^{-1}DAx = D^{-1}0 = 0$, so $Ax = 0$. □

Proposition 3.19. *Let $A \in \mathcal{M}_{m\times n}(k)$, $D \in \mathcal{M}_{m\times m}(k)$, and $E \in \mathcal{M}_{n\times n}(k)$. Assume that both D and E are invertible. Then*

1. $\mathcal{C}(DA) = L_D(\mathcal{C}(A)) \cong \mathcal{C}(A)$,
2. $\mathcal{C}(AE) = \mathcal{C}(A)$,
3. $\mathcal{N}(DA) = \mathcal{N}(A)$,
4. $\mathcal{N}(AE) = L_{E^{-1}}\mathcal{N}(A) \cong \mathcal{N}(A)$,
5. $\mathcal{R}(DA) = \mathcal{R}(A)$,
6. $\mathcal{R}(AE) = L_{E^t}\mathcal{R}(A) \cong \mathcal{R}(A)$.

Proof. Statements (2), (3), and (5) were proved in Proposition 3.18. We give alternative proofs below.

Since E is an invertible matrix, it follows that E^t and E^{-1} are also invertible matrices. Since D is also invertible, it follows that L_D, L_E, L_{E^t}, and $L_{E^{-1}}$ are isomorphisms. (See Exercise 5.) Since L_D is an isomorphism, we have

$$\mathcal{C}(DA) = \mathrm{im}(L_{DA}) = \mathrm{im}(L_D \circ L_A) = L_D(\mathrm{im}(L_A)) = L_D(\mathcal{C}(A)) \cong \mathcal{C}(A).$$

Since L_E is surjective, we have

$$\mathcal{C}(AE) = \mathrm{im}(L_{AE}) = \mathrm{im}(L_A \circ L_E) = \mathrm{im}(L_A) = \mathcal{C}(A).$$

Since L_D is injective, we have

$$\mathcal{N}(DA) = \ker(L_{DA}) = \ker(L_D \circ L_A) = \ker(L_A) = \mathcal{N}(A).$$

We have

$$b \in \mathcal{N}(AE) \iff AEb = 0 \iff Eb \in \mathcal{N}(A) \iff b \in L_{E^{-1}}\mathcal{N}(A).$$

Thus $\mathcal{N}(AE) = L_{E^{-1}}\mathcal{N}(A) \cong \mathcal{N}(A)$, because $L_{E^{-1}}$ is an isomorphism.

From (1) and (2), we now have

$$\mathcal{R}(DA) = \mathcal{C}((DA)^t) = \mathcal{C}(A^tD^t) = \mathcal{C}(A^t) = \mathcal{R}(A),$$

and

$$\mathcal{R}(AE) = \mathcal{C}((AE)^t) = \mathcal{C}(E^tA^t) \cong \mathcal{C}(A^t) = \mathcal{R}(A).$$

□

3.4 Elementary Matrices

We will define elementary matrices in $\mathcal{M}_{m\times n}(k)$ and study their properties.

Let e_{ij} denote the matrix in $\mathcal{M}_{m\times n}(k)$ that has a 1 in the (i,j)-entry and zeros in all other entries. The collection of these matrices forms a basis of $\mathcal{M}_{m\times n}(k)$ called the standard basis of $\mathcal{M}_{m\times n}$. Multiplication of these simple matrices satisfies the formula

$$e_{ij}e_{kl} = \delta(j,k)e_{il},$$

where $\delta(j,k)$ is defined by $\delta(j,k) = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k. \end{cases}$

When $m = n$, there are three types of matrices that we single out. They are known as the *elementary matrices.*

Definition 3.20.

1. *Let $E_{ij}(a)$ be the matrix $I_n + ae_{ij}$, where $a \in k$ and $i \neq j$.*

2. *Let $D_i(a)$ be the matrix $I_n + (a-1)e_{ii}$, where $a \in k$, $a \neq 0$.*

3. *Let P_{ij}, $i \neq j$, be the matrix $I_n - e_{ii} - e_{jj} + e_{ij} + e_{ji}$.*

Note that $E_{ij}(0)$ is the identity matrix I_n, and that $D_i(a)$ is the matrix obtained from I_n by replacing the (i,i)-entry of I_n with a. Also note that P_{ij} can be obtained from the identity matrix by either interchanging the i^{th} and j^{th} rows, or by interchanging the i^{th} and j^{th} columns.

The notation we are using for the elementary matrices, which is quite standard, does not indicate that the matrices under consideration are $n \times n$ matrices. We will always assume that the elementary matrices under consideration always have the appropriate size in our computations.

We now relate the basic row and column operations of matrices to the elementary matrices.

Proposition 3.21. *Let $A \in \mathcal{M}_{m\times n}(k)$. Let $A_1 = E_{ij}(a)A$, $A_2 = D_i(a)A$, $A_3 = P_{ij}A$, where the elementary matrices lie in $\mathcal{M}_{m\times m}(k)$. Then*

A_1 is obtained from A by adding a times row j of A to row i of A.

A_2 is obtained from A by multiplying row i of A by a.

A_3 is obtained from A by interchanging rows i and j of A.

Proposition 3.22. *Let $A \in \mathcal{M}_{m\times n}(k)$. Let $A_4 = AE_{ij}(a)$, $A_5 = AD_i(a)$, $A_6 = AP_{ij}$, where the elementary matrices lie in $\mathcal{M}_{n\times n}(k)$. Then*

A_4 is obtained from A by adding a times column i of A to column j of A.

A_5 is obtained from A by multiplying column i of A by a.

A_6 is obtained from A by interchanging columns i and j of A.

Exercise 8 asks for a proof of the last two propositions. The operations described in Propositions 3.21, 3.22 are called the elementary row and column operations.

Proposition 3.23.

1. *$E_{ij}(a)E_{ij}(b) = E_{ij}(a+b)$ for all $a, b \in k$. Therefore, $E_{ij}(a)^{-1} = E_{ij}(-a)$.*

2. *$D_i(a)D_i(b) = D_i(ab)$ for all nonzero $a, b \in k$. Therefore, $D_i(a)^{-1} = D_i(a^{-1})$, for nonzero $a \in k$.*

3. *$P_{ij}^{-1} = P_{ij}$.*

Proof. We will prove these results by applying Proposition 3.21. We have $E_{ij}(a)E_{ij}(b)I = E_{ij}(a+b)I$ because each expression adds $a+b$ times row j of I to row i of I. Since $I = E_{ij}(0) = E_{ij}(a+(-a)) = E_{ij}(a)E_{ij}(-a)$, it follows that $E_{ij}(a)^{-1} = E_{ij}(-a)$.

Similarly,$D_i(a)D_i(b)I = D_i(ab)I$ because each expression multiplies row i of I by ab.

Finally, $P_{ij}P_{ij}I = I$, because the left side interchanges rows i and j of I twice, which has the net effect of doing nothing. Therefore, $P_{ij}P_{ij} = I$ and (3) follows from this. □

Here are several useful cases when products of elementary matrices commute.

Proposition 3.24.

1. *$E_{ij}(a)E_{il}(b) = E_{il}(b)E_{ij}(a)$ for all $a, b \in k$.*

2. *$E_{il}(a)E_{jl}(b) = E_{jl}(b)E_{il}(a)$ for all $a, b \in k$.*

3. *$D_i(a)D_j(b) = D_j(b)D_i(a)$ for all nonzero $a, b \in k$.*

Proof. (1) Since $i \neq j$ and $i \neq l$, we have

$$\begin{aligned} E_{ij}(a)E_{il}(b) &= (I_n + ae_{ij})(I_n + be_{il}) \\ &= I_n + ae_{ij} + be_{il} = (I_n + be_{il})(I_n + ae_{ij}) = E_{il}(b)E_{ij}(a). \end{aligned}$$

(2) Since $i \neq l$ and $j \neq l$, we have

$$\begin{aligned} E_{il}(a)E_{jl}(b) &= (I_n + ae_{il})(I_n + be_{jl}) \\ &= I_n + ae_{il} + be_{jl} = (I_n + be_{jl})(I_n + ae_{il}) = E_{jl}(b)E_{il}(a). \end{aligned}$$

(3) Since $e_{ii}e_{jj} = e_{jj}e_{ii}$, we have

$$\begin{aligned} D_i(a)D_j(b) &= (I_n + (a-1)e_{ii})(I_n + (b-1)e_{jj}) \\ &= I_n + (a-1)e_{ii} + (b-1)e_{jj} + (a-1)(b-1)e_{ii}e_{jj} \\ &= (I_n + (b-1)e_{jj})(I_n + (a-1)e_{ii}) = D_j(b)D_i(a). \end{aligned}$$

□

3.5 Permutations and Permutation Matrices

This section is important for Chapter 6 as well as for further results on elementary matrices.

Definition 3.25. *A function $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is called a* permutation *if σ is bijective. The set of all permutations $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is denoted S_n. If $\sigma(i) = i$, $1 \leq i \leq n$, then σ is called the* identity permutation *and is written $\sigma = id$.*

There are $n!$ permutations in S_n. If $\sigma \in S_m$ and $m < n$, then we can consider $\sigma \in S_n$ by setting $\sigma(i) = i$ for all i satisfying $m+1 \leq i \leq n$. In this way we have $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n \subseteq \cdots$.

If $\sigma_1, \sigma_2 \in S_n$, then $\sigma_1 \circ \sigma_2 \in S_n$. We call $\sigma_1 \circ \sigma_2$ the product of σ_1 and σ_2.

Definition 3.26. *A permutation $\sigma \in S_n$ is called a* transposition *if there exist $i, j \in \{1, \ldots, n\}$, $i \neq j$, such that $\sigma(i) = j$, $\sigma(j) = i$, and $\sigma(l) = l$ for all $l \in \{1, \ldots, n\}$ with $l \neq i, j$. Let σ_{ij} denote this transposition.*

Proposition 3.27. *Each $\sigma \in S_n$, $n \geq 2$, is a product of transpositions in S_n.*

Proof. If $\sigma = id$, then $\sigma = \sigma_{12} \circ \sigma_{12}$. We now prove the result by induction on n. Assume that $n \geq 2$. Let $\sigma(n) = i$ and let $\sigma' = \sigma_{in} \circ \sigma$. Then $\sigma'(n) = \sigma_{in}(i) = n$, so we can regard $\sigma' \in S_{n-1}$. By induction, σ' is a product of transpositions in $S_{n-1} \subseteq S_n$. Therefore, $\sigma = \sigma_{in} \circ \sigma_{in} \circ \sigma = \sigma_{in} \circ \sigma'$ is a product of transpositions in S_n. □

In Theorem 6.4 of Chapter 6, we will prove that the number of transpositions in any representation of σ as a product of transpositions is uniquely determined modulo 2.

Definition 3.28. *A matrix $P = (p_{ij}) \in \mathcal{M}_{n\times n}(k)$ is called a* permutation matrix *if there exists a permutation $\sigma \in S_n$ such that*

$$p_{ij} = \begin{cases} 0, & \textit{if } i \neq \sigma(j) \\ 1, & \textit{if } i = \sigma(j). \end{cases}$$

Denote this permutation matrix by P_σ.

Each column of a permutation matrix $P \in \mathcal{M}_{n\times n}(k)$ consists of exactly one 1 and $n-1$ zeros. The same holds for each row of P because if $p_{ij} = p_{il} = 1$, then $\sigma(j) = i = \sigma(l)$. Then $j = l$ because σ is injective.

Conversely, if each row and each column of an $n \times n$ matrix has exactly one 1 and $n-1$ zeros, then the matrix is a permutation matrix. (See Exercise 14.)

The elementary matrix P_{ij} defined in Section 3.4 is the same as the permutation matrix P_σ where σ is the transposition σ_{ij}. We call P_{ij} an elementary permutation matrix.

Let $\mathcal{P}_{n\times n}$ denote the set of permutation matrices.

Proposition 3.29. *There is a bijection $f : S_n \to \mathcal{P}_{n\times n}$ given by $\sigma \mapsto P_\sigma$ with the property $f(\sigma_1 \circ \sigma_2) = f(\sigma_1)f(\sigma_2)$. In other words, $P_{\sigma_1\circ\sigma_2} = P_{\sigma_1}P_{\sigma_2}$.*

Proof. The definition of matrix multiplication shows that

$$\begin{aligned} (i,j)\text{-entry of } P_{\sigma_1}P_{\sigma_2} &= \begin{cases} 1, & \text{if } i = \sigma_1(l) \text{ and } l = \sigma_2(j) \text{ for some } l, \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} 1, & \text{if } i = \sigma_1 \circ \sigma_2(j), \\ 0, & \text{otherwise,} \end{cases} \\ &= (i,j)\text{-entry of } P_{\sigma_1\circ\sigma_2}. \end{aligned}$$

Thus $P_{\sigma_1 \circ \sigma_2} = P_{\sigma_1} P_{\sigma_2}$.

Suppose $f(\sigma_1) = f(\sigma_2)$. Then $P_{\sigma_1} = P_{\sigma_2}$. This implies $\sigma_1(j) = \sigma_2(j)$ for all $j = 1, 2, \ldots, n$. Therefore $\sigma_1 = \sigma_2$, so f is injective.

Finally, f is surjective because every permutation matrix has the form $P_\sigma = f(\sigma)$. Therefore f is a bijection. □

Proposition 3.30. *Let P_σ be a permutation matrix. Then*

1. $P_\sigma^{-1} = P_{\sigma^{-1}}$.

2. $P_\sigma^t = P_\sigma^{-1} = P_{\sigma^{-1}}$, *so P_σ^t is a permutation matrix.*

3. *P_σ is a product of elementary permutation matrices P_{ij}.*

4. *If P_σ is an elementary permutation matrix, then $P_\sigma = P_\sigma^t = P_\sigma^{-1}$.*

Proof. 1. Proposition 3.29 implies that $P_{\sigma^{-1}} P_\sigma = P_{id} = P_\sigma P_{\sigma^{-1}}$. Since $P_{id} = I_n$, it follows that $P_\sigma^{-1} = P_{\sigma^{-1}}$.

2. Direct matrix multiplication shows that $P_\sigma P_\sigma^t = P_\sigma^t P_\sigma = I_n$. Therefore, $P_\sigma^t = P_\sigma^{-1}$. Alternatively, it follows from Definition 3.28 that $P_\sigma^t = (q_{ij})$ where

$$q_{ij} = \begin{cases} 0, & \text{if } j \neq \sigma(i) \\ 1, & \text{if } j = \sigma(i) \end{cases} = \begin{cases} 0, & \text{if } i \neq \sigma^{-1}(j) \\ 1, & \text{if } i = \sigma^{-1}(j). \end{cases}$$

It follows that $P_\sigma^t = P_{\sigma^{-1}}$.

3. This follows from Propositions 3.27 and 3.29.

4. If P_σ is an elementary permutation matrix, then σ is a transposition. Then $\sigma^{-1} = \sigma$ and the result follows from (1) and (2).

□

The following result will be useful later.

Proposition 3.31.

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

Proof. We have

$$P_\sigma \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{j=1}^{n} x_j(\ j^{th} \text{ column of } P_\sigma).$$

Suppose that $\sigma(j) = i$. Then the j^{th} column of P_σ has a single 1 in the i^{th} row. Thus the i^{th} row of $P_\sigma \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is $x_j = x_{\sigma^{-1}(i)}$. □

3.6 The Rank of a Matrix

Proposition 3.32. *Let $A \in \mathcal{M}_{m\times n}(k)$. Then there exists an invertible matrix $D \in \mathcal{M}_{m\times m}(k)$, an invertible matrix $E \in \mathcal{M}_{n\times n}(k)$, and an integer $r \geq 0$ such that the matrix DAE has the form*

$$DAE = \begin{pmatrix} I_r & B_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}.$$

In addition we have

1. *D is a product of elementary matrices.*

2. *E is a permutation matrix.*

3. *$0 \leq r \leq \min\{m, n\}$.*

Proof. We will show that A can be put in the desired form by a sequence of elementary row operations and a permutation of the columns. It follows from Proposition 3.21 that performing elementary row operations on A results in multiplying A on the left by $m \times m$ elementary matrices. Each elementary matrix is invertible by Proposition 3.23. Let D be the product of these elementary matrices. Then D is an invertible $m \times m$ matrix. Similarly, it follows from Proposition 3.22 and Proposition 3.30 (3) that permuting the columns of A results in multiplying A on the right by an $n \times n$ permutation matrix E.

If $A = 0$, then let $r = 0$, $D = I_m$, and $E = I_n$. Now assume that $A \neq 0$. Then $a_{ij} \neq 0$ for some i, j.

Interchange rows 1 and i and then interchange columns 1 and j. (Replace A with $P_{1i}AP_{1j}$.) This lets us assume that $a_{11} \neq 0$.

Multiply row 1 by a_{11}^{-1}. (Replace A with $D_1(a_{11}^{-1})A$.) This lets us assume that $a_{11} = 1$.

Add $-a_{i1}$ times row 1 to row i, $2 \leq i \leq m$. (Replace A with

$$E_{m1}(-a_{m1}) \cdots E_{31}(-a_{31})E_{21}(-a_{21})A.)$$

This lets us assume that $a_{11} = 1$ and $a_{i1} = 0$ for $2 \leq i \leq m$.

Suppose by induction that we have performed elementary row operations on A and permuted the columns of A such that for some l satisfying $1 \leq l < \min\{m, n\}$, we have

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

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

If $a_{ij} = 0$ for all $l+1 \leq i \leq m$ and $l+1 \leq j \leq n$, then let $r = l$ and we are done. Otherwise assume that $a_{ij} \neq 0$ for some i, j with $i \geq l+1$, $j \geq l+1$. Then interchange rows $l+1$ and i, and then interchange columns $l+1$ and j. This lets us assume that $a_{l+1,l+1} \neq 0$. Multiply row $l+1$ by $a_{l+1,l+1}^{-1}$. This lets us assume that $a_{l+1,l+1} = 1$. Add $-a_{i,l+1}$ times row $l+1$ to row i, for i with $1 \leq i \leq m$, but $i \neq l+1$. This lets us assume that $a_{l+1,l+1} = 1$ and $a_{i,l+1} = 0$ for all i satisfying $1 \leq i \leq m$ but $i \neq l+1$. Observe that columns j, where $1 \leq j \leq l$, remain unchanged throughout this step.

By induction, this procedure either eventually ends as above or $r = \min\{m, n\}$. This puts A in the required form. □

The matrix D in Proposition 3.32 can be found by applying to I_m the same sequence of elementary row operations that were used in the proof of the proposition. The matrix E in Proposition 3.32 can be found by applying to I_n the same sequence of column interchanges that were used in the proof of the proposition.

Corollary 3.33. *Let $A \in \mathcal{M}_{n\times n}(k)$ be an invertible matrix. Then A is a product of elementary matrices.*

Proof. By Proposition 3.32, there exist invertible matrices $D, E \in \mathcal{M}_{n\times n}(k)$ such that D and E are each products of elementary matrices and DAE has the form given in Proposition 3.32. Since A is invertible, DAE is also invertible, so we must have $m - r = 0$. Thus $r = m = n$. Thus $DAE = I_n$. Then $A = D^{-1}E^{-1}$, so A is a product of elementary matrices. □

Corollary 3.34. *Let $A \in \mathcal{M}_{n\times n}(k)$ be an invertible matrix and suppose that D and E are as in Proposition 3.32 such that $DAE = I_n$. Then $A^{-1} = ED$.*

Proof. If $DAE = I_n$, then $A = D^{-1}E^{-1}$, so $A^{-1} = ED$. □

Proposition 3.35. *Let $A \in \mathcal{M}_{m\times n}(k)$. Then* $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$.

Proof. We find invertible matrices D, E as in Proposition 3.32 Then Proposition 3.18 implies that $\dim(\mathcal{C}(A)) = \dim(\mathcal{C}(DA)) = \dim(\mathcal{C}(DAE))$, and $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(DA)) = \dim(\mathcal{R}(DAE))$. This shows that we can assume from the start that

$$A = \begin{pmatrix} I_r & B_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}.$$

The first r rows of A are linearly independent and the remaining $m-r$ rows of A are zero. Therefore, $\dim(\mathcal{R}(A)) = r$. The first r columns of A are linearly independent. The columns of A can be regarded as lying in $k^{(r)}$. Therefore $\dim(\mathcal{C}(A)) = r$. □

Definition 3.36. *Let $A \in \mathcal{M}_{m\times n}(k)$. The* rank *$r(A)$ of A is the dimension of the row space or column space of A. Thus $r(A) = \dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$.*

It is a very surprising result that $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$. Since $\mathcal{R}(A) \subseteq k^{(n)}$ and $\mathcal{C}(A) \subseteq k^{(m)}$, there seems to be no apparent reason why the two subspaces should have the same dimension. In Chapter 4, we will see additional connections between these two subspaces that will give more insight into the situation.

We end this section with a theorem that gives a number of equivalent conditions that guarantee an $n \times n$ matrix is invertible.

Theorem 3.37. *Let $A \in \mathcal{M}_{n\times n}(k)$. The following statements are equivalent.*

1. *A is invertible.*
2. *$\mathcal{R}(A) = k^{(n)}$.*
3. *The rows of A are linearly independent.*
4. *Dim$(\mathcal{R}(A)) = n$.*
5. *$\mathcal{C}(A) = k^{(n)}$.*

6. *The columns of A are linearly independent.*

7. *$Dim(\mathcal{C}(A)) = n$.*

8. *There exists $B \in \mathcal{M}_{n\times n}(k)$ such that $BA = I_n$.*

9. *There exists $C \in \mathcal{M}_{n\times n}(k)$ such that $AC = I_n$.*

10. *$L_A : k^{(n)} \to k^{(n)}$ is an isomorphism.*

11. *$L_A : k^{(n)} \to k^{(n)}$ is an injection.*

12. *$L_A : k^{(n)} \to k^{(n)}$ is a surjection.*

13. *A is a product of elementary matrices.*

Proof. Statements (2)-(4) are equivalent by Proposition 1.14 and Exercise 13 of Chapter 1. The same reasoning shows that statements (5)-(7) are equivalent. Statements (10)-(12) are equivalent by Proposition 2.9.

(8) ⇒ (2): We have $k^{(n)} = \mathcal{R}(I_n) = \mathcal{R}(BA) \subseteq \mathcal{R}(A) \subseteq k^{(n)}$. Therefore equality holds throughout and (2) holds.

(2) ⇒ (8): Since the $\mathcal{R}(A) = k^{(n)}$, there exist constants $b_{i1}, \ldots, b_{in}$ in k such that $b_{i1}(\text{first row of } A) + \cdots + b_{in}(n^{th} \text{ row of } A) = e_i$, where e_i is the i^{th} vector in the standard basis of $k^{(n)}$. If we let B be the $n \times n$ matrix whose (i,j)-entry is b_{ij}, then $BA = I_n$.

Similarly, if we apply the same reasoning to columns and column operations as we did to rows and row operations in the proof that (2) and (8) are equivalent, then we easily see that (5) and (9) are equivalent.

Since $[L_A]_{\epsilon_n}^{\epsilon_n} = A$ by Lemma 3.11, it follows that (1) and (10) are equivalent by Exercise 5.

The next step is to prove that (1), (8), (10), (11) are equivalent.

(8) ⇒ (11): Lemma 3.11 and Propositions 3.4, 3.5 imply that $L_B L_A = L_{BA} = L_{I_n} = 1_{k^{(n)}}$. Then (11) follows from Exercise 9b in Chapter 2.

(11) ⇒ (10) ⇒ (1) comes from above and (1) ⇒ (8) is trivial.

Statements (1), (9), (10), (12) are proved equivalent in the same way as the equivalence of (1), (8), (10), (11).

It remains to prove that (1) and (13) are equivalent. Since elementary matrices are invertible and products of invertible matrices are invertible by Proposition 3.8, it is clear that (13) implies (1). Finally, (13) implies (1) by Corollary 3.33. □

3.7 Systems of Linear Equations

We begin by describing an algorithm to solve systems of linear equations. Then we will apply these results to finding bases of $\ker(f)$ and $\operatorname{im}(f)$ for a linear transformation f. This will extend the results given at the end of Chapter 2.

Consider the following system of m linear equations in n variables with coefficients in k.

$$\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m
\end{gathered}$$

Let $A = (a_{ij})_{m\times n}$, $x = (x_j)_{n\times 1}$, and $c = (c_i)_{m\times 1}$. Then the system of equations above is equivalent to the matrix equation $Ax = c$. When $c_1 = \cdots = c_m = 0$, this system is called a *homogeneous* system of linear equations. Otherwise, the system is called an *inhomogeneous* system of linear equations.

In Chapter 2, we constructed a linear transformation $f : k^{(n)} \to k^{(m)}$ with the property that $f(b) = c$ if and only if $b = (b_1, \ldots, b_n)$ is a solution for $x = (x_1, \ldots, x_n)$ in the system of linear equations above. It was shown that if b is one solution, then all solutions are given by $b+\ker(f)$. It is easy to check that f is the linear transformation L_A defined before Lemma 3.11. That is, $f(b) = Ab$. Thus, $\ker(f) = \{b \in k^{(n)} | Ab = 0\} = \ker(L_A) = \mathcal{N}(A)$. It follows that finding the solutions of a homogeneous system of linear equations is equivalent to finding either the null space of a matrix A or the kernel of the associated linear transformation L_A. Also,

$$\operatorname{im}(f) = \{c \in k^{(m)} | Ab = c \text{ for some } b \in k^{(n)}\} = \operatorname{im}(L_A) = \mathcal{C}(A).$$

Thus

$$n = \dim(\ker(f)) + \dim(\operatorname{im}(f)) = \dim(\mathcal{N}(A)) + \dim(\mathcal{C}(A)).$$

Proposition 3.38. *Let $A \in \mathcal{M}_{m\times n}(k)$ and $c \in k^{(m)}$. Suppose that D and E are invertible matrices as in Proposition 3.32 and suppose that r is the rank of A. Then $Ax = c$ has a solution if and only if the last $m - r$ coordinates of Dc are zero.*

Proof. Since D is an invertible $m \times m$ matrix, the equation $Ax = c$ has a solution if and only if the equation $DAx = Dc$ has a solution, and this holds if and only if $Dc \in \mathcal{C}(DA) = \mathcal{C}(DAE)$. The column space of DAE consists of those vectors in $k^{(m)}$ whose last $m - r$ coordinates are zero. Therefore, $Ax = c$ has a solution if and only if the last $m - r$ coordinates of Dc are zero. □

Proposition 3.39. *Let $A \in \mathcal{M}_{m\times n}(k)$ and $c \in k^{(m)}$. Suppose that D and E are invertible matrices as in Proposition 3.32 and suppose that r is the rank of A. Assume that the equation $Ax = c$ has a solution. Let*

$$Dc = \begin{pmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{m\times 1} \quad \textit{and} \quad d = \begin{pmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{n\times 1},$$

where the last $m-r$ coordinates of Dc equal zero and the last $n-r$ coordinates of d equal zero. (The notation is valid since $r = \dim(\mathcal{R}(A)) \le \min\{m, n\}$.) Then a solution of $Ax = c$ is given by $x = Ed$.

Proof. We have

$$DAEd = d_1 \begin{pmatrix} 1^{st}\text{ column} \\ \text{of } DAE \end{pmatrix} + \cdots + d_r \begin{pmatrix} r^{th}\text{ column} \\ \text{of } DAE \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{m\times 1} = Dc.$$

Multiply by D^{-1} on the left to conclude $AEd = c$. Thus, $x = Ed$ is a solution of the equation $Ax = c$. □

We now describe $\mathcal{N}(A)$. We have $\mathcal{N}(A) = \mathcal{N}(DA) = L_E\mathcal{N}(DAE)$. (See Proposition 3.19.) We first compute a basis of $\mathcal{N}(DAE)$ and $\dim(\mathcal{N}(DAE))$. Since DAE is an $m \times n$ matrix and $r(DAE) = \dim(\mathcal{C}(DAE)) = r$, it follows that $\dim(\mathcal{N}(DAE)) = n - r$.

Let $DAE = (a_{ij})_{m\times n}$. It follows from Proposition 3.32 that $a_{ij} = 0$ for $r+1 \leq i \leq m$, and the upper $r \times r$ submatrix of DAE is the $r \times r$ identity matrix I_r. The null space of DAE corresponds to the solutions of the following homogeneous system of r linear equations in n variables.

$$\begin{gathered} x_1 + a_{1,r+1}x_{r+1} + \cdots + a_{1n}x_n = 0 \\ x_2 + a_{2,r+1}x_{r+1} + \cdots + a_{2n}x_n = 0 \\ \vdots \\ x_r + a_{r,r+1}x_{r+1} + \cdots + a_{rn}x_n = 0 \end{gathered}$$

The vectors

$$\begin{pmatrix} -a_{1,r+1} \\ \vdots \\ -a_{r,r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -a_{1,r+2} \\ \vdots \\ -a_{r,r+2} \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} -a_{1n} \\ \vdots \\ -a_{rn} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

are $n-r$ linearly independent vectors in $\mathcal{N}(DAE)$, which therefore form a basis of $\mathcal{N}(DAE)$.

Finally we have $\mathcal{N}(A) = L_E\mathcal{N}(DAE)$. Since E is a permutation matrix, the coordinates of the basis vectors of $\mathcal{N}(DAE)$ are permuted to obtain a basis of $\mathcal{N}(A)$. The complete solution for the original system is $b + \mathcal{N}(A)$, where b is any solution to $Ax = c$. If a solution exists, then we saw above that we may take $b = Ed$.

3.8 Bases of Subspaces

A number of problems can be solved using the techniques above. Here are several types of problems and some general procedures based on elementary row operations.

1. Let V be a vector space with basis $\beta = \{v_1, \ldots, v_n\}$. Let $Y \subset V$ be a nonzero subspace of V and suppose $Y =< y_1, \ldots, y_m >$. Find a basis of Y.

   Although Corollary 1.10(3) implies some subset of $\{y_1, \ldots, y_m\}$ contains a basis of Y, the Corollary does not produce a method to actually find the basis. Here is a method.

   We follow the notation that precedes Lemma 3.11. Suppose that $y_i = \sum_{j=1}^{n} a_{ij} v_j$, $1 \leq i \leq m$. Let $A = (a_{ij})_{m \times n}$. Then

   $$\phi_\beta(y_i) = (a_{i1}, a_{i2}, \ldots, a_{in}),$$

   the i^{th} row of A. Thus $\phi_\beta(Y) = \mathcal{R}(A)$, the row space of A. Since Y is a nonzero subspace of V, it is clear that A is a nonzero matrix. Therefore a basis of Y is given by ϕ_β^{-1}(a basis of $\mathcal{R}(A)$).

   Find invertible matrices D, E as in Proposition 3.32. Then $\mathcal{R}(A) = \mathcal{R}(DA)$. The matrix DA has r nonzero rows which are linearly independent. These rows form a basis of $\mathcal{R}(A)$.

2. Let $y_1, \ldots, y_m \in k^{(n)}$. Determine if these vectors are linearly independent.

   Let A be an $m \times n$ matrix whose rows are $y_1, \ldots, y_m$. Find invertible matrices D, E as in Proposition 3.32. The vectors $y_1, \ldots, y_m$ are linearly independent if and only if $r(A) = r(DAE) = m$, i.e., each row of DAE is nonzero. The rows of A are linearly dependent if and only if elementary row operations can be performed on A to obtain a row of zeros.

3. Let $W \subset k^{(m)}$ be a subspace. Let $\{u_1, \ldots, u_n\}$ be a set of generators of W. Let $c \in k^{(m)}$. Determine whether $c \in W$. If $c \in W$, find $b_1, \ldots, b_n$ in k such that $c = b_1 u_1 + \cdots + b_n u_n$.

   Let S be a basis of W. Then $c \in W$ if and only if $S \bigcup \{c\}$ is a linearly dependent set. (This statement would be false if we replaced S by an arbitrary set of generators of W. See Exercise 13.) To find $b_1, \ldots, b_n$, it is more convenient to change the problem to one about solving a system of linear equations. Let A be the $m \times n$ matrix whose columns are the vectors $\{u_1, \ldots, u_n\}$. Then $c \in W$ if and only if the equation $Ax = c$ has a solution.

4. Let $f : V \to W$ be a linear transformation. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V and let $\gamma = \{w_1, \ldots, w_m\}$ be a basis of W. Let $A = [f]_\beta^\gamma = (a_{ij})$. Find a basis of $\text{im}(f)$.

   Proposition 3.12 implies that

$$\begin{aligned}\phi_\gamma(\text{im}(f)) &= \phi_\gamma(f(V)) = L_A(\phi_\beta(V)) \\ &= L_A(k^{(n)}) = \{Ab \mid b \in k^{(n)}\} = \mathcal{C}(A).\end{aligned}$$

   Then a basis of $\text{im}(f)$ is given by ϕ_γ^{-1}(a basis of $\mathcal{C}(A)$). A basis of $\mathcal{C}(A)$ can be computed by applying elementary column operations on A in the same manner as was used to compute a basis of $\mathcal{R}(A)$ in (1).

Exercises

1. Prove Proposition 3.3 and find a basis of $\mathcal{M}_{m\times n}(k)$.

2. Prove Proposition 3.6. (An easy method is to interpret each matrix as the matrix of a linear transformation with respect to some basis, and then to recall that composition of functions is associative. For a second proof, compute the entries of $(AB)C$ and $A(BC)$ directly.)

3. Matrix multiplication is not commutative, even for square matrices.

4. Prove Proposition 3.13.

5. Using our usual notation, $f \in \mathcal{L}(V, W)$ is an isomorphism if and only if $[f]_\beta^\gamma$ is an invertible matrix.

6. Let $A \in \mathcal{M}_{m\times n}(k)$ and let $B \in \mathcal{M}_{n\times m}(k)$.

   (a) Suppose that $AB = I_m$. Then $m \leq n$.

   (b) Suppose that $AB = I_m$ and $m = n$. Then $BA = I_m$. Thus, if $m = n$, then $AB = I_m \iff BA = I_m$.

   (c) Suppose that $AB = I_m$ and $m < n$. Then $BA \neq I_n$.

7. Prove Proposition 3.15 (1), (2).

8. Prove Propositions 3.21 and 3.22.

9. The elementary matrix P_{ij} can be expressed as a product of the other two types of elementary matrices $E_{ij}(a)$ and $D_i(a)$. (First solve this problem for 2×2 matrices.)

10. Give another proof of Proposition 3.23 by using the formulas $E_{ij}(a) = I_n + ae_{ij}$, $D_i(a) = I_n + (a-1)e_{ii}$, and $P_{ij} = I_n - e_{ii} - e_{jj} + e_{ij} + e_{ji}$.

11. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V and let $\gamma = \{w_1, \ldots, w_m\}$ be a basis of W. Set $V_j =< v_j >$ and $W_i =< w_i >$ so that $V = \bigoplus_{j=1}^{n} V_j$ and $W = \bigoplus_{i=1}^{m} W_i$. Let $f \in \mathcal{L}(V, W)$ and suppose $[f]_\beta^\gamma = (a_{ij})$. Then show $f = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}$ where $f_{ij} \in \mathcal{L}(V_j, W_i)$ and $f_{ij}(v_j) = a_{ij}w_i$. Compare this exercise to Exercise 12 of Chapter 2.

12. Justify the procedures given in Section 3.8 to solve the problems that are posed.

13. Justify the statement about $S \bigcup \{c\}$ in problem 3 of Section 3.8.

14. Suppose that each row and each column of $A \in \mathcal{M}_{n \times n}(k)$ has exactly one 1 and $n-1$ zeros. Prove that A is a permutation matrix.

15. Let A be an $m \times n$ matrix and let $r = \text{rank}(A)$. Then

    (a) $r \leq \min\{m, n\}$.

    (b) $r = \dim(\operatorname{im} L_A)$.

    (c) $\dim(\ker L_A) = n - r$.

16. In Proposition 3.12, show that $\phi_\beta(\ker(f)) = \ker(L_A)$ and $\phi_\gamma(\operatorname{im}(f)) = \operatorname{im}(L_A)$.

17. Suppose that S and T are two sets and that $f : S \to T$ is a bijection. Suppose that S is a vector space. Then there is a unique way to make T a vector space such that f is an isomorphism of vector spaces. Similarly, if T is a vector space, then there is a unique way to make S a vector space such that f is an isomorphism of vector spaces.

18. Use Exercise 17 to develop a new proof of Proposition 3.4. Either start with the vector space structure of $\mathcal{M}_{m \times n}(k)$ to develop the vector space structure of $\mathcal{L}(V, W)$, or start with the vector space structure of $\mathcal{L}(V, W)$ to develop the vector space structure of $\mathcal{M}_{m \times n}(k)$.

19. Let $A, A_1, A_2 \in \mathcal{M}_{m\times n}(k)$, $B, B_1, B_2 \in \mathcal{M}_{n\times p}(k)$, $C \in \mathcal{M}_{p\times q}(k)$, and let $a \in k$. Then the following statements hold. (Compare with Proposition 2.12.)

    (a) $(A_1 + A_2)B = A_1B + A_2B$

    (b) $A(B_1 + B_2) = AB_1 + AB_2$

    (c) $a(AB) = (aA)B = A(aB)$

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

    Note that when $m = n = p = q$, statements (1) - (4) imply that $\mathcal{M}_{n\times n}(k)$ is an associative algebra.

20. A *diagonal* matrix $A \in \mathcal{M}_{n\times n}(k)$ is a matrix that can be written $\sum_{i=1}^n a_i e_{ii}$. In other words, $A = (a_{ij})$ is a diagonal matrix if $a_{ij} = 0$ for all i, j where $i \neq j$.

    Let $D_1 = \sum_{i=1}^n a_i e_{ii}$ and $D_2 = \sum_{i=1}^n b_i e_{ii}$ be two diagonal matrices. Then $D_1 D_2 = \sum_{i=1}^n a_i b_i e_{ii}$. In particular, $D_1 D_2 = D_2 D_1$ for any two diagonal $n \times n$ matrices.

Chapter 4

The Dual Space

4.1 Basic Results

Let V be a vector space over k, $\dim V = n$. We shall consider k as a vector space of dimension one over k by identifying k with the one-dimensional vector space $k^{(1)}$. That is, the element $a \in k$ is identified with the vector $(a) \in k^{(1)}$.

Definition 4.1. *The vector space $\mathcal{L}(V, k)$ is called the* dual space *of V, or simply the* dual *of V and is written V^*. The elements of V^* are called* linear functionals.

Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V. We use Proposition 3.1 to define elements $\phi_1, \ldots, \phi_n \in \mathcal{L}(V, k) = V^*$ as follows. Let ϕ_i denote the element in V^* such that

$$\phi_i(v_j) = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases}$$

Proposition 4.2. *The elements $\phi_1, \ldots, \phi_n$ form a basis of V^*.*

Proof. Let $f \in V^*$ and suppose that $f(v_j) = c_j$, $c_j \in k$, $1 \leq j \leq n$. Let $g = \sum_{i=1}^n c_i\phi_i$. Then $g(v_j) = \sum_{i=1}^n c_i\phi_i(v_j) = c_j$. Thus $f = g$ by Proposition 3.1(1), because both functions agree on the basis β. This shows that $\phi_1, \ldots, \phi_n$ span V^*.

Suppose that $\sum_{i=1}^n b_i\phi_i = 0$. Evaluation at v_j gives $0 = \sum_{i=1}^n b_i\phi_i(v_j) = b_j$. Therefore each $b_j = 0$ and so $\phi_1, \ldots, \phi_n$ are linearly independent. Therefore, the elements $\phi_1, \ldots, \phi_n$ form a basis of V^*. □

The basis in Proposition 4.2 is called the dual basis of V^* to β and we write $\beta^* = \{\phi_1, \ldots, \phi_n\}$ to denote this. Since $\dim V^* = \dim \mathcal{L}(V, k) = n$, by Proposition 3.4, we could have shortened the proof of Proposition 4.2 by making use of Proposition 1.14.

Corollary 4.3. *Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V and let $\beta^* = \{\phi_1, \ldots, \phi_n\}$ be the dual basis of V^* to β. Let $v \in V$ and let $f \in V^*$. Then $v = \sum_{j=1}^n \phi_j(v)v_j$ and $f = \sum_{i=1}^n f(v_i)\phi_i$.*

Proof. The proof of Proposition 4.2 shows that $f = \sum_{i=1}^n f(v_i)\phi_i$. Let $v \in V$. Then $v = \sum_{i=1}^n a_i v_i$, with each $a_i \in k$ uniquely determined. Then $\phi_j(v) = \sum_{i=1}^n a_i \phi_j(v_i) = a_j$. Therefore $v = \sum_{j=1}^n \phi_j(v)v_j$. □

We now begin to develop the connections between dual spaces, matrix representations of linear transformations, transposes of matrices, and exact sequences.

Let V, W be finite dimensional vector spaces over k. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of V and let $\gamma = \{w_1, \ldots, w_m\}$ be an ordered basis of W. Let $\beta^* = \{\phi_1, \ldots, \phi_n\}$ and $\gamma^* = \{\psi_1, \ldots, \psi_m\}$ be the corresponding ordered dual bases of V^* and W^*. Let $f \in \mathcal{L}(V, W)$. Define a function $f^* : W^* \to V^*$ by the rule $f^*(\tau) = \tau \circ f$. Then $f^*(\tau)$ lies in V^* because $\tau \circ f$ is the composite of the two linear transformations $V \xrightarrow{f} W \xrightarrow{\tau} k$.

Proposition 4.4. *Using the notation from above, the following statements hold.*

1. *$f^* \in \mathcal{L}(W^*, V^*)$.*
2. *$[f^*]_{\gamma^*}^{\beta^*} = ([f]_\beta^\gamma)^t$.*

Proof. (1) $f^* \in \mathcal{L}(W^*, V^*)$ because

$$f^*(\tau_1 + \tau_2) = (\tau_1 + \tau_2) \circ f = \tau_1 \circ f + \tau_2 \circ f = f^*(\tau_1) + f^*(\tau_2)$$

and

$$f^*(a\tau) = (a\tau) \circ f = a(\tau \circ f) = af^*(\tau).$$

(2) Let $[f]_\beta^\gamma = (a_{ij})_{m \times n}$ and let $[f^*]_{\gamma^*}^{\beta^*} = (b_{ij})_{n \times m}$. Then

$$f^*(\psi_j) = \psi_j \circ f = \sum_{i=1}^n (\psi_j \circ f)(v_i) \cdot \phi_i,$$

by Corollary 4.3. This shows that $b_{ij} = (\psi_j \circ f)(v_i)$. But,

$$(\psi_j \circ f)(v_i) = \psi_j(f(v_i)) = \psi_j(\sum_{l=1}^{m} a_{li}w_l) = \sum_{l=1}^{m} a_{li}\psi_j(w_l) = a_{ji}.$$

Therefore $b_{ij} = a_{ji}$ and this proves (2). □

Let $U \xrightarrow{f} V \xrightarrow{g} W$ be a sequence of linear transformations of vector spaces. The linear transformations f, g induce corresponding linear transformations on the dual spaces. Namely, we have $W^* \xrightarrow{g^*} V^* \xrightarrow{f^*} U^*$. Then $(g \circ f)^* = f^* \circ g^*$ because for every $\psi \in W^*$, we have

$$(g \circ f)^*(\psi) = \psi \circ (g \circ f) = (\psi \circ g) \circ f = g^*(\psi) \circ f = f^* \circ g^*(\psi).$$

Proposition 4.5. *In the situation above, assume that* $\dim U = p$, $\dim V = n$, *and* $\dim W = m$. *Let* α, β, γ *be ordered bases of* U, V, W, *respectively, and let* $\alpha^*, \beta^*, \gamma^*$ *be the dual bases of* U^*, V^*, W^* *to* α, β, γ. *Let* $A = [g]_\beta^\gamma$, $B = [f]_\alpha^\beta$, *and* $C = [g \circ f]_\alpha^\gamma$.

Then $(AB)^t = B^tA^t$. *In fact,* $(AB)^t = B^tA^t$ *for any pair of matrices having compatible sizes.*

Proof. $AB = [g]_\beta^\gamma[f]_\alpha^\beta = [g \circ f]_\alpha^\gamma = C$. Therefore, Proposition 4.4(2) implies that

$$(AB)^t = C^t = ([g \circ f]_\alpha^\gamma)^t = [(g \circ f)^*]_{\gamma^*}^{\alpha^*} = [f^* \circ g^*]_{\gamma^*}^{\alpha^*} = [f^*]_{\beta^*}^{\alpha^*}[g^*]_{\gamma^*}^{\beta^*} = B^tA^t.$$

Now let A, B be any pair of matrices having compatible sizes so that A, B can be multiplied. We may assume that $A \in \mathcal{M}_{m\times n}(k)$ and $B \in \mathcal{M}_{n\times p}(k)$. Then $A = [g]_\beta^\gamma$ for some $g \in \mathcal{L}(V, W)$ and $B = [f]_\alpha^\beta$ for some $f \in \mathcal{L}(U, V)$. The rest of the proof follows from above. □

Proposition 4.5 furnishes another proof of Proposition 3.15(3). Propositions 4.4 and 4.5 give a more intrinsic meaning of the transpose of a matrix.

4.2 Exact sequences and more on the rank of a matrix

Lemma 4.6. *Let* Y *be a vector space and let* $W \subseteq Y$ *be a subspace. Then there exists a subspace* $X \subseteq Y$ *such that* $W \bigoplus X = Y$.

Proof. Let $\{v_1, \dots, v_l\}$ be a basis of W. Extend this basis to a basis

$$\{v_1, \dots, v_l, v_{l+1}, \dots, v_m\}$$

of Y. Let $X = \langle v_{l+1}, \dots, v_m \rangle$, the span of $\{v_{l+1}, \dots, v_m\}$. Then $Y = W + X$ and $W \cap X = (0)$, and so $Y = W \bigoplus X$ by Proposition 1.16 and Definition 1.17. □

Proposition 4.7 (Fundamental Mapping Property of Linear Transformations)**.** *Let V, Y, Z be vector spaces over k and let $\phi : V \to Z$, $g : V \to Y$ be linear transformations. Assume that* $\ker(g) \subseteq \ker(\phi)$.

1. *If g is surjective, then there exists a unique linear transformation $\psi : Y \to Z$ such that $\phi = \psi \circ g$.*

2. *In general, there always exists at least one linear transformation $\psi : Y \to Z$ such that $\phi = \psi \circ g$.*

Proof. (1) First assume that g is surjective. Define $\psi : Y \to Z$ as follows. Let $y \in Y$. Since g is surjective, we may choose $v \in V$ such that $g(v) = y$. Define $\psi(y)$ by $\psi(y) = \phi(v)$. We must show that ψ is well-defined, ψ is a linear transformation, $\phi = \psi \circ g$ and ψ is uniquely determined.

Suppose that $v' \in V$ and $g(v') = y$. Then

$$v - v' \in \ker(g) \subseteq \ker(\phi).$$

Thus $\phi(v) = \phi(v')$, and this shows that ψ is well-defined.

To show that $\psi : Y \to Z$ is a linear transformation, let $y, y' \in Y$ and $a \in k$. Let $v, v' \in V$ such that $g(v) = y$ and $g(v') = y'$. Then $g(v + v') = g(v) + g(v') = y + y'$ and $g(av) = ag(v) = ay$. Thus,

$$\psi(y + y') = \phi(v + v') = \phi(v) + \phi(v') = \psi(y) + \psi(y'),$$

and

$$\psi(ay) = \phi(av) = a\phi(v) = a\psi(y).$$

This shows that ψ is a linear transformation.

Let $v \in V$. Then $(\psi \circ g)(v) = \psi(g(v)) = \phi(v)$ from the definition of ϕ.

To show that the linear transformation $\psi : Y \to Z$ is unique, suppose that $\psi' : Y \to Z$ is a linear transformation such that $\phi = \psi' \circ g$. Let $y \in Y$ and choose $v \in V$ such that $g(v) = y$. Then

$$\psi(y) = \phi(v) = \psi'(g(v)) = \psi'(y)$$

for all $y \in Y$. Therefore, $\psi = \psi'$.

(2) Now we remove the assumption that g is surjective. Let $W = \text{im}(g)$. Then (1) implies that there exists a unique linear transformation $\psi' : W \to Z$ such that $\phi = \psi' \circ g$. Since $Y = W \bigoplus X$ for some subspace $X \subseteq Y$, there exists a linear transformation $\psi : Y \to Z$ such that $\psi|_W = \psi'$. (For example, we can take $\psi(X) = 0$.) It follows that $\phi = \psi \circ g$. □

Remark: Note that the statement and proof of (1) extends to the case of modules over a commutative ring R. The proof of (2) uses special properties of vector spaces, namely the existence of a basis.

Proposition 4.8. *If $U \xrightarrow{f} V \xrightarrow{g} W$ is an exact sequence of linear transformations of vector spaces, then $W^* \xrightarrow{g^*} V^* \xrightarrow{f^*} U^*$ is also an exact sequence of linear transformations.*

Proof. First we show that $\text{im}(g^*) \subseteq \ker(f^*)$. Let $\phi \in \text{im}(g^*)$. Then $g^*(\psi) = \phi$, for some $\psi \in W^*$. It follows that

$$f^*(\phi) = f^*(g^*(\psi)) = (g \circ f)^*(\psi) = \psi \circ (g \circ f) = 0,$$

because $\text{im}(f) \subseteq \ker(g)$ and so $g \circ f = 0$. Thus $\phi \in \ker(f^*)$, so $\text{im}(g^*) \subseteq \ker(f^*)$.

Now let $\phi \in \ker(f^*)$. Then $f^*(\phi) = 0$ and so $\phi \circ f = 0$. Thus $\ker(g) = \text{im}(f) \subseteq \ker(\phi)$. Proposition 4.7 implies that there exists $\psi : W \to k$ such that $\psi \circ g = \phi$. Thus $\psi \in W^*$ and $g^*(\psi) = \psi \circ g = \phi$. Thus $\phi \in \text{im}(g^*)$, and so $\ker(f^*) \subseteq \text{im}(g^*)$. Therefore, $\text{im}(g^*) = \ker(f^*)$. □

Corollary 4.9. *Let $f \in \mathcal{L}(V, W)$ and let $f^* \in \mathcal{L}(W^*, V^*)$ be the induced linear transformation on dual spaces.*

1. *If f is injective, then f^* is surjective.*

2. *If f is surjective, then f^* is injective.*

Proof. If f is injective, then $0 \to V \to W$ is exact at V and so $W^* \to V^* \to 0$ is an exact sequence. Thus f^* is surjective.

If f is surjective, then $V \to W \to 0$ is exact at W and so $0 \to W^* \to V^*$ is an exact sequence. Thus f^* is injective. □

Corollary 4.10. *Suppose that $W \subseteq V$ is a subspace and $\iota : W \to V$ is the inclusion map. Let $\iota^* : V^* \to W^*$ denote the induced linear transformation on dual spaces. Let $\tau \in V^*$. Then $\iota^*(\tau) = \tau|_W$.*

In particular, if $\rho : V^* \to W^*$ is the function given by restricting the domain to W, then $\rho = \iota^*$ and ρ is surjective.

Proof. Let $\tau \in V^*$. Then $\iota^*(\tau) = \tau \circ \iota = \tau|_W = \rho(\tau)$. Therefore $\rho = \iota^*$. The Corollary 4.9 implies that ι^* is surjective because ι is injective. Thus ρ is surjective. □

Proposition 4.11. *Let $f \in \mathcal{L}(V, W)$ and let $f^* \in \mathcal{L}(W^*, V^*)$ be the induced linear transformation on dual spaces. Then $\dim(\operatorname{im}(f)) = \dim(\operatorname{im}(f^*))$.*

Proof. The linear transformation $f : V \to W$ induces the linear transformations $V \xrightarrow{g} \operatorname{im}(f) \xrightarrow{\iota} W$ where ι is the inclusion map and g is the same as f except that the range of f has been restricted to $\operatorname{im}(f)$. Then $f = \iota \circ g$.

This induces the linear transformations $W^* \xrightarrow{\iota^*} \operatorname{im}(f)^* \xrightarrow{g^*} V^*$ where $f^* = g^* \circ \iota^*$. We have that g^* is injective because g is surjective, and ι^* is surjective because ι is injective. Since $f^* : W^* \to V^*$, this gives

$$\begin{aligned}
\dim(W^*) - \dim(\operatorname{im}(f^*)) &= \dim(\ker(f^*)) \\
&= \dim(\ker(\iota^*)), \text{ because } g^* \text{ is injective,} \\
&= \dim(W^*) - \dim(\operatorname{im}(\iota^*)) \\
&= \dim(W^*) - \dim(\operatorname{im}(f)^*), \text{ because } \iota^* \text{ is surjective.}
\end{aligned}$$

Therefore, $\dim(\operatorname{im}(f^*)) = \dim(\operatorname{im}(f)^*) = \dim(\operatorname{im}(f))$. □

Lemma 4.12. *Let $f : V \to W$ be a linear transformation. Let β and γ be ordered bases of V and W, respectively, and let $A = [f]_\beta^\gamma$. Then $\dim(\operatorname{im}(f)) = \dim(\mathcal{C}(A))$.*

Proof. Proposition 3.12 and Exercise 16 of Chapter 3 imply that

$$\phi_\gamma(\operatorname{im}(f)) = \operatorname{im}(L_A) = \mathcal{C}(A).$$

Thus $\dim(\operatorname{im}(f)) = \dim(\phi_\gamma(\operatorname{im}(f))) = \dim(\mathcal{C}(A))$. □

We now can give a new proof of Proposition 3.35.

Theorem 4.13. *Let $A \in \mathcal{M}_{m\times n}(k)$. Then* $\dim(\mathcal{C}(A)) = \dim(\mathcal{R}(A))$.

Proof. Let $L_A : k^{(n)} \to k^{(m)}$ be the linear transformation given by $L_A(b) = Ab$. Let ϵ_n and ϵ_m be the standard bases of $k^{(n)}$ and $k^{(m)}$. Then $[L_A]_{\epsilon_n}^{\epsilon_m} = A$, by Lemma 3.11.

Let ϵ_n^* and ϵ_m^* be the dual bases of $(k^{(n)})^*$ and $(k^{(m)})^*$ to ϵ_n and ϵ_m. We have an induced linear transformation

$$(L_A)^* : (k^{(m)})^* \to (k^{(n)})^*.$$

Proposition 4.4 (2) implies that

$$[(L_A)^*]_{\epsilon_m^*}^{\epsilon_n^*} = ([L_A]_{\epsilon_n}^{\epsilon_m})^t = A^t.$$

It follows from Lemma 4.12 and Proposition 4.11 that

$$\dim(\mathcal{C}(A)) = \dim(\operatorname{im}(L_A)) = \dim(\operatorname{im}((L_A)^*)) = \dim(\mathcal{C}(A^t)) = \dim(\mathcal{R}(A)).$$

□

Let $f : V \to W$ be a linear transformation, with $\dim V = n$ and $\dim W = m$. Let $f^* : W^* \to V^*$ be the induced linear transformation of dual spaces where $f^*(\tau) = \tau \circ f$, for $\tau \in W^*$. We now construct bases of the subspaces $\ker(f)$, $\operatorname{im}(f)$, $\ker(f^*)$, and $\operatorname{im}(f^*)$.

Let $r = \dim(\operatorname{im}(f))$. Then $n - r = \dim(\ker(f))$. Let $\{v_{r+1}, \dots, v_n\}$ be a basis of $\ker(f)$. Extend this basis to a basis $\beta = \{v_1, \dots, v_r, v_{r+1}, \dots, v_n\}$ of V. Let $w_i = f(v_i)$, $1 \le i \le r$.

We show now that $\{w_1, \dots, w_r\}$ is a basis of $\operatorname{im}(f)$. First, $\{w_1, \dots, w_r\}$ spans $\operatorname{im}(f)$ because

$$\operatorname{im}(f) = \langle f(v_1), \dots, f(v_n)\rangle = \langle f(v_1), \dots, f(v_r)\rangle = \langle w_1, \dots, w_r\rangle.$$

Since $\dim(\operatorname{im}(f)) = r$, it follows that $\{w_1, \dots, w_r\}$ is a basis of $\operatorname{im}(f)$.

Extend $\{w_1, \dots, w_r\}$ to a basis $\gamma = \{w_1, \dots, w_r, \dots, w_m\}$ of W. Let $\gamma^* = \{\psi_1, \dots, \psi_m\}$ be the dual basis of W^* to γ. We will now show that $\{\psi_{r+1}, \dots, \psi_m\}$ is a basis of $\ker(f^*)$. If $r + 1 \le j \le m$ then $\psi_j \in \ker(f^*)$ because $f^*(\psi_j) = \psi_j \circ f$ and

$$\psi_j \circ f(V) = \psi_j(\operatorname{im}(f)) = \psi_j(\langle w_1, \dots, w_r\rangle) = 0.$$

Now let $\psi = \sum_{i=1}^m c_i\psi_i \in \ker(f^*)$. Then $0 = f^*(\sum_{i=1}^m c_i\psi_i) = (\sum_{i=1}^m c_i\psi_i) \circ f$. Therefore, if $1 \le j \le r$, then

$$0 = \left(\left(\sum_{i=1}^m c_i\psi_i\right) \circ f\right)(v_j) = \left(\sum_{i=1}^m c_i\psi_i\right)(f(v_j)) = \left(\sum_{i=1}^m c_i\psi_i\right)(w_j) = c_j.$$

Therefore $\psi = \sum_{i=r+1}^m c_i\psi_i$. This shows that $\{\psi_{r+1}, \ldots, \psi_m\}$ spans $\ker(f^*)$. Since $\{\psi_{r+1}, \ldots, \psi_m\}$ is part of a linearly independent set, it follows that $\{\psi_{r+1}, \ldots, \psi_m\}$ is a basis of $\ker(f^*)$.

Let $\beta^* = \{\phi_1, \ldots, \phi_n\}$ be the dual basis of V^* to β. We will show that $\{\phi_1, \ldots, \phi_r\}$ is a basis of $\mathrm{im}(f^*)$.

Since $\{\psi_1, \ldots, \psi_m\}$ is a basis of W^* and since $\{\psi_{r+1}, \ldots, \psi_m\} \subseteq \ker(f^*)$, it follows that

$$\mathrm{im}(f^*) = \langle f^*(\psi_1), \ldots, f^*(\psi_m)\rangle = \langle f^*(\psi_1), \ldots, f^*(\psi_r)\rangle = \langle \psi_1 \circ f, \ldots, \psi_r \circ f\rangle.$$

We will show that $f^*(\psi_i) = \psi_i \circ f = \phi_i$ for $1 \le i \le r$. This will imply that $\mathrm{im}(f^*) = \langle \phi_1, \ldots, \phi_r\rangle$. Since the set $\{\phi_1, \ldots, \phi_r\}$ is a linearly independent set, being a subset of β^*, it will follow that $\{\phi_1, \ldots, \phi_r\}$ is a basis of $\mathrm{im}(f^*)$.

Assume that $1 \le i \le r$. We have

$$\begin{aligned}(\psi_i \circ f)(v_j) = \psi_i(f(v_j)) &= \begin{cases}\psi_i(w_j), & \text{if } 1 \le j \le r\\ \psi_i(0), & \text{if } r+1 \le j \le n\end{cases} = \begin{cases}0, & \text{if } j \ne i\\ 1, & \text{if } j = i\end{cases}\\ &= \phi_i(v_j).\end{aligned}$$

Therefore, $f^*(\psi_i) = \psi_i \circ f = \phi_i$ for $1 \le i \le r$, because the two linear transformations agree on a basis of V.

4.3 The double dual of a vector space

$\mathcal{L}(V^*, k)$ is the dual space of V^*. It is often written V^{**} and called the double dual of V. Each element $v \in V$ determines an element $f_v \in V^{**}$ as follows. Let $f_v : V^* \to k$ be the function defined by $f_v(\phi) = \phi(v)$ for any $\phi \in V^*$. The function f_v is a linear transformation because for ϕ, $\tau \in V^*$ and $a \in k$, we have

$f_v(\phi + \tau) = (\phi + \tau)(v) = \phi(v) + \tau(v) = f_v(\phi) + f_v(\tau)$ and

$f_v(a\phi) = (a\phi)(v) = a(\phi(v)) = af_v(\phi)$.

Thus, $f_v \in \mathcal{L}(V^*, k)$. This gives us a function $f : V \to V^{**}$ where $f(v) = f_v$.

Proposition 4.14. *The function $f : V \to V^{**}$ is an injective linear transformation. If* $\dim V$ *is finite, then f is an isomorphism.*

Proof. First we show that f is a linear transformation. Let $v, w \in V$, let $a \in k$, and let $\phi \in V^*$. Since $f_{v+w}(\phi) = \phi(v+w) = \phi(v)+\phi(w) = f_v(\phi)+f_w(\phi)$, it follows that $f_{v+w} = f_v + f_w$ in V^{**} Thus,

$$f(v+w) = f_{v+w} = f_v + f_w = f(v) + f(w).$$

Since $f_{av}(\phi) = \phi(av) = a\phi(v) = af_v(\phi)$, it follows that $f_{av} = af_v$ in V^{**}. Thus,

$$f(av) = f_{av} = af_v = af(v).$$

This shows that f is a linear transformation.

Suppose that $v \in \ker f$. Then $f_v = f(v) = 0$. That means $f_v(\phi) = \phi(v) = 0$ for all $\phi \in V^*$. If $v \neq 0$, then V has a basis that contains v. Then there would exist an element $\phi \in V^*$ such that $\phi(v) \neq 0$, because a linear transformation can be defined by its action on basis elements, and this is a contradiction. Therefore $v = 0$ and f is injective.

If $\dim V = n$, then $n = \dim V = \dim V^* = \dim V^{**}$. Therefore f is also surjective and so f is an isomorphism. □

4.4 Annihilators of subspaces

Definition 4.15. *Let S be a subset of a vector space V. The* annihilator *S° of S is the set $S^\circ = \{\tau \in V^* \mid \tau(v) = 0$ for all $v \in S\}$.*

Thus $S^\circ = \{\tau \in V^* \mid S \subseteq \ker(\tau)\}$.

Lemma 4.16. *Let S be a subset of V and let $W = \langle S \rangle$.*

1. *S° is a subspace of V^*.*

2. *$S^\circ = W^\circ$.*

Proof. If τ_1, $\tau_2 \in S^\circ$ and $a \in k$, then it is easy to see that $\tau_1 + \tau_2 \in S^\circ$ and $a\tau_1 \in S^\circ$. Therefore (1) holds.

Since $S \subseteq W$, it follows easily that $W^\circ \subseteq S^\circ$. Now let $\tau \in S^\circ$. Then $\tau(S) = 0$ implies $S \subseteq \ker(\tau)$. Since $\ker(\tau)$ is a subspace, it follows that $W = \langle S \rangle \subseteq \ker(\tau)$. Therefore, $\tau \in W^\circ$ and (2) holds. □

Proposition 4.17. *Let W be a subspace of V. Consider the function $\rho : V^* \to W^*$ given by restricting the domain to W. That is, $\rho(\tau) = \tau|_W$.*

1. $\ker(\rho) = W^\circ$.

2. $\dim(W^\circ) = \dim(V) - \dim(W)$.

Proof. We know from Corollary 4.10 that ρ is a surjective linear transformation. Then $\ker(\rho) = \{\tau \in V^* \mid \tau|_W = 0\} = W^\circ$. Thus

$$\begin{aligned}\dim(W^\circ) &= \dim(\ker(\rho)) = \dim V^* - \dim(\mathrm{im}(\rho))\\ &= \dim V^* - \dim W^* = \dim V - \dim W.\end{aligned}$$

□

Here is a further look at the situation in Proposition 4.17. Let W be a subspace of V. The following sequence is exact.

$$0 \to W \xrightarrow{\iota} V \xrightarrow{\pi} V/W \to 0$$

Proposition 4.8 implies that the following sequence is exact.

$$0 \to (V/W)^* \xrightarrow{\pi^*} V^* \xrightarrow{\iota^*} W^* \to 0$$

Since $\iota^* = \rho$ (see Corollary 4.10), the exactness implies

$$\pi^*((V/W)^*) = \mathrm{im}(\pi^*) = \ker(\iota^*) = \ker(\rho) = W^\circ.$$

Thus $\pi^* : (V/W)^* \to W^\circ$ is an isomorphism and so

$$\dim(W^\circ) = \dim((V/W)^*) = \dim(V/W) = \dim(V) - \dim(W).$$

Since $\ker(\iota^*) = W^\circ$, we have $V^*/W^\circ \cong W^*$.

See Exercise 7 for additional proofs of some of these results.

Proposition 4.18. *Let $f : V \to W$ be a linear transformation and let $f^* : W^* \to V^*$ be the corresponding linear transformation of dual spaces. Then*

1. $\ker(f^*) = (\mathrm{im}(f))^\circ$

2. $(\ker(f))^\circ = \mathrm{im}(f^*)$

Proof. (1) $h \in \ker(f^*) \iff h \circ f = 0 \iff h(\mathrm{im}(f)) = 0 \iff h \in (\mathrm{im}(f))^\circ$.

(2) Let $g \in \mathrm{im}(f^*)$. Then $g = f^*(h) = h \circ f$, for some $h \in W^*$. Since $g(\ker(f)) = (h \circ f)(\ker(f)) = h(f(\ker(f))) = 0$, it follows that $g \in (\ker(f))^\circ$. Thus, $\mathrm{im}(f^*) \subseteq (\ker(f))^\circ$. Proposition 4.11 and Proposition 4.17 (2) imply that

$$\dim(\mathrm{im}(f^*)) = \dim(\mathrm{im}(f)) = \dim V - \dim(\ker(f)) = \dim(\ker(f))^\circ.$$

It follows that $(\ker(f))^\circ = \mathrm{im}(f^*)$. □

Here is a direct proof that $(\ker(f))^\circ \subseteq \mathrm{im}(f^*)$ in Proposition 4.18 (2). Let $\phi \in (\ker(f))^\circ$. Then $\phi \in V^*$ and $\phi(\ker(f)) = 0$. Since $\ker(f) \subseteq \ker(\phi)$, Proposition 4.7 (2) implies that there exists $\psi \in W^*$ such that $f^*(\psi) = \psi \circ f = \phi$. Thus $\phi \in \mathrm{im}(f^*)$, so $(\ker(f))^\circ \subseteq \mathrm{im}(f^*)$.

4.5 Subspaces of V and V^*

The following definition includes Definition 4.15.

Definition 4.19. *Let V be a vector space over k of dimension n. Let W be a subspace of V and let Y be a subspace of V^*.*

1. *Let $W^\circ = \{f \in V^* \mid W \subseteq \ker(f)\}$.*
2. *Let $Y_\circ = \bigcap_{f \in Y} \ker(f)$.*

Thus, $W^\circ \subseteq V^*$, $Y_\circ \subseteq V$, and $Y_\circ = \{v \in V \mid f(v) = 0, \text{ for all } f \in Y\}$. From Proposition 4.17 (2), we have $\dim(W^\circ) = \dim(V) - \dim(W)$.

It is easy to see that if $W_1 \subseteq W_2$, then $W_1^\circ \supseteq W_2^\circ$, and if $Y_1 \subseteq Y_2$, then $(Y_1)_\circ \supseteq (Y_2)_\circ$.

Proposition 4.20. *$W \subseteq (W^\circ)_\circ$ and $Y \subseteq (Y_\circ)^\circ$.*

Proof. If $x \in W$, then $f(x) = 0$ for all $f \in W^\circ$. Thus,

$$x \in \bigcap_{f \in W^\circ} \ker(f) = (W^\circ)_\circ.$$

Therefore, $W \subseteq (W^\circ)_\circ$.

If $f \in Y$, then $f(v) = 0$ for all $v \in Y_\circ$. Thus $f(Y_\circ) = 0$ and this implies that $f \in (Y_\circ)^\circ$. Therefore, $Y \subseteq (Y_\circ)^\circ$. □

Lemma 4.21. *If Y is spanned by $\{f_1, \dots, f_r\}$, then $Y_\circ = \bigcap_{i=1}^r \ker(f_i)$.*

Proof. We have $Y_\circ = \bigcap_{f \in Y} \ker(f) \subseteq \bigcap_{i=1}^r \ker(f_i)$, because $f_i \in Y$. Now let $x \in \bigcap_{i=1}^r \ker(f_i)$. For each $f \in Y$, we have $f = a_1 f_1 + \cdots + a_r f_r$ with $a_i \in k$. It follows that $f(x) = \sum_{i=1}^r a_i f_i(x) = 0$. Then $x \in \ker(f)$ and it follows that $x \in \bigcap_{f \in Y} \ker(f) = Y_\circ$. Therefore, $Y_\circ = \bigcap_{i=1}^r \ker(f_i)$. □

Proposition 4.22. $\dim(Y_\circ) = \dim(V^*) - \dim(Y)$.

Proof. We have $\dim(Y) \le \dim((Y_\circ)^\circ) = \dim(V) - \dim(Y_\circ)$. Thus, $\dim(Y_\circ) \le \dim(V) - \dim(Y) = \dim(V^*) - \dim(Y)$.

Suppose that $\dim(Y) = r$ and let $\{f_1, \dots, f_r\}$ be a basis of Y. Then Lemma 4.21 implies that $Y_\circ = \bigcap_{i=1}^r \ker(f_i)$. Since $f_i \neq 0$, it follows that $f_i : V \to k$ is surjective and so $\dim(\ker(f_i)) = n-1$. Thus $\operatorname{codim}(\ker(f_i)) = 1$ and $\operatorname{codim}(\bigcap_{i=1}^r \ker(f_i)) \le \sum_{i=1}^r \operatorname{codim}(\ker(f_i)) = r$. (See Exercise 4 in Chapter 2.) Then

$$\begin{aligned}\dim(Y_\circ) &= \dim\left(\bigcap_{i=1}^r \ker(f_i)\right) \\ &\ge \dim(V) - r = \dim(V) - \dim(Y) = \dim(V^*) - \dim(Y).\end{aligned}$$

Therefore, $\dim(Y_\circ) = \dim(V^*) - \dim(Y)$. □

Proposition 4.23. *$W = (W^\circ)_\circ$ and $Y = (Y_\circ)^\circ$.*

Proof. $\dim((W^\circ)_\circ) = \dim(V^*) - \dim(W^\circ) = \dim(V) - (\dim(V) - \dim(W)) = \dim(W)$. Since $W \subseteq (W^\circ)_\circ$, it follows that $W = (W^\circ)_\circ$.

$\dim((Y_\circ)^\circ) = \dim(V) - \dim(Y_\circ) = \dim(V) - (\dim(V^*) - \dim(Y)) = \dim(Y)$. Since $Y \subseteq (Y_\circ)^\circ$, it follows that $Y = (Y_\circ)^\circ$. □

Corollary 4.24. *Let $W_1, W_2 \subseteq V$ and $Y_1, Y_2 \subseteq V^*$. If $(W_1)^\circ = (W_2)^\circ$, then $W_1 = W_2$. If $(Y_1)_\circ = (Y_2)_\circ$, then $Y_1 = Y_2$.*

Proof. If $(W_1)^\circ = (W_2)^\circ$, then $W_1 = ((W_1)^\circ)_\circ = ((W_2)^\circ)_\circ = W_2$. If $(Y_1)_\circ = (Y_2)_\circ$, then $Y_1 = ((Y_1)_\circ)^\circ = ((Y_2)_\circ)^\circ = Y_2$. □

The results above imply the following Theorem.

Theorem 4.25. *Let $\Delta : \{$Subspaces of $V\} \to \{$Subspaces of $V^*\}$, where $W \to W^\circ$.*

Let $\Omega : \{$Subspaces of $V^*\} \to \{$Subspaces of $V\}$, where $Y \to Y_\circ$.

Then Δ and Ω are inclusion-reversing bijections, and Δ and Ω are inverse functions of each other.

4.6 A calculation

The next result gives a computation in V^* that doesn't involve a dual basis of V^*.

Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V and let $\gamma = \{\tau_1, \ldots, \tau_n\}$ be a basis of V^* (not necessarily the dual basis of V^* to $\{v_1, \ldots, v_n\}$). Let $\tau_i(v_j) = b_{ij}$, for $1 \leq i \leq n$ and $1 \leq j \leq n$.

Let $v \in V$ and $\tau \in V^*$. We wish to compute $\tau(v)$. Let $\tau = \sum_{i=1}^n c_i \tau_i$ and let $v = \sum_{j=1}^n d_j v_j$. Let

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \ d = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}, \ B = (b_{ij})_{n \times n}.$$

Then

$$\begin{aligned} \tau(v) &= \left(\sum_{i=1}^n c_i \tau_i\right)\left(\sum_{j=1}^n d_j v_j\right) = \sum_{i=1}^n \sum_{j=1}^n c_i d_j \tau_i(v_j) = \sum_{i=1}^n \sum_{j=1}^n c_i b_{ij} d_j \\ &= \sum_{i=1}^n c_i \left(\sum_{j=1}^n b_{ij} d_j\right) = c^t B d. \end{aligned}$$

The last equality holds because the i^{th} coordinate of Bd equals $\sum_{j=1}^n b_{ij} d_j$.

If $\{\tau_1, \ldots, \tau_n\}$ is the dual basis of V^* to $\{v_1, \ldots, v_n\}$, then $B = I_n$ because we would have

$$b_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases}$$

In this case, the formula reads $\tau(v) = c^t d$.

Exercises

1. Suppose that $f(v) = 0$ for all $f \in V^*$. Then $v = 0$.

2. Suppose that V is infinite dimensional over k. Let $\beta = \{v_\alpha\}_{\alpha \in J}$ be a basis of V where the index set J is infinite. For each $\alpha \in J$, define a linear transformation $\phi_\alpha : V \to k$ by the rule $\phi_\alpha(v_\lambda) = 0$ if $\alpha \neq \lambda$ and $\phi_\alpha(v_\lambda) = 1$ if $\alpha = \lambda$. Show that $\{\phi_\alpha\}_{\alpha \in J}$ is a linearly independent set in V^* but is not a basis of V^*.

3. In Proposition 4.14, show that f is not surjective if V is infinite dimensional over k. (As in the previous exercise assume that V has a basis, even though the proof of this fact in this case was not given in Chapter 1.)

4. Let $\{v_1, \ldots, v_n\}$ be a basis of V and let $\{\phi_1, \ldots, \phi_n\}$ be the dual basis of V^* to $\{v_1, \ldots, v_n\}$. In the notation of Proposition 4.14, show that $\{f_{v_1}, \ldots, f_{v_n}\}$ is the dual basis of V^{**} to $\{\phi_1, \ldots, \phi_n\}$. Give a second proof that the linear transformation f in Proposition 4.14 is an isomorphism, in the case that $\dim V$ is finite, by observing that f takes a basis of V to a basis of V^{**}.

5. Let V, W be finite dimensional vector spaces. Let $f : V \to W$ be a linear transformation. Let $f^* : W^* \to V^*$ and $f^{**} : V^{**} \to W^{**}$ be the induced linear transformations on the dual spaces and double dual spaces. Let $h : V \to V^{**}$ and $h' : W \to W^{**}$ be the isomorphisms from Proposition 4.14. Show that $f^{**} \circ h = h' \circ f$.

6. Let $U \xrightarrow{f} V \xrightarrow{g} W$ be a sequence of linear transformations of finite dimensional vector spaces. Suppose that $W^* \xrightarrow{g^*} V^* \xrightarrow{f^*} U^*$ is an exact sequence of linear transformations. Then prove that the original sequence is exact. (Hint: One method is to apply Proposition 4.8 to the given exact sequence and use Proposition 4.14 along with the previous exercise. One can also give a direct argument.)

7. Give a direct proof that ρ is surjective in Proposition 4.10 and give a direct proof of the dimension formula in Proposition 4.17 (2) as follows. Let $\{v_1, \ldots, v_r\}$ be a basis of W. Extend this basis to a basis $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ of V. Let $\{\phi_1, \ldots, \phi_n\}$ be the dual basis of V^* to $\{v_1, \ldots, v_n\}$.

   (a) Show that $\{\phi_1|_W, \ldots, \phi_r|_W\}$ is the dual basis of W^* to $\{v_1, \ldots, v_r\}$. Let $\tau \in W^*$. Show that $\tau = \sum_{i=1}^r a_i\phi_i|_W = \rho(\sum_{i=1}^r a_i\phi_i)$. Conclude that $\rho : V^* \to W^*$ is surjective.

   (b) Show that $\phi_{r+1}, \ldots, \phi_n \in W^\circ$. If $\tau = \sum_{i=1}^n a_i\phi_i \in W^\circ$, then show $a_j = 0$, $1 \leq j \leq r$. Conclude that $\{\phi_{r+1}, \ldots, \phi_n\}$ spans W°. Since $\{\phi_{r+1}, \ldots, \phi_n\}$ is a linearly independent set, conclude that $\{\phi_{r+1}, \ldots, \phi_n\}$ is a basis of W°. Therefore $\dim(W^\circ) = \dim(V) - \dim(W)$.

Chapter 5

Inner Product Spaces

5.1 The Basics

In this chapter, k denotes either the field of real numbers $\mathbf{R}$ or the field of complex numbers $\mathbf{C}$.

If $\alpha \in \mathbf{C}$, let $\overline{\alpha}$ denote the complex conjugate of α. Thus if $\alpha = a + bi$, where $a, b \in \mathbf{R}$ and $i = \sqrt{-1}$, then $\overline{\alpha} = a - bi$. Recall the following facts about complex numbers and complex conjugation. If $\alpha, \beta \in \mathbf{C}$, then $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$, $\overline{\alpha\beta} = \overline{\alpha}\overline{\beta}$, and $\overline{\overline{\alpha}} = \alpha$. A complex number α is real if and only if $\alpha = \overline{\alpha}$. The absolute value of $\alpha \in \mathbf{C}$ is defined to be the nonnegative square root of $\alpha\overline{\alpha}$. This is well defined because $\alpha\overline{\alpha} = a^2 + b^2$ is a nonnegative real number. We see that this definition of the absolute value of a complex number extends the usual definition of the absolute value of a real number by considering the case $b = 0$.

Definition 5.1. *Let V be a vector space over k. An* inner product *on V is a function $\langle\ \rangle : V \times V \to k$, where $(v, w) \mapsto \langle v, w\rangle$, satisfying the following four statements. Let $v_1, v_2, v, w \in V$ and $c \in k$.*

1. $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$.

2. $\langle cv, w\rangle = c\langle v, w\rangle$.

3. $\langle w, v\rangle = \overline{\langle v, w\rangle}$.

4. $\langle v, v\rangle > 0$ *for all* $v \neq 0$.

Statement 3 in Definition 5.1 implies that $\langle v, v\rangle$ is a real number. Statements 1 and 2 in the definition imply that for any fixed $w \in V$, the function $\langle\ , w\rangle : V \to k$, given by $v \mapsto \langle v, w\rangle$, is a linear transformation. It follows that $\langle 0, w\rangle = 0$, for all $w \in V$.

Proposition 5.2. *Let $\langle\ \rangle : V \times V \to k$ be an inner product on k. Let $v, w, w_1, w_2 \in V$ and $c \in k$. Then*

5. $\langle v, w_1 + w_2\rangle = \langle v, w_1\rangle + \langle v, w_2\rangle$.

6. $\langle v, cw\rangle = \overline{c}\langle v, w\rangle$.

Proof. $\langle v, w_1 + w_2\rangle = \overline{\langle w_1 + w_2, v\rangle} = \overline{\langle w_1, v\rangle + \langle w_2, v\rangle} = \overline{\langle w_1, v\rangle} + \overline{\langle w_2, v\rangle} = \langle v, w_1\rangle + \langle v, w_2\rangle$.

$\langle v, cw\rangle = \overline{\langle cw, v\rangle} = \overline{c\langle w, v\rangle} = \overline{c}\overline{\langle w, v\rangle} = \overline{c}\langle v, w\rangle$. □

If $k = \mathbf{R}$, then $\langle w, v\rangle = \langle v, w\rangle$ and $\langle v, cw\rangle = c\langle v, w\rangle$. If $k = \mathbf{C}$, then it is necessary to introduce complex conjugation in Definition 5.1 in order statements (1)-(4) to be consistent. For example, suppose $\langle v, cw\rangle = c\langle v, w\rangle$ for all $c \in \mathbf{C}$. Then $-\langle v, v\rangle = \langle iv, iv\rangle > 0$ when $v \neq 0$. But $\langle v, v\rangle > 0$ by (4) when $v \neq 0$.

Example. Let $V = k^{(n)}$, where k is $\mathbf{R}$ or $\mathbf{C}$, and let $v, w \in k^{(n)}$. Consider v, w to be column vectors in $k^{(n)}$ (or $n \times 1$ matrices). Define an inner product $\langle\rangle : V \times V \to k$ by $(v, w) \mapsto v^t\overline{w}$. If v is the column vector given by $(a_1, \ldots, a_n)$ and w is the column vector given by $(b_1, \ldots, b_n)$, $a_i, b_i \in k$, then $\langle v, w\rangle = \sum_{i=1}^n a_i\overline{b_i}$. It is easy to check that the first two statements in Definition 5.1 are satisfied. The third statement follows from $\overline{\sum_{i=1}^n a_i\overline{b_i}} = \sum_{i=1}^n b_i\overline{a_i}$. The fourth statement follows from $\sum_{i=1}^n a_i\overline{a_i} = \sum_{i=1}^n |a_i|^2 > 0$ as long as some a_i is nonzero. This inner product is called the *standard inner product* on $k^{(n)}$.

Definition 5.3. *Let $\langle\ \rangle$ be an inner product on V and let $v \in V$. The* norm *of v, written $||v||$, is defined to be the nonnegative square root of $\langle v, v\rangle$.*

The norm on V extends the absolute value function on k in the following sense. Suppose $V = k^{(1)}$ and let $v \in V$. Then $v = (a)$, $a \in k$. Assume that $\langle\ \rangle$ is the standard inner product on V. Then $||v||^2 = \langle v, v\rangle = a\overline{a} = |a|^2$. Thus $||v|| = |a|$.

Proposition 5.4. *Let $\langle\ \rangle$ be an inner product on V and let $v, w \in V$, $c \in k$. Then the following statements hold.*

1. $||cv|| = |c|\,||v||$

2. $||v|| > 0$, *if* $v \neq 0$.

3. $|\langle v, w\rangle| \leq ||v||\,||w||$ *(Cauchy-Schwarz inequality)*

4. $||v + w|| \leq ||v|| + ||w||$ *(triangle inequality)*

Proof. 1. $||cv||^2 = \langle cv, cv\rangle = c\bar{c}\langle v, v\rangle = |c|^2||v||^2$.

2. If $v \neq 0$, then $||v||^2 = \langle v, v\rangle > 0$.

3. Let $y = ||w||^2v - \langle v, w\rangle w$. Then

$$\begin{aligned} 0 \leq \langle y, y\rangle &= \langle ||w||^2v - \langle v, w\rangle w, ||w||^2v - \langle v, w\rangle w\rangle \\ &= ||w||^4\langle v, v\rangle - \langle v, w\rangle||w||^2\langle w, v\rangle - ||w||^2\overline{\langle v, w\rangle}\langle v, w\rangle + \langle v, w\rangle\overline{\langle v, w\rangle}\langle w, w\rangle \\ &= ||w||^2(||v||^2||w||^2 - |\langle v, w\rangle|^2 - |\langle v, w\rangle|^2 + |\langle v, w\rangle|^2) \\ &= ||w||^2(||v||^2||w||^2 - |\langle v, w\rangle|^2). \end{aligned}$$

If $w = 0$, then 3. is clear. If $w \neq 0$, then this calculation shows that $0 \leq ||v||^2||w||^2 - |\langle v, w\rangle|^2$ and this gives 3.

4. First we note that 3 implies that

$$\langle v, w\rangle + \langle w, v\rangle = \langle v, w\rangle + \overline{\langle v, w\rangle} = 2\mathrm{Re}(\langle v, w\rangle) \leq 2|\langle v, w\rangle| \leq 2||v||\,||w||.$$

Therefore, 4 follows from

$$\begin{aligned} ||v + w||^2 = \langle v + w, v + w\rangle &= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\ &\leq ||v||^2 + 2||v||\,||w|| + ||w||^2 = (||v|| + ||w||)^2. \end{aligned}$$

□

Proposition 5.5. *The inner product* $\langle\ \ \rangle$ *is completely determined by* $||\ \ ||^2$. *The following formulas hold.*

1. If $k = \mathbf{R}$, *then* $\langle v, w\rangle = \frac{1}{4}||v + w||^2 - \frac{1}{4}||v - w||^2$.

2. If $k = \mathbf{C}$, *then* $\langle v, w\rangle = \frac{1}{4}||v+w||^2 - \frac{1}{4}||v-w||^2 + \frac{i}{4}||v+iw||^2 - \frac{i}{4}||v-iw||^2$.

Proof. See Exercise 2. □

5.2 Orthogonality

In this section, let $\langle\ \ \rangle$ denote a fixed inner product on V.

Definition 5.6. *Let $\langle\ \ \rangle$ be an inner product on V.*

1. *Vectors $v, w \in V$ are* orthogonal *if $\langle v, w\rangle = 0$.*

2. *A subset S of vectors in V is an* orthogonal set *if each pair of distinct vectors in S is orthogonal.*

3. *An* orthonormal set *S in V is an orthogonal set with the additional property that $||v|| = 1$ for each $v \in S$.*

Note that the relation of orthogonality is symmetric since $\langle v, w\rangle = 0$ if and only if $\langle w, v\rangle = 0$.

Proposition 5.7. *An orthogonal set of nonzero vectors in V is linearly independent.*

Proof. Let S be an orthogonal set in V and let $v_1, \ldots, v_n$ be distinct nonzero vectors in S. Suppose $c_1v_1 + \cdots + c_nv_n = 0$, $c_i \in k$. Then for each j, $0 = \langle 0, v_j\rangle = \langle \sum_{i=1}^n c_iv_i, v_j\rangle = \sum_{i=1}^n c_i\langle v_i, v_j\rangle = c_j\langle v_j, v_j\rangle$. Therefore $c_j = 0$ since $\langle v_j, v_j\rangle \neq 0$. Thus $v_1, \ldots, v_n$ is a linearly independent set and so S is a linearly independent set. □

Definition 5.8. *Let $S \subseteq V$ be a subset in V. The* orthogonal complement *of S in V, written $S^\perp$, is $\{v \in V | \langle v, w\rangle = 0 \, for \ all \, w \in S\}$.*

Note that $S^\perp$ depends on the given inner product on V, although the notation does not indicate that.

Proposition 5.9. *1. $\{0\}^\perp = V$, $V^\perp = \{0\}$.*

2. *$S^\perp$ is a subspace of V.*

3. *If $S_1 \subseteq S_2$, then $S_2^\perp \subseteq S_1^\perp$.*

4. *Let $W = Span(S)$. Then $W^\perp = S^\perp$.*

Proof. If $v \in V^\perp$, then $\langle v, v\rangle = 0$ and so $v = 0$.

2 and 3 are clear.

4. Let $v \in S^\perp$. If $w \in W$, then $w = \sum_{i=1}^n c_is_i$, where $c_i \in k$ and $s_i \in S$. Then $\langle v, w\rangle = \langle v, \sum_{i=1}^n c_is_i\rangle = \sum_{i=1}^n \overline{c_i}\langle v, s_i\rangle = 0$, since $v \in S^\perp$. Thus, $v \in W^\perp$ and so $S^\perp \subseteq W^\perp$. The other inclusion follows easily from 3. □

Lemma 5.10. *Let $Y \subseteq V$ be a subspace.*

1. *If Y is spanned by $\{v_1, \dots, v_l\}$, then $Y^\perp = \bigcap_{i=1}^l \langle v_i \rangle^\perp$.*

2. *If $\dim V = n$ and $v \in V$, with $v \neq 0$, then $\dim(\langle v \rangle^\perp) = n - 1$.*

Proof. Since $\langle v_i \rangle \subseteq Y$, we have $Y^\perp \subseteq \langle v_i \rangle^\perp$ and it follows $Y^\perp \subseteq \bigcap_{i=1}^l \langle v_i \rangle^\perp$. Let $w \in \bigcap_{i=1}^l \langle v_i \rangle^\perp$ and let $y \in Y$. Then $y = \sum_{i=1}^l a_i v_i$ and

$$\langle y, w \rangle = \sum_{i=1}^l a_i \langle v_i, w \rangle = 0.$$

Therefore, $w \in Y^\perp$ and we have $Y^\perp = \bigcap_{i=1}^l \langle v_i \rangle^\perp$.

Now suppose $\dim V = n$ and let $v \in V$, with $v \neq 0$. Then $\langle v \rangle^\perp$ is the kernel of the linear transformation $f : V \to k$ given by $f(w) = \langle w, v \rangle$. The linear transformation f is nonzero since $f(v) \neq 0$. Therefore, f is surjective (as k is a one-dimensional vector space over k) and Proposition 2.5 implies $\dim(\ker(f)) = \dim(V) - \dim(\operatorname{im}(f)) = n - 1$. □

Proposition 5.11. *Let $W \subseteq V$ be a subspace,* $\dim V = n$. *Then*

1. $W \cap W^\perp = (0)$,

2. $\dim(W^\perp) = \dim V - \dim W$,

3. $V = W \bigoplus W^\perp$,

4. $W^{\perp\perp} = W$, *where $W^{\perp\perp}$ means $(W^\perp)^\perp$.*

Proof. Let $v \in W \cap W^\perp$. Then $\langle v, v \rangle = 0$ and this implies $v = 0$. Therefore, $W \cap W^\perp = (0)$. This proves (1).

Let $\{v_1, \dots, v_l\}$ be a basis of W. Lemma 5.10 implies that $W^\perp = \bigcap_{i=1}^l \langle v_i \rangle^\perp$ and $\dim(\langle v_i \rangle)^\perp = n - 1$, $1 \leq i \leq l$. Then

$$\operatorname{codim}(\bigcap_{i=1}^l \langle v_i \rangle^\perp) \leq \sum_{i=1}^l \operatorname{codim}(\langle v_i \rangle^\perp) = l.$$

This implies

$$\dim(V) - \dim(\bigcap_{i=1}^l \langle v_i \rangle^\perp) \leq l$$

and so

$$\dim(\bigcap_{i=1}^{l} \langle v_i \rangle^{\perp}) \geq n - l.$$

Therefore, $\dim(W^{\perp}) \geq n - l$. Since $W \cap W^{\perp} = (0)$, we have

$$\begin{aligned} n = l + (n-l) &\leq \dim W + \dim W^{\perp} = \dim(W + W^{\perp}) + \dim(W \cap W^{\perp}) \\ &= \dim(W + W^{\perp}) \leq \dim V = n. \end{aligned}$$

Thus $\dim(W^{\perp}) = n - l$ and $\dim(W + W^{\perp}) = n$. This implies $W + W^{\perp} = V$. This proves (2) and (3).

It is easy to check that $W \subseteq W^{\perp\perp}$. We have equality since

$$\dim W^{\perp\perp} = \dim V - \dim W^{\perp} = \dim V - (\dim V - \dim W) = \dim W.$$

This proves (4). □

The next two results give a computational procedure to find the orthogonal complement of a subspace.

Proposition 5.12 (Gram-Schmidt Orthogonalization). *Let $\langle\ \ \rangle$ be an inner product on V and let $\{v_1, \ldots, v_n\}$ be a linearly independent set in V. Then there exists $w_1, \ldots, w_n \in V$ such that the following two statements hold.*

1. *$\{w_1, \ldots, w_n\}$ is an orthogonal set of nonzero vectors.*

2. *$Span\{v_1, \ldots, v_l\} = Span\{w_1, \ldots, w_l\}$, $1 \leq l \leq n$.*

In particular, if $\{v_1, \ldots, v_n\}$ is a basis of V, then $\{w_1, \ldots, w_n\}$ is an orthogonal basis of V. Therefore, every finite dimensional inner product space has an orthogonal basis.

Proof. The proof is by induction on n. If $n = 1$, let $w_1 = v_1$. Now assume by induction that the result has been proved for all values $m < n$. That means we can assume there exist $w_1, \ldots, w_{n-1} \in V$ such that $\{w_1, \ldots, w_{n-1}\}$ is an orthogonal set of nonzero vectors and $Span\{v_1, \ldots, v_l\} = Span\{w_1, \ldots, w_l\}$, $1 \leq l \leq n - 1$.

Let

$$w_n = v_n - \sum_{j=1}^{n-1} \frac{\langle v_n, w_j \rangle}{||w_j||^2} w_j.$$

If $w_n = 0$, then $v_n \in Span\{w_1, \ldots, w_{n-1}\} = Span\{v_1, \ldots, v_{n-1}\}$, which would contradict the linear independence of $\{v_1, \ldots, v_n\}$. Therefore, $w_n \neq 0$.

If $1 \leq i \leq n-1$, then

$$\begin{aligned} \langle w_n, w_i \rangle &= \langle v_n, w_i \rangle - \sum_{j=1}^{n-1} \frac{\langle v_n, w_j \rangle}{||w_j||^2} \langle w_j, w_i \rangle \\ &= \langle v_n, w_i \rangle - \frac{\langle v_n, w_i \rangle}{||w_i||^2} \langle w_i, w_i \rangle = 0. \end{aligned}$$

Therefore, $\{w_1, \ldots, w_n\}$ is an orthogonal set of nonzero vectors and it is easy to check that $Span\{v_1, \ldots, v_n\} = Span\{w_1, \ldots, w_n\}$. The remaining statements now follow easily. □

Corollary 5.13. *Assume* $\dim V = n$.

1. *If $\{v_1, \ldots, v_n\}$ is a basis of V, then an orthogonal basis $\{w_1, \ldots, w_n\}$ of V is given by the following formulas.*

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \frac{\langle v_2, w_1 \rangle}{||w_1||^2} w_1 \\ w_3 &= v_3 - \frac{\langle v_3, w_1 \rangle}{||w_1||^2} w_1 - \frac{\langle v_3, w_2 \rangle}{||w_2||^2} w_2 \\ &\vdots \\ w_i &= v_i - \sum_{j=1}^{i-1} \frac{\langle v_i, w_j \rangle}{||w_j||^2} w_j \\ &\vdots \\ w_n &= v_n - \sum_{j=1}^{n-1} \frac{\langle v_n, w_j \rangle}{||w_j||^2} w_j \end{aligned}$$

2. *Let $y_i = \frac{1}{||w_i||} w_i$. Then $\{y_1, \ldots, y_n\}$ is an orthonormal basis of V. In particular, every finite dimensional inner product space has an orthonormal basis.*

Proof. 1. The w_i's are well defined since each w_i is defined in terms of $v_i, w_1, \ldots, w_{i-1}$. The proof of the previous proposition shows $\{w_1, \ldots, w_n\}$ is an orthogonal basis of V.

2. Proposition 5.4(1) implies that $||y_i|| = 1$. □

5.3 The Matrix of an Inner Product on V

Definition 5.14. *Let $A \in \mathcal{M}_{m\times n}(k)$, $A = (a_{ij})_{m\times n}$.*

1. *The* adjoint *of A, written A^*, is defined to be $\overline{A^t}$. That is, $A^* \in \mathcal{M}_{n\times m}(k)$ and if $A^* = (b_{ij})_{n\times m}$, then $b_{ij} = \overline{a_{ji}}$.*

2. *If $A^* = A$, then A is called a* hermitian *matrix. Note that if A is hermitian, then $m = n$.*

Observe that $\overline{A^t} = \left(\overline{A}\right)^t$.

Proposition 5.15. *Let $A, B \in \mathcal{M}_{m\times n}(k)$ and let $C \in \mathcal{M}_{n\times p}(k)$. Let $c \in k$. Then*

1. *$(A+B)^* = A^* + B^*$ and $(cA)^* = \overline{c}A^*$,*

2. *$(AC)^* = C^*A^*$,*

3. *$A^{**} = A$, where A^{**} means $(A^*)^*$.*

Proof. (1) and (3) are easy using basic facts about complex conjugation and transposes of matrices. Here is a proof of (2).

$$(AC)^* = \overline{(AC)^t} = \overline{C^tA^t} = \overline{C^t}\,\overline{A^t} = C^*A^*.$$

□

Since we sometimes work with more than one inner product at a time, we modify our notation for an inner product function in the following way. We denote an inner product by a function $B : V \times V \to k$, where $B(x, y) = \langle x, y\rangle_B$. If there is no danger of confusion, we write instead $B(x, y) = \langle x, y\rangle$.

Definition 5.16. *Let $B : V \times V \to k$ be an inner product on V and let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of V. Let $[B]_\beta \in \mathcal{M}_{n\times n}(k)$ be the matrix whose (i, j)-entry is given by $\langle v_i, v_j\rangle_B$. Then $[B]_\beta$ is called the matrix of B with respect to the basis β.*

Proposition 5.17. *Let $B : V \times V \to k$ be an inner product on V and let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of V. Let $G = [B]_\beta$, the matrix of B with respect to the basis β.*

Let $x = \sum_{i=1}^n a_iv_i$ and $y = \sum_{j=1}^n b_jv_j$. Then the following statements hold.

1. $\langle x, y\rangle = [x]^t_\beta G\overline{[y]_\beta}$.

2. G is a hermitian matrix.

3. G is an invertible matrix.

Proof. 1.

$$\begin{aligned}\langle x, y\rangle &= \langle\sum_{i=1}^n a_iv_i, \sum_{j=1}^n b_jv_j\rangle = \sum_{i=1}^n\sum_{j=1}^n\langle a_iv_i, b_jv_j\rangle\\ &= \sum_{i=1}^n\sum_{j=1}^n a_i\overline{b_j}\langle v_i, v_j\rangle = \sum_{i=1}^n a_i(\sum_{j=1}^n\langle v_i, v_j\rangle\overline{b_j}) = [x]^t_\beta G\overline{[y]_\beta}.\end{aligned}$$

2. The (i, j)-entry of $G^* = \overline{(j, i) - \text{entry of } G} = \overline{\langle v_j, v_i\rangle} = \langle v_i, v_j\rangle =$ (i, j)-entry of G. Thus, $G^* = G$.

3. Suppose $G[y]_\beta = 0$ for some $y \in V$. Then $G\overline{[x]_\beta} = 0$ where $[x]_\beta = \overline{[y]_\beta}$. Now $\langle x, x\rangle = [x]^t_\beta G\overline{[x]_\beta} = 0$. This implies $x = 0$ and thus $y = 0$. Therefore, the null space of G equals zero. This implies that $\ker(L_G) = 0$ and thus G is invertible by Theorem 3.24. □

Proposition 5.18. *Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V. Let $\mathcal{B}$ denote the set of all inner products $B : V \times V \to k$ on V. Consider the function $g_\beta : \mathcal{B} \to \mathcal{M}_{n\times n}(k)$ given by $g_\beta(B) = [B]_\beta$ where $B \in \mathcal{B}$. Then g_β is injective. The image of g_β is contained in the set of all hermitian matrices in $\mathcal{M}_{n\times n}(k)$.*

Proof. The function g_β is injective since the entries of $[B]_\beta$, $\langle v_i, v_j\rangle$, uniquely determine B. The image of g_β is contained in the set of hermitian matrices in $\mathcal{M}_{n\times n}(k)$ by Proposition 5.17. □

The precise image of g_β is determined in Corollary 5.21 below.

If β and γ are two bases of V, the next result gives the relation between $[B]_\beta$ and $[B]_\gamma$.

Proposition 5.19. *Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_n\}$ be two bases of V. Let $G = [B]_\beta$ and $H = [B]_\gamma$. Let $P = [1_V]^\beta_\gamma$. Then $H = P^tG\overline{P}$.*

Proof. The formula $[x]_\beta = [1_V]^\beta_\gamma[x]_\gamma = P[x]_\gamma$ and Proposition 5.17 imply for any $x, y \in V$ that $[x]^t_\gamma H\overline{[y]_\gamma} = \langle x, y\rangle = [x]^t_\beta G\overline{[y]_\beta} = [x]^t_\gamma P^tG\overline{P[y]_\gamma}$. Since this holds for all $x, y \in V$, it follows that $H = P^tG\overline{P}$. □

Exercise 5 asks for another proof of Proposition 5.19.

Theorem 5.20. *Let V be a vector space over k of dimension n and let $\beta = \{v_1, \dots, v_n\}$ be a basis of V. Let $B : V \times V \to k$ be an inner product on V. Then $G = [B]_\beta = P^t\overline{P}$ for some invertible matrix $P \in \mathcal{M}_{n\times n}(k)$.*

Conversely, let $P \in \mathcal{M}_{n\times n}(k)$ be an invertible matrix. Then there exists a unique inner product $B : V \times V \to k$ such that $[B]_\beta = P^t\overline{P}$.

Proof. First assume $B : V \times V \to k$ is an inner product on V. Let $\gamma = \{w_1, \dots, w_n\}$ be an orthonormal basis of V with respect to the inner product B. Let $P = [1_V]_\beta^\gamma$. Then $[x]_\gamma = [1_V]_\beta^\gamma[x]_\beta = P[x]_\beta$. Thus, P is invertible since 1_V is invertible. We have $[B]_\gamma = I_n$ because γ is an orthonormal basis of V with respect to B. We now have

$$\begin{aligned}[x]_\beta^t[B]_\beta\overline{[y]_\beta} &= \langle x, y\rangle_B = [x]_\gamma^t[B]_\gamma\overline{[y]_\gamma} = [x]_\gamma^t I_n\overline{[y]_\gamma} = [x]_\beta^t P^t I_n \overline{P}\,\overline{[y]_\beta} \\ &= [x]_\beta^t P^t\overline{P}\,\overline{[y]_\beta}.\end{aligned}$$

Since this holds for all $x, y \in V$, it follows that $[B]_\beta = P^t\overline{P}$.

Now let $P \in \mathcal{M}_{n\times n}(k)$ be an arbitrary invertible matrix. We will show that the function $B : V \times V \to k$ defined by $B(x, y) = [x]_\beta^t P^t\overline{P}\overline{[y]_\beta}$ is an inner product on V. Statements (1),(2) of Definition 5.1 are easily seen to hold. For statement (3) of Definition 5.1, we have

$$B(y, x) = [y]_\beta^t P^t\overline{P}\overline{[x]_\beta} = \overline{[x]_\beta}^t\overline{P}^t P[y]_\beta = \overline{[x]_\beta^t P^t\overline{P}\overline{[y]_\beta}} = \overline{B(x, y)}.$$

For statement (4), we have

$$B(x, x) = [x]_\beta^t P^t\overline{P}\overline{[x]_\beta} = (P[x]_\beta)^t\overline{P[x]_\beta} = [y]_\beta^t\overline{[y]_\beta},$$

for the unique vector $y \in V$ satisfying $[y]_\beta = P[x]_\beta$. Thus, $B(x, x) = [y]_\beta^t\overline{[y]_\beta} \geq 0$, with equality holding if and only if $[y]_\beta = 0$. But $[y]_\beta = 0$ if and only if $[x]_\beta = 0$ since P is invertible. Therefore $B(x, x) > 0$ if $x \neq 0$. This shows that statements (1)-(4) of Definition 5.1 hold, and thus B is an inner product.

To see that $[B]_\beta = P^t\overline{P}$, we observe that the (i, j)-entry of $[B]_\beta = B(v_i, v_j) = e_i^t P^t\overline{P}\overline{e_j} =$ the (i, j)-entry of $P^t\overline{P}$. The uniqueness follows from Proposition 5.18. □

We have now proved the following Corollary.

Corollary 5.21. $\mathrm{im}(g_\beta) = \{P^t\overline{P} | P \in \mathcal{M}_{n\times n}(k),\ P$ *invertible*$\}$. *Thus a matrix* G *has the form* $G = [B]_\beta$ *if and only if* $G = P^t\overline{P}$ *for some invertible matrix* $P \in \mathcal{M}_{n\times n}(k)$.

Corollary 5.22. *Let* $V = k^{(n)}$ *and let* $\beta = \epsilon_n$, *the standard basis of* $k^{(n)}$. *Let* $G \in \mathcal{M}_{n\times n}(k)$. *Consider the function* $B : V \times V \to k$, *where* $(x, y) \mapsto x^tG\overline{y}$. *Then* B *is an inner product on* $k^{(n)}$ *if and only if* $G = P^t\overline{P}$ *for some invertible matrix* $P \in \mathcal{M}_{n\times n}(k)$.

Let us now check how $P^t\overline{P}$ depends on the choice of the orthonormal basis γ of V in the proof of Theorem 5.20. Suppose $\delta = \{y_1, \ldots, y_n\}$ is another orthonormal basis of V. Let $Q = [1_V]_\gamma^\delta$. Then $[x]_\delta = [1_V]_\gamma^\delta[x]_\gamma = Q[x]_\gamma$. Since δ is an orthonormal basis of V, we have $\langle x, y\rangle_B = [x]_\delta^t I_n\overline{[y]_\delta} = [x]_\gamma^t Q^t I_n\overline{Q}\,\overline{[y]_\gamma}$. Since this equation holds for all $x, y \in V$, and we saw earlier that $\langle x, y\rangle_B = [x]_\gamma^t I_n\overline{[y]_\gamma}$, it follows that $Q^t\overline{Q} = I_n$.

Now we repeat the computation of $\langle x, y\rangle_B$ above with δ in place of γ. We have $[1_V]_\beta^\delta = [1_V]_\gamma^\delta[1_V]_\beta^\gamma = QP$. Therefore, $\langle x, y\rangle_B = [x]_\beta^t(QP)^t\overline{QP}\,\overline{[y]_\beta}$. But $(QP)^t\overline{QP} = P^tQ^t\overline{Q}\,\overline{P} = P^t\overline{P}$. This shows that a different choice of γ would replace P by QP where $Q^t\overline{Q} = I_n$.

5.4 The Adjoint of a Linear Transformation

Lemma 5.23. *Let* $(V, \langle\ \rangle)$ *be given and assume that* $\dim V = n$. *Let* $T : V \to V$ *be a linear transformation. Let* $w \in V$ *be given. Then*

1. *The function* $h : V \to k$, *given by* $v \mapsto \langle Tv, w\rangle$, *is a linear transformation.*

2. *There is a unique* $w' \in V$ *such that* $h(v) = \langle v, w'\rangle$ *for all* $v \in V$. *That is,* $h = \langle\ , w'\rangle$.

Proof. 1. $h(v_1 + v_2) = \langle T(v_1 + v_2), w\rangle = \langle Tv_1 + Tv_2, w\rangle = \langle Tv_1, w\rangle + \langle Tv_2, w\rangle = h(v_1) + h(v_2)$.

$h(cv) = \langle T(cv), w\rangle = \langle cTv, w\rangle = c\langle Tv, w\rangle = ch(v)$.

2. Let $\{y_1, \ldots, y_n\}$ be an orthonormal basis of V. Let $c_i = h(y_i)$, $1 \le i \le n$, and let $w' = \sum_{i=1}^n \overline{c_i}y_i$. Then $\langle y_j, w'\rangle = \langle y_j, \sum_{i=1}^n \overline{c_i}y_i\rangle = \sum_{i=1}^n c_i\langle y_j, y_i\rangle = c_j = h(y_j)$, $1 \le j \le n$. Therefore $h = \langle , w'\rangle$, since both linear transformations agree on a basis of V.

To show that w' is unique, suppose that $\langle v, w' \rangle = \langle v, w'' \rangle$ for all $v \in V$. Then $\langle v, w' - w'' \rangle = 0$ for all $v \in V$. Let $v = w' - w''$. Then $w' - w'' = 0$ and so $w' = w''$. □

Proposition 5.24. *Let $(V, \langle\ \rangle)$ be given and assume that $\dim V = n$. Let $T : V \to V$ be a linear transformation. Then there exists a unique linear transformation $T^* : V \to V$ such that $\langle Tv, w \rangle = \langle v, T^* w \rangle$ for all $v, w \in V$.*

Proof. Define $T^* : V \to V$ by $w \mapsto w'$ where $\langle Tv, w \rangle = \langle v, w' \rangle$, for all $v \in V$. Therefore, $\langle Tv, w \rangle = \langle v, T^* w \rangle$ for all $v, w \in V$. It remains to show that T^* is a linear transformation. The uniqueness of T^* is clear from the previous result.

For all $v \in V$,

$$\begin{aligned}\langle v, T^*(w_1 + w_2) \rangle &= \langle Tv, w_1 + w_2 \rangle = \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\ &= \langle v, T^* w_1 \rangle + \langle v, T^* w_2 \rangle = \langle v, T^* w_1 + T^* w_2 \rangle.\end{aligned}$$

Therefore, $T^*(w_1 + w_2) = T^* w_1 + T^* w_2$. For all $v \in V$,

$$\langle v, T^*(cw) \rangle = \langle Tv, cw \rangle = \overline{c} \langle Tv, w \rangle = \overline{c} \langle v, T^* w \rangle = \langle v, cT^* w \rangle.$$

Therefore, $T^*(cw) = cT^* w$. □

Definition 5.25. *The linear transformation T^* constructed in Proposition 5.24 is called the* adjoint *of T.*

Proposition 5.26. *Let $T, U \in \mathcal{L}(V, V)$ and let $c \in k$. Then*

1. *$(T + U)^* = T^* + U^*$.*
2. *$(cT)^* = \overline{c} T^*$.*
3. *$(TU)^* = U^* T^*$.*
4. *$T^{**} = T$. ($T^{**} means (T^*)^*$.)*

Proof. We have

$$\begin{aligned}\langle v, (T + U)^* w \rangle &= \langle (T + U)v, w \rangle = \langle Tv + Uv, w \rangle = \langle Tv, w \rangle + \langle Uv, w \rangle \\ &= \langle v, T^* w \rangle + \langle v, U^* w \rangle = \langle v, T^* w + U^* w \rangle = \langle v, (T^* + U^*) w \rangle.\end{aligned}$$

Thus, $(T + U)^* w = (T^* + U^*) w$ for all $w \in V$ and so $(T + U)^* = T^* + U^*$.

Next, we have

$$\langle v, (cT)^*w\rangle = \langle (cT)v, w\rangle = \langle c(Tv), w\rangle = c\langle Tv, w\rangle = c\langle v, T^*w\rangle = \langle v, \overline{c}T^*w\rangle.$$

Thus, $(cT)^*w = \overline{c}T^*w$ for all $w \in V$ and so $(cT)^* = \overline{c}T^*$.

Since $\langle v, (TU)^*w\rangle = \langle (TU)v, w\rangle = \langle Uv, T^*w\rangle = \langle v, U^*T^*w\rangle$, it follows that $(TU)^*w = U^*T^*w$ for all $w \in V$. Thus $(TU)^* = U^*T^*$.

Since $\langle v, T^{**}w\rangle = \langle T^*v, w\rangle = \overline{\langle w, T^*v\rangle} = \overline{\langle Tw, v\rangle} = \langle v, Tw\rangle$, it follows that $T^{**}w = Tw$ for all $w \in V$. Thus $T^{**} = T$. □

Proposition 5.27. *Let $T \in \mathcal{L}(V, V)$ and assume that T is invertible. Then T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.*

Proof. Let $1_V : V \to V$ be the identity linear transformation. Then $(1_V)^* = 1_V$, since $\langle v, (1_V)^*w\rangle = \langle 1_V(v), w\rangle = \langle v, w\rangle$. Thus $(1_V)^*w = w$ for all $w \in V$ and so $(1_V)^* = 1_V$.

Then $(T^{-1})^*T^* = (TT^{-1})^* = (1_V)^* = 1_V$, and $T^*(T^{-1})^* = (T^{-1}T)^* = (1_V)^* = 1_V$. Therefore, $(T^*)^{-1} = (T^{-1})^*$. □

Proposition 5.28. *Let $(V, \langle\ \rangle)$ be given and assume that* $\dim V = n$. *Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered orthonormal basis of V. Let $T : V \to V$ be a linear transformation. Let $A = [T]_\beta^\beta$ and let $B = [T^*]_\beta^\beta$. Then $B = A^*(= \overline{A}^t)$.*

Proof. Let $A = (a_{ij})_{n\times n}$ and let $B = (b_{ij})_{n\times n}$ Then $T(v_j) = \sum_{i=1}^n a_{ij}v_i$, $1 \le j \le n$, and $T^*(v_j) = \sum_{i=1}^n b_{ij}v_i$, $1 \le j \le n$. But

$$T(v_j) = \sum_{i=1}^n \langle T(v_j), v_i\rangle v_i$$

by Exercise 3 at the end of this chapter, since β is an orthonormal basis. Thus, $a_{ij} = \langle Tv_j, v_i\rangle$. Similar reasoning applied to T^* implies that $b_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle} = \overline{\langle Tv_i, v_j\rangle} = \overline{a_{ji}}$. Therefore $B = A^*$. □

5.5 Normal Transformations

Definition 5.29. *Let $(V, \langle\ \rangle)$ be given and let $T \in \mathcal{L}(V, V)$.*

1. *T is* normal *if $TT^* = T^*T$.*

2. *T is* hermitian *if $T^* = T$.*

3. *T is* skew-hermitian *if $T^* = -T$.*

4. *T is* unitary *if $TT^* = T^*T = 1_V$.*

5. *T is* positive *if $T = SS^*$ for some $S \in \mathcal{L}(V, V)$.*

Clearly, hermitian, skew-hermitian, and unitary transformations are special cases of normal transformations.

Definition 5.30. *Let $T \in \mathcal{L}(V, V)$.*

1. *An element $\alpha \in k$ is an* eigenvalue *of T if there is a nonzero vector $v \in V$ such that $Tv = \alpha v$.*

2. *An* eigenvector *of T is a nonzero vector $v \in V$ such that $Tv = \alpha v$ for some $\alpha \in k$.*

3. *For $\alpha \in k$, let $V_\alpha = \ker(T - \alpha 1_V)$. Then V_α is called the* eigenspace *of V associated to α.*

The definition of V_α depends on T although the notation doesn't indicate this. If α is not an eigenvalue of T, then $V_\alpha = \{0\}$.

The three special types of normal transformations above are distinguished by their eigenvalues.

Proposition 5.31. *Let α be an eigenvalue of T.*

1. *If T is hermitian, then $\alpha \in \mathbf{R}$.*

2. *If T is skew-hermitian, then α is pure imaginary. That is, $\alpha = ri$ where $r \in \mathbf{R}$ and $i = \sqrt{-1}$.*

3. *If T is unitary, then $|\alpha| = 1$.*

4. *If T is positive, then $\alpha \in \mathbf{R}$ and α is positive.*

Proof. Let $Tv = \alpha v$, $v \neq 0$.
1. $\alpha\langle v, v\rangle = \langle \alpha v, v\rangle = \langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \langle v, \alpha v\rangle = \overline{\alpha}\langle v, v\rangle$. Since $\langle v, v\rangle \neq 0$, it follows that $\alpha = \overline{\alpha}$ and so $\alpha \in \mathbf{R}$.
2. Following the argument in (1), we have $\alpha\langle v, v\rangle = \langle v, T^*v\rangle = \langle v, -Tv\rangle = \langle v, -\alpha v\rangle = -\overline{\alpha}\langle v, v\rangle$. Therefore $\alpha = -\overline{\alpha}$ and so $2\mathrm{Re}(\alpha) = \alpha + \overline{\alpha} = 0$. This implies $\alpha = ri$ where $r \in \mathbf{R}$ and $i = \sqrt{-1}$.

3. $\langle v,v\rangle = \langle v,T^*Tv\rangle = \langle Tv,Tv\rangle = \langle \alpha v,\alpha v\rangle = \alpha\overline{\alpha}\langle v,v\rangle$. Thus, $\alpha\overline{\alpha}=1$ and so $|\alpha|=1$.
4. As in (1), we have $\alpha\langle v,v\rangle = \langle v,T^*v\rangle = \langle v,SS^*v\rangle = \langle S^*v,S^*v\rangle > 0$. Since $\langle v,v\rangle > 0$, it follows that $\alpha > 0$. □

Definition 5.32. *Let $W\subseteq V$ be a subspace and let $T\in\mathcal{L}(V,V)$. Then W is a T-*invariant subspace *of V if $T(W)\subseteq W$.*

Proposition 5.33. *Let $T\in\mathcal{L}(V,V)$ and let W be a subspace of V. Assume that* $\dim V = n$.

1. *If W is a T-invariant subspace of V, then $W^\perp$ is a T^*-invariant subspace of V.*

2. *If W is a T-invariant subspace of V and T is normal, then $W^\perp$ is a T-invariant subspace of V.*

3. *Assume that W is a T-invariant subspace of V and T is normal. Using (2), let $T|_W : W\to W$ and $T|_{W^\perp} : W^\perp\to W^\perp$ be the linear transformations induced by T by restriction of domains. Then $T|_W$ and $T|_{W^\perp}$ are each normal.*

Proof. 1. Let $x\in W$ and let $y\in W^\perp$. Then $Tx\in W$ and $\langle x,T^*y\rangle = \langle Tx,y\rangle = 0$, since $y\in W^\perp$. Since x is an arbitrary vector in W, it follows that $T^*y\in W^\perp$ and so $T^*(W^\perp)\subseteq W^\perp$.
2. Let $\{v_1,\dots,v_l\}$ be an orthonormal basis of W and let $\{v_{l+1},\dots,v_n\}$ be an orthonormal basis of $W^\perp$. Then $\beta=\{v_1,\dots,v_l,v_{l+1},\dots,v_n\}$ is an orthonormal basis of V. Let $A=[T]_\beta^\beta$. Then $A^*=[T^*]_\beta^\beta$ by Proposition 5.28, since β is an orthonormal basis of V. We have $AA^*=A^*A$, since $TT^*=T^*T$. Since W is a T-invariant subspace of V, we have

$$A = \begin{pmatrix} C & D \\ 0 & E \end{pmatrix},$$

where

$$C\in\mathcal{M}_{l\times l}(k),\ D\in\mathcal{M}_{l\times(n-l)}(k),\ 0\in\mathcal{M}_{(n-l)\times l}(k),\ \text{and}\, E\in\mathcal{M}_{(n-l)\times(n-l)}(k).$$

Note that the submatrix 0 appears because W is T-invariant. The equation $AA^*=A^*A$ gives

$$\begin{pmatrix} C & D \\ 0 & E \end{pmatrix}\begin{pmatrix} C^* & 0 \\ D^* & E^* \end{pmatrix} = \begin{pmatrix} C^* & 0 \\ D^* & E^* \end{pmatrix}\begin{pmatrix} C & D \\ 0 & E \end{pmatrix}$$

$$\begin{pmatrix} CC^* + DD^* & DE^* \\ ED^* & EE^* \end{pmatrix} = \begin{pmatrix} C^*C & C^*D \\ D^*C & D^*D + E^*E \end{pmatrix}.$$

It is necessary to check carefully that the block matrix multiplication shown here works as indicated. The upper $l \times l$ matrix shows that $CC^* + DD^* = C^*C$ and so $DD^* = C^*C - CC^*$. It follows from this that $D = 0$. (See Exercise 10 for the details of this conclusion.) Therefore,

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

Since $A = [T]_\beta^\beta$, it follows that $W^\perp$ is T-invariant.
3. Since $D = 0$, the matrix equations above show that $CC^* = C^*C$ and $EE^* = E^*E$. If we let $\beta' = \{v_1, \ldots, v_l\}$ and $\beta'' = \{v_{l+1}, \ldots, v_n\}$, then we have $C = [T|_W]_{\beta'}^{\beta'}$ and $E = [T|_{W^\perp}]_{\beta''}^{\beta''}$. It follows that $T|_W$ and $T|_{W^\perp}$ are both normal. □

In order to apply Proposition 5.33 we require the following two facts. These two facts will be proved later in these notes.

Facts: Let $T : V \to V$ be a linear transformation. (There is no assumption here that V has an inner product.)

1. If $k = \mathbf{C}$, then V contains a one-dimensional T-invariant subspace.

2. If $k = \mathbf{R}$, then V contains a T-invariant subspace of dimension ≤ 2.

When $k = \mathbf{C}$, statement 1 implies that V contains an eigenvector. To see this, let W be a one-dimensional T-invariant subspace generated by the nonzero vector v. Then $Tv = \alpha v$ for some $\alpha \in k$ since $Tv \in W$. Thus v is an eigenvector with eigenvalue α.

Theorem 5.34. *Let $(V, \langle\ \rangle)$ be given and assume $\dim V = n$. Let $T : V \to V$ be a linear transformation and assume that T is normal.*

1. *If $k = \mathbf{C}$, then V can be written as an orthogonal direct sum of one-dimensional T-invariant subspaces. Thus, V has an orthonormal basis β consisting of eigenvectors of T and $[T]_\beta^\beta$ is a diagonal matrix.*

2. *If $k = \mathbf{R}$, then V can be written as an orthogonal direct sum of T-invariant subspaces of dimension ≤ 2. Let*

$$V = W_1 \bigoplus \cdots \bigoplus W_l \bigoplus W_{l+1} \bigoplus \cdots \bigoplus W_{l+s},$$

where

$$\dim W_i = \begin{cases} 1, & \textit{if } 1 \leq i \leq l \\ 2, & \textit{if } l+1 \leq i \leq l+s. \end{cases}$$

Each $T|_{W_i} : W_i \to W_i$ is a normal transformation.

3. *If $k = \mathbf{R}$, then V can be written as an orthogonal direct sum of one-dimensional T-invariant subspaces if and only if T is hermitian ($T^* = T$). That is, V has an orthonormal basis β consisting of eigenvectors of T if and only if $T^* = T$.*

Proof. 1. The proof is by induction on n. If $n = 1$, the result is clear. Now assume the result has been proved if $\dim V < n$ and assume $\dim V = n$. Using Fact 1, there exists a one-dimensional T-invariant subspace W_1. Then $V = W_1 \bigoplus W_1^\perp$, where $W_1^\perp$ is T-invariant and $T|_{W_1^\perp}$ is normal. Since $\dim W_1^\perp = n-1$, the induction hypothesis implies that $W_1^\perp$ can be written as an orthogonal direct sum of one-dimensional $T|_{W_1^\perp}$-invariant subspaces. Thus, $W_1^\perp = W_2 \bigoplus \cdots \bigoplus W_n$, where $\dim W_i = 1$, $2 \leq i \leq n$, W_i is $T|_{W_1^\perp}$-invariant, and the W_i's are pairwise orthogonal. Therefore $V = W_1 \bigoplus W_2 \bigoplus \cdots \bigoplus W_n$ where each W_i is a one-dimensional T-invariant subspace and the W_i's are pairwise orthogonal. Let $W_i = \langle v_i \rangle$, where $||v_i|| = 1$, $1 \leq i \leq n$. Then $\beta = \{v_1, \ldots, v_n\}$ is an orthonormal basis of V and $[T]_\beta^\beta$ is a diagonal matrix with (i,i)-entry equal to α_i, where $Tv_i = \alpha_i v_i$.

2. The proof of (2) is also by induction and is similar to the proof of (1) except we use Fact 2 in place of Fact 1. Let $W_i = \langle v_i \rangle$, $1 \leq i \leq l$, where $||v_i|| = 1$ and $Tv_i = \alpha_i v_i$. Let $\{y_i, z_i\}$ be an orthonormal basis of W_i, $l+1 \leq i \leq l+s$. Then $\beta = \{v_1, \ldots, v_l, y_{l+1}, z_{l+1}, \ldots, y_{l+s}, z_{l+s}\}$ is an orthonormal basis of V. We know that $T|_{W_i}$ is normal by Proposition 5.33.

It remains to describe normal transformations $T : V \to V$ when $k = \mathbf{R}$ and $\dim V = 2$. We will do that now before proving 3.

Proposition 5.35. *Let $k = \mathbf{R}$, $\dim V = 2$, $T \in \mathcal{L}(V, V)$, and assume that T is normal. Then either $T^* = T$ or $T^* = \lambda T^{-1}$ for some $\lambda \in \mathbf{R}$, $\lambda > 0$.*

Proof. Let β be an orthonormal basis of V and let $A = [T]_\beta^\beta$. Then $A^* = [T^*]_\beta^\beta$, since β is an orthonormal basis of V. Since $k = \mathbf{R}$, we have $A^* = A^t$. Let

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

Since T is normal, we have $TT^* = T^*T$ and so $AA^t = AA^* = A^*A = A^tA$. Then $AA^t = A^tA$ gives

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

$$\begin{pmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{pmatrix} = \begin{pmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{pmatrix}.$$

Comparing the $(1,1)$-entry gives $b^2 = c^2$. If $b = c$, then $A = A^t = A^*$ and so $T^* = T$. Now assume that $b = -c \neq 0$. Then $ac + bd = ab + cd$ implies $ac + (-c)d = a(-c) + cd$ and so $2c(a - d) = 0$. Since $2c \neq 0$, we must have $a = d$. We now have

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

It is easy to check that $AA^t = A^tA = (a^2 + c^2)I_2$. Therefore, $A^* = A^t = (a^2 + c^2)A^{-1}$ and so $T^* = \lambda T^{-1}$ for some $\lambda \in \mathbf{R}$. □

Proposition 5.36. *Let $k = \mathbf{R}$, $\dim V = 2$, $T \in \mathcal{L}(V, V)$, and assume that T is normal. Then $T^* = T$ if and only if V has a one-dimensional T-invariant subspace.*

Proof. First assume that V has a one-dimensional T-invariant subspace W. Then $V = W \bigoplus W^\perp$ and $W^\perp$ is also a T-invariant subspace by Proposition 5.33. Let $W = \langle v_1 \rangle$ and let $W^\perp = \langle v_2 \rangle$, where $||v_i|| = 1$, $i = 1, 2$. Then $\beta = \{v_1, v_2\}$ is an orthonormal basis of V and $A = [T]_\beta^\beta$ is a diagonal matrix, since v_1, v_2 are each eigenvectors of T. Since $A^* = A^t = A$ and β is an orthonormal basis of V, it follows that $T^* = T$ by Proposition 5.28.

Now assume that $T^* = T$. Let β be an orthonormal basis of V and let

$$A = [T]_\beta^\beta = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Since β is an orthonormal basis of V, we have $A = A^* = A^t$ and so $b = c$. A straight forward calculation shows that $A^2 - (a + d)A + (ad - b^2)I_2 = 0$. It follows that $T^2 - (a + d)T + (ad - b^2)1_V = 0$. Let $f(x) = x^2 - (a + d)x + (ad - b^2)$. Then $f(x) = 0$ has two (not necessarily distinct) real roots since the discriminant $(a + d)^2 - 4(ad - b^2) = (a - d)^2 + 4b^2 \geq 0$. Thus, $f(x) = (x-\alpha)(x-\gamma)$, where $\alpha, \gamma \in \mathbf{R}$. It follows that $(T-\alpha 1_V)(T-\gamma 1_V) = 0$. We may assume that $\ker(T - \alpha 1_V) \neq 0$. That means there exists a nonzero vector $v \in V$ such that $Tv = \alpha v$. Then $W = \langle v \rangle$ is a one-dimensional T-invariant subspace of V. □

Now we can finish the proof of 3 in Theorem 5.34. Let β be the orthonormal basis from the proof of 2 and let $A = [T]_\beta^\beta$. Let $\beta_i = \{v_i\}$, if $1 \le i \le l$, and let $\beta_i = \{y_i, z_i\}$, if $l+1 \le i \le l+s$. Then β_i is an orthonormal basis of W_i, $1 \le i \le l+s$ and $\beta = \beta_1 \cup \cdots \cup \beta_{l+s}$. Let $A_i = [T|_{W_i}]_{\beta_i}^{\beta_i}$. Since each W_i is a T-invariant subspace of V, β_i is an orthonormal basis of W_i, β is an orthonormal basis of V, each $T|_{W_i}$ is normal (by 2), Proposition 5.36, and $k = \mathbf{R}$, we have the following conclusions.

T is hermitian $\Leftrightarrow T^* = T \Leftrightarrow A^* = A \Leftrightarrow A^t = A \Leftrightarrow A_i^t = A_i \Leftrightarrow A_i^* = A_i \Leftrightarrow (T|_{W_i})^* = T|_{W_i} \Leftrightarrow$ each W_i has a one-dimensional $T|_{W_i}$-invariant subspace $\Leftrightarrow$ each W_i can be written as an orthogonal direct sum of one-dimensional $T|_{W_i}$-invariant subspaces $\Leftrightarrow$ V can be written as an orthogonal direct sum of one-dimensional T-invariant subspaces. □

Some special terminology for the case $k = \mathbf{R}$:

Definition 5.37. *Assume $k = \mathbf{R}$ and let $T : V \to V$ be a linear transformation.*

1. *If $T^* = T$, then T is* symmetric.
2. *If $T^* = -T$, then T is* skew-symmetric.
3. *If $TT^* = T^*T = 1_V$, then T is* orthogonal.

Some terminology for matrices over $\mathbf{C}$ and $\mathbf{R}$:

Definition 5.38. *Let $A \in \mathcal{M}_{n\times n}(k)$. Recall that $A^* = \overline{A}^t = \overline{A^t}$.*

1. *A is* normal *$\Leftrightarrow AA^* = A^*A \Leftrightarrow \overline{A}A^t = A^t\overline{A}$.*
2. *A is* hermitian *$\Leftrightarrow A^* = A \Leftrightarrow A^t = \overline{A}$. If $k = \mathbf{R}$ and A is hermitian, then $A^t = A$ and A is called* symmetric.
3. *A is* skew-hermitian *$\Leftrightarrow A^* = -A \Leftrightarrow A^t = -\overline{A}$. If $k = \mathbf{R}$ and A is skew-hermitian, then $A^t = -A$ and A is called* skew-symmetric.
4. *A is* unitary *$\Leftrightarrow AA^* = A^*A = I_n \Leftrightarrow A^t\overline{A} = \overline{A}A^t = I_n$. If $k = \mathbf{R}$ and A is unitary, then $AA^t = A^tA = I_n$ and A is called* orthogonal.

If A is a unitary matrix, then the columns of A are orthonormal with respect to the standard inner product (pairwise orthogonal and have norm 1) because $A^*A = I_n$ implies $A^t\overline{A} = I_n$. The rows of A are also orthonormal with respect to the standard inner product because $AA^* = I_n$. If $k = \mathbf{R}$ and A is orthogonal, note that this requires the rows and columns to be orthonormal, not just orthogonal. Tradition requires us to live with this confusing terminology.

Theorem 5.39. *Let $A \in \mathcal{M}_{n\times n}(k)$.*

1. *Assume $k = \mathbf{C}$. Then A is normal if and only if there exists a unitary matrix P such that P^*AP is a diagonal matrix.*
2. *Assume $k = \mathbf{R}$. Then A is a symmetric matrix if and only if there exists an orthogonal matrix P such that P^tAP is a diagonal matrix.*

Proof. 1. First assume that P is a unitary matrix such that $P^*AP = D$ is a diagonal matrix. Then

$$(P^*AP)(P^*AP)^* = DD^* = D^*D = (P^*AP)^*(P^*AP)$$

$$P^*APP^*A^*P = P^*A^*PP^*AP.$$

Then $AA^* = A^*A$, since $PP^* = I_n$ and P, P^* are both invertible.

Now assume that A is normal. We may assume that $A = [T]_\gamma^\gamma$ for some $T \in \mathcal{L}(V, V)$ and some orthonormal basis γ of V. Since $k = \mathbf{C}$, V contains an orthonormal basis β consisting of eigenvectors of T, by Theorem 5.34. Let $P = [1_V]_\beta^\gamma$. Then P is a unitary matrix since β and γ are both orthonormal bases of V. (See Exercise 8(b).) Then $[T]_\beta^\beta$ is a diagonal matrix and $[T]_\beta^\beta = [1_V]_\gamma^\beta[T]_\gamma^\gamma[1_V]_\beta^\gamma = P^{-1}AP = P^*AP$.

2. First assume that P is an orthogonal matrix such that $P^tAP = D$ is a diagonal matrix. Then $(P^tAP)^t = D^t = D = P^tAP$. Thus, $P^tA^tP = P^tAP$ and so $A^t = A$, since P is invertible. Therefore A is symmetric.

Now assume that A is symmetric. We may assume that $A = [T]_\gamma^\gamma$ for some $T \in \mathcal{L}(V, V)$ and some orthonormal basis γ of V. Since $A = A^t = A^*$ and γ is an orthonormal basis of V, it follows that $T = T^*$ and T is hermitian. By Theorem 5.34(3), V contains an orthonormal basis β consisting of eigenvectors of T. Then $[T]_\beta^\beta$ is a diagonal matrix. As in 1, let

$P = [1_V]_\beta^\gamma$. Then P is an orthogonal matrix since β and γ are both orthonormal bases of V. (See Exercise 8(b).) Then $[T]_\beta^\beta$ is a diagonal matrix and $[T]_\beta^\beta = [1_V]_\gamma^\beta [T]_\gamma^\gamma [1_V]_\beta^\gamma = P^{-1}AP = P^tAP$. □

If k is an arbitrary field, char k different from 2, and $A \in \mathcal{M}_{n\times n}(k)$ is a symmetric matrix, then it can be shown that there exists an invertible matrix P such that P^tAP is a diagonal matrix. The field of real numbers $\mathbf{R}$ has the very special property that one can find an *orthogonal* matrix P such that P^tAP is diagonal.

Let $A \in \mathcal{M}_{n\times n}(\mathbf{R})$, A symmetric. Here is a nice way to remember how to find the orthogonal matrix P in Theorem 5.39(2). Let ϵ_n be the standard basis of $\mathbf{R}^{(n)}$. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis of $\mathbf{R}^{(n)}$ consisting of eigenvectors of A. Let P be the matrix whose j^{th} column is the eigenvector v_j expressed in the standard basis ϵ_n. Then $P = [1_V]_\beta^{\epsilon_n}$. Recall that P is an orthogonal matrix since the columns of P are orthonormal. Let $Av_j = \alpha_j v_j$, $1 \le j \le n$. Let $D \in \mathcal{M}_{n\times n}(\mathbf{R})$ be the diagonal matrix whose (j,j)-entry is α_j. Then it is easy to see that $AP = PD$ and so $D = P^{-1}AP = P^tAP$.

Exercises

1. In proposition 5.4(3), equality in the Cauchy-Schwarz inequality holds if and only if v, w are linearly dependent.

2. Prove Proposition 5.5.

3. Let $\{v_1, \ldots, v_n\}$ be an orthogonal set of nonzero vectors. If $w = \sum_{i=1}^n c_i v_i$, $c_i \in k$, then
$$c_j = \frac{\langle w, v_j\rangle}{||v_j||^2}.$$

4. Let $(V, \langle\ \rangle)$ be given, with $\dim V$ finite. Let S_1, S_2 be subsets of V and let $W_1 = \text{Span}(S_1)$, $W_2 = \text{Span}(S_2)$. Recall that $S_i^\perp = W_i^\perp$, $i = 1, 2$. Assume that $0 \in S_i$, $i = 1, 2$. Show the following are true.

   (a) $$\begin{aligned} (S_1+S_2)^\perp &= (W_1+W_2)^\perp = W_1^\perp \cap W_2^\perp = S_1^\perp \cap S_2^\perp \\ &\ (W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp = S_1^\perp + S_2^\perp. \end{aligned}$$

(b) It is possible that $(S_1 \cap S_2) \subsetneq (W_1 \cap W_2)$ and $(W_1 \cap W_2)^\perp \subsetneq (S_1 \cap S_2)^\perp$.

5. Give another proof of Proposition 5.19 by directly computing the (r, s)-entry of H, namely $\langle w_r, w_s \rangle$, and showing that it is equal to the (r, s)-entry of $P^t G \overline{P}$.

6. Let $G, H \in \mathcal{M}_{n\times n}(k)$ and suppose $x^t G \overline{y} = x^t H \overline{y}$ for all $x, y \in k^{(n)}$. Then show $G = H$.

7. Suppose $R^t \overline{R} = P^t \overline{P}$, where P, R are invertible matrices in $\mathcal{M}_{n\times n}(k)$. Let $Q = PR^{-1}$. Then $Q^t \overline{Q} = I_n$.

8. (a) Suppose that β, γ are bases of V. Then $[B]_\gamma = ([1_V]_\gamma^\beta)^t [B]_\beta \overline{[1_V]_\gamma^\beta}$.

   (b) Assume that β, γ are orthonormal bases of V. Let $R = [1_V]_\beta^\gamma$. Then $R^t \overline{R} = I_n$.

9. Let $T \in \mathcal{L}(V, V)$ and let W be a subspace of V. Assume that W is a T-invariant subspace of V and that T is normal. Then W is a T^*-invariant subspace of V.

10. Let $A \in \mathcal{M}_{n\times n}(k)$, $A = (a_{ij})_{n\times n}$. Define the trace of A to be $\text{tr}A = a_{11} + a_{22} + \cdots a_{nn}$, the sum of the diagonal entries of A.

    (a) If $A, B \in \mathcal{M}_{n\times n}(k)$, then $\text{tr}(A + B) = \text{tr}A + \text{tr}B$.

    (b) If $A, B \in \mathcal{M}_{n\times n}(k)$, then $\text{tr}(AB) = \text{tr}(BA)$.

    (c) If $A \in \mathcal{M}_{n\times n}(k)$, compute $\text{tr}(AA^*)$.

    (d) If $\text{tr}(AA^*) = 0$, then $A = 0$.

    (e) If $BB^* = A^*A - AA^*$, then $B = 0$.

11. Let $T : V \to V$ be a normal linear transformation. Use the following parts to show that if $T^m v = 0$ for some vector $v \in V$ and some positive integer m, then $Tv = 0$. ($T^m = T \circ \cdots \circ T$.)

    (a) If $R : V \to V$ is a linear transformation and $R^* R v = 0$, then $Rv = 0$.

    (b) If $S : V \to V$ is hermitian and $S^2 v = 0$, then $Sv = 0$.

    (c) If $S : V \to V$ is hermitian and $S^m v = 0$, then $Sv = 0$.

(d) Let $S = T^*T$. Then $S^m = (T^*)^m T^m$.

(e) If $T^m v = 0$, then $Tv = 0$.

12. Suppose $k = \mathbf{R}$, $\dim V = 2$, $T \in \mathcal{L}(V, V)$, and T is normal. Describe all such T's such that $T^* = T$ and $T^* = \lambda T^{-1}$. In Proposition 5.35, if $b = -c \neq 0$, then V does not contain a one-dimensional T-invariant subspace.

13. Describe all orthogonal matrices $A \in \mathcal{M}_{2\times 2}(\mathbf{R})$.

14. Let $A \in \mathcal{M}_{2\times 2}(\mathbf{R})$ be a symmetric matrix. Find an orthogonal matrix $P \in \mathcal{M}_{2\times 2}(\mathbf{R})$ such that $P^t AP$ is diagonal.

Chapter 6

Determinants

6.1 The Existence of n-Alternating Forms

In this chapter, k denotes an arbitrary field.

Definition 6.1. *Let $V_1, \ldots, V_n, W$ be vector spaces over k. A function*

$$f : V_1 \times \cdots \times V_n \to W$$

is multilinear *(or n-*multilinear*) if f is linear in each variable. That is, if j is given and $v_i \in V_i$, $i \neq j$, then the function $f_j : V_j \to W$, given by $f_j(x) = f(v_1, \ldots, v_{j-1}, x, v_{j+1}, \ldots, v_n)$ is a linear transformation.*

We will usually assume that $V = V_1 = \cdots = V_n$ and $W = k$. In this case, a multilinear function (or n-multilinear function) is called a multilinear form (or n-multilinear form) on V.

Definition 6.2. *An n-multilinear function $f : V \times \cdots \times V \to W$ is called an* alternating function *(or n-*alternating function*) if*

1. *f is a multilinear function and*
2. *$f(v_1, \ldots, v_n) = 0$ whenever two adjacent arguments of f are equal. That is, if $v_i = v_{i+1}$, for some i, $1 \leq i \leq n-1$, then $f(v_1, \ldots, v_n) = 0$.*

If $W = k$, we call f an alternating form *(or n-*alternating form*) on V.*

If $f : V \times \cdots \times V \to W$ is defined by $f(v_1, \ldots, v_n) = 0$ for all $v_1, \ldots, v_n \in V$, then f is clearly an n-alternating form on V.

Proposition 6.3. *Let $f : V \times \cdots \times V \to W$ be an n-alternating function. Let $v_1, \dots, v_n \in V$. Then the following statements hold.*

1. $f(v_1, \dots, v_{i+1}, v_i, \dots, v_n) = -f(v_1, \dots, v_{i-1}, v_i, v_{i+1}, v_{i+2}, \dots, v_n)$. *(Interchanging two adjacent arguments of f multiplies $f(v_1, \dots, v_n)$ by -1.)*

2. *If $v_i = v_j$, $i \neq j$, then $f(v_1, \dots, v_i, \dots v_j, \dots, v_n) = 0$.*

3. *If $i < j$, then*

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -f(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

(Interchanging any two distinct arguments of f multiplies the value of f by -1.)

4. *Let $a \in k$ and let $i \neq j$. Then*

$$f(v_1, \dots, v_n) = f(v_1, \dots, v_{i-1}, v_i + av_j, v_{i+1}, \dots, v_n).$$

Proof. (1) Let $g(x, y) = f(v_1, \dots, v_{i-1}, x, y, v_{i+2}, \dots, v_n)$, $x, y \in V$. Then g is also multilinear and alternating, and so we have

$$0 = g(x + y, x + y) = g(x, x) + g(x, y) + g(y, x) + g(y, y) = g(x, y) + g(y, x).$$

Thus, $g(y, x) = -g(x, y)$, and so $g(v_{i+1}, v_i) = -g(v_i, v_{i+1})$. This gives (1).

(2) We successively interchange adjacent arguments of f until two adjacent arguments are equal. Then (1) implies that this changes the value of $f(v_1, \dots, v_i, \dots v_j, \dots, v_n)$ by a factor of ±1. Since f is alternating, the new value of f is 0. Therefore $f(v_1, \dots, v_i, \dots v_j, \dots, v_n) = 0$.

(3) This follows immediately from (2) using the argument in (1).

(4) The multilinearity of f and the result in (2) give

$$\begin{aligned} f(v_1, \dots, v_{i-1}, &v_i + av_j, v_{i+1}, \dots, v_n) \\ &= f(v_1, \dots, v_i, \dots, v_n) + af(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_n) \\ &= f(v_1, \dots, v_i, \dots, v_n). \end{aligned}$$

□

If $\operatorname{char} k \neq 2$, then condition (1) of Proposition (6.3) can be taken as the definition of an alternating function. To see this, suppose that

$$f : V \times \cdots \times V \to W$$

is an n-multilinear function and suppose that condition (1) of Proposition (6.3) holds. Assume that $v_i = v_{i+1}$, where $1 \leq i \leq n-1$, and define $g(x, y)$ as in the proof of condition (1) of Proposition 6.3. Then condition (1) implies that $g(x, x) = -g(x, x)$ for all $x \in V$. Then $2g(x, x) = 0$. If $\operatorname{char} k \neq 2$, then $g(x, x) = 0$, and thus the original defining condition of an alternating function holds.

If $\operatorname{char} k = 2$, then there are n-multilinear functions that satisfy condition (1) of Proposition 6.3, but don't satisfy the original defining condition of an alternating function.

We have not yet given an example of a nonzero alternating function as it requires some work to produce such examples. We seemingly digress for a moment to discuss some results about permutations and then return to the problem of producing nontrivial alternating functions. (Permutations were introduced in Chapter 3, Section 5.)

We now recall some concepts from Chapter 3, Section 5. Let T_n denote the set of all functions $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. Let S_n be the subset of T_n consisting of all bijective functions

$$\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}.$$

The elements of S_n are called permutations of $\{1, 2, \ldots, n\}$ since each $\sigma \in S_n$ corresponds to a rearrangement or reordering of $\{1, 2, \ldots, n\}$. An easy calculation shows that $|T_n| = n^n$ and $|S_n| = n!$.

Let $\sigma \in S_n$. The ordered set $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}$ can be rearranged to the ordered set $\{1, 2, \ldots, n\}$ by a sequence of steps, called transpositions, that only interchange two elements at time. Let $\epsilon(\sigma)$ denote the number of interchanges required to do this. Although $\epsilon(\sigma)$ is not necessarily well defined, Theorem 6.4 below states that $\epsilon(\sigma)$ is well defined modulo 2. We'll see below that the result in Theorem 6.4 is essentially equivalent to the existence of a nonzero n-alternating function on V. Namely, in Proposition 6.6, we will show that if $\operatorname{char} k \neq 2$ and if there exists a nonzero n-alternating function on V, then $\epsilon(\sigma)$ is well defined modulo 2. Conversely, if $\epsilon(\sigma)$ is well defined modulo 2, then Exercise 1 shows that there is a nonzero n-alternating form on V. We will prove that $\epsilon(\sigma)$ is well defined modulo 2 as a consequence of Proposition 6.6 and Theorem 6.10.

Theorem 6.4. *Using the notation above, $\epsilon(\sigma)$ is well defined modulo 2.*

We will prove Theorem 6.4 later. Theorem 6.4 allows us to make the following definition.

Definition 6.5. *A permutation $\sigma \in S_n$ is called* even *(or* odd*), if $\epsilon(\sigma)$ is even (or odd).*

Proposition 6.6. *Let V, W be vector spaces defined over k and assume that* $\operatorname{char} k \neq 2$. *Suppose that there exists a nonzero n-alternating function*

$$f : V \times \cdots \times V \to W.$$

Then $\epsilon(\sigma)$ is well defined modulo 2 for every $\sigma \in S_n$.

Proof. Choose $v_1, \ldots, v_n \in V$ such that $f(v_1, \ldots, v_n) \neq 0$ and let $\sigma \in S_n$. We apply Proposition 6.3(3) repeatedly to see that $f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = (-1)^{\epsilon(\sigma)} f(v_1, \ldots, v_n)$. Since $f(v_1, \ldots, v_n) \neq -f(v_1, \ldots, v_n)$, it follows that for any sequence of transpositions bringing $(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$ back to the original order $v_1, \ldots, v_n$, the value $(-1)^{\epsilon(\sigma)}$ is always the same. Therefore, $\epsilon(\sigma)$ well defined modulo 2. □

The statement that $\epsilon(\sigma)$ is well defined modulo 2 has nothing to do with whether a field has characteristic different from 2. Thus, although we will show in Theorem 6.10 that there exist n-alternating functions over all fields, Proposition 6.6 implies that in order to show that $\epsilon(\sigma)$ is well defined modulo 2, we need only show the existence of an n-alternating function over just one field k having characteristic different from 2.

The following Proposition contains a calculation that is central to the rest of this chapter.

Proposition 6.7. *Assume that $\epsilon(\sigma)$ is well defined modulo 2 and let*

$$f : V \times \cdots \times V \to W$$

be an n-alternating function on V. Let $v_1, \ldots, v_n \in V$ be arbitrary vectors. Let $w_j = \sum_{i=1}^{n} a_{ij} v_i$, $1 \leq j \leq n$, $a_{ij} \in k$. Then

$$f(w_1, \ldots, w_n) = \sum_{\sigma \in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1} \cdots a_{\sigma(n)n} f(v_1, \ldots, v_n).$$

Proof.

$$\begin{aligned}
f(w_1, \ldots, w_n) &= f(a_{11}v_1 + \cdots + a_{n1}v_n, \ldots, a_{1n}v_1 + \cdots + a_{nn}v_n) \\
&= \sum_{\sigma \in T_n} a_{\sigma(1)1} \cdots a_{\sigma(n)n} f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} a_{\sigma(1)1} \cdots a_{\sigma(n)n} f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) \\
&= \sum_{\sigma \in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1} \cdots a_{\sigma(n)n} f(v_1, \ldots, v_n).
\end{aligned}$$

□

Notations. Let $A \in \mathcal{M}_{n\times n}(k)$.

1. Let A_j denote the j^{th} column of A. Sometimes we will write $A = (A_1, \ldots, A_n)$
2. Let A_{ij} denote the $(n-1)\times(n-1)$ submatrix of A obtained by deleting the i^{th} row and j^{th} column of A.

Definition 6.8. *Let $A \in \mathcal{M}_{n\times n}(k)$ and consider the columns of A as elements of $k^{(n)}$. An $n \times n$* determinant *is a function* $\det : \mathcal{M}_{n\times n}(k) \to k$*, where $A \to \det A$, that satisfies the following two conditions.*

1. $\det : k^{(n)} \times \cdots \times k^{(n)} \to k$ *is an n-alternating form on the columns of A.*

2. $\det I_n = 1$.

If $A \in \mathcal{M}_{n\times n}(k)$, we will write either $\det A$ or $\det(A_1, \ldots, A_n)$, whichever is more convenient.

Proposition 6.9. *Suppose that char $k \neq 2$. If a determinant function* $\det : \mathcal{M}_{n\times n}(k) \to k$ *exists, then it is unique. In particular, $\epsilon(\sigma)$ is well defined modulo 2, and if $A = (a_{ij})_{n\times n}$, then*

$$\det A = \sum_{\sigma \in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1} \cdots a_{\sigma(n)n}.$$

Proof. The existence of a determinant function implies that $\epsilon(\sigma)$ is well defined modulo 2 by Definition 6.8 and Proposition 6.6.

Let $\{e_1, \dots, e_n\}$ be the standard basis of $k^{(n)}$ and let $A \in \mathcal{M}_{n\times n}(k)$. Then $A_j = \sum_{i=1}^n a_{ij}e_i$. Proposition 6.7 implies that

$$\begin{aligned}\det A = \det(A_1, \dots, A_n) &= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1}\cdots a_{\sigma(n)n} \det(e_1, \dots, e_n) \\ &= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1}\cdots a_{\sigma(n)n},\end{aligned}$$

because $\det(e_1, \dots, e_n) = \det I_n = 1$. This shows that det is uniquely determined by the entries of A. □

We now show that a determinant function $\det : \mathcal{M}_{n\times n}(k) \to k$ exists. One possibility would be to show that the formula given in Proposition 6.9 satisfies the two conditions in Definition 6.8. (See Exercise 1.) Instead we shall follow a different approach that will yield more results in the long run.

Theorem 6.10. *For each $n \geq 1$ and every field k, there exists a determinant function* $\det : \mathcal{M}_{n\times n}(k) \to k$.

Proof. The proof is by induction on n. If $n = 1$, let $\det : \mathcal{M}_{1\times 1}(k) \to k$, where $\det(a)_{1\times 1} = a$. Then det is linear (1-multilinear) on the column of $(a)_{1\times 1}$ and $\det I_1 = 1$. The alternating condition is satisfied vacuously.

Now let $n \geq 2$ and assume that we have proved the existence of a determinant function $\det : \mathcal{M}_{(n-1)\times(n-1)}(k) \to k$. Let $A \in \mathcal{M}_{n\times n}(k)$, $A = (a_{ij})_{n\times n}$. Choose an integer i, $1 \leq i \leq n$, and define $\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$. This expression makes sense because $\det A_{ij}$ is defined by our induction hypothesis. Showing that $\det A$ is a multilinear form on the n columns of A is the same as showing that $\det A$ is a linear function (or transformation) on the l^{th} column of A, $1 \leq l \leq n$. Since a sum of linear transformations is another linear transformation, it is sufficient to consider separately each term $(-1)^{i+j} a_{ij} \det A_{ij}$.

If $j \neq l$, then a_{ij} doesn't depend on the l^{th} column of A and $\det A_{ij}$ is a linear function on the l^{th} column of A, by the induction hypothesis, because A_{ij} is an $(n-1) \times (n-1)$ matrix. Therefore, $(-1)^{i+j} a_{ij} \det A_{ij}$ is a linear function on the l^{th} column of A.

If $j = l$, then $a_{ij} = a_{il}$ is a linear function on the l^{th} column of A and $\det A_{ij} = \det A_{il}$ doesn't depend on the l^{th} column of A because A_{il} is obtained from A by deleting the l^{th} column of A. Therefore, $(-1)^{i+j} a_{ij} \det A_{ij}$ is a linear function on the l^{th} column of A in this case also.

We conclude that $\det A$ is a multilinear form on the n columns of A. Next we show that $\det A$ is an alternating form on the n columns of A. Suppose $A_l = A_{l+1}$. If $j \neq l, l+1$, then A_{ij} has two adjacent columns that are equal. Therefore $\det A_{ij} = 0$ by the induction hypothesis. The remaining two terms in the expression for $\det A$ are

$$(-1)^{i+l} a_{il} \det A_{il} + (-1)^{i+(l+1)} a_{i(l+1)} \det A_{i(l+1)}.$$

We have $A_{il} = A_{i(l+1)}$ and $a_{il} = a_{i(l+1)}$ because the l^{th}, $(l+1)^{th}$ columns of A are equal. Therefore,

$$\begin{aligned}(-1)^{i+l} a_{il} \det A_{il} &+ (-1)^{i+(l+1)} a_{i(l+1)} \det A_{i(l+1)} \\ &= a_{il} \det A_{il}((-1)^{i+l} + (-1)^{i+(l+1)}) = 0.\end{aligned}$$

Thus $\det A = 0$ and so $\det A$ is an alternating form on the n columns of A.

Now we show that $\det I_n = 1$. Letting $I_n = A$ in the formula for $\det A$ and noting that $a_{ij} = 0$ when $i \neq j$ and $a_{ii} = 1$, we have

$$\det I_n = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = (-1)^{i+i} a_{ii} \det A_{ii} = \det A_{ii} = 1,$$

by the induction hypothesis, because $A_{ii} = I_{n-1}$. Therefore det is a determinant function. □

The formula given for $\det A$ in Theorem 6.10 is called the expansion of $\det A$ along the i^{th} row of A.

We have now established the existence of a nonzero n-alternating form on $k^{(n)}$ for an arbitrary field k. Since any vector space V of dimension n over k is isomorphic to $k^{(n)}$, it follows that we have established the existence of nonzero n-alternating forms on V for any n-dimensional vector space V over k. By Proposition 6.6, we have now proved that $\epsilon(\sigma)$ is well defined modulo 2 for any $\sigma \in S_n$.

The uniqueness part of Proposition 6.9 shows that our definition of $\det A$ does not depend on the choice of i given in the proof of Theorem 6.10. Therefore, we can compute $\det A$ by expanding along any row of A.

We now restate Proposition 6.7 in the following Corollary using the result from Proposition 6.9.

Corollary 6.11. *Let $f : V \times \cdots \times V \to W$ be an n-alternating function on V. Let $v_1, \ldots, v_n \in V$ be arbitrary vectors. Let $w_j = \sum_{i=1}^n a_{ij} v_i$, $1 \leq j \leq n$, $a_{ij} \in k$. Let $A = (a_{ij})_{n\times n} \in \mathcal{M}_{n\times n}(k)$. Then*

$$f(w_1, \ldots, w_n) = (\det A) f(v_1, \ldots, v_n).$$

The next result summarizes the information we need for the function $\epsilon(\sigma)$, where $\sigma \in S_n$.

Proposition 6.12. *1. $\epsilon(\sigma)$ is well defined modulo 2 for all $\sigma \in S_n$.*

2. $\epsilon(\sigma\tau) \equiv \epsilon(\sigma) + \epsilon(\tau) \pmod 2$, for all $\sigma, \tau \in S_n$.

3. $\epsilon(\sigma^{-1}) \equiv \epsilon(\sigma) \pmod 2$, for all $\sigma \in S_n$.

Proof. (1) This follows from Proposition 6.6 and Theorem 6.10.

(2) This follows from counting the number of interchanges needed to pass through the sequences $\{\sigma\tau(1), \ldots, \sigma\tau(n)\}$, $\{\tau(1), \ldots, \tau(n)\}$, $\{1, 2, \ldots, n\}$.

(3) $\epsilon(\sigma) + \epsilon(\sigma^{-1}) \equiv \epsilon(\sigma\sigma^{-1}) \equiv \epsilon(\text{identity}) \equiv 0 \pmod 2$. Therefore, $\epsilon(\sigma^{-1}) \equiv \epsilon(\sigma) \pmod 2$, for all $\sigma \in S_n$. □

Proposition 6.13. *Let $A \in \mathcal{M}_{n\times n}(k)$. Then $\det(A^t) = \det A$.*

Proof. Let $A = (a_{ij})_{n\times n}$ and let $A^t = (b_{ij})_{n\times n}$. Then $b_{ij} = a_{ji}$ and

$$\begin{aligned}
\det(A^t) &= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma)} b_{\sigma(1)1} \cdots b_{\sigma(n)n} \\
&= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma)} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\
&= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma^{-1})} a_{\sigma^{-1}(\sigma(1))\sigma(1)} \cdots a_{\sigma^{-1}(\sigma(n))\sigma(n)} \\
&= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma^{-1})} a_{\sigma^{-1}(1)1} \cdots a_{\sigma^{-1}(n)n} \\
&= \sum_{\sigma\in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1} \cdots a_{\sigma(n)n} \\
&= \det A.
\end{aligned}$$

□

Proposition 6.14.

1. *The function* $\det\colon \mathcal{M}_{n\times n}(k) \to k$ *is an n-alternating form on the rows of $n \times n$ matrices.*

2. *The determinant of a matrix $A \in \mathcal{M}_{n\times n}(k)$ can be computed by expanding along any row or any column of A.*

Proof. (1) Since $\det(A^t) = \det A$ and the rows of A are the columns of A^t, the result follows from the properties of det applied to the columns of A^t.

(2) We have already seen that $\det A$ can be computed by expanding along any row of A. Now let $B = A^t$, $B = (b_{ij})_{n\times n}$. Then $b_{ij} = a_{ji}$ and $B_{ij} = (A_{ji})^t$. If we expand along the j^{th} column of A, we have

$$\begin{aligned}\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij} &= \sum_{i=1}^{n}(-1)^{i+j}b_{ji}\det(B_{ji})^t \\ &= \sum_{i=1}^{n}(-1)^{i+j}b_{ji}\det B_{ji} = \det B = \det(A^t) = \det A,\end{aligned}$$

where we have expanded along the j^{th} row of B. □

6.2 Further Results and Applications

Theorem 6.15. *Let $A, B \in \mathcal{M}_{n\times n}(k)$. Then* $\det(AB) = (\det A)(\det B)$.

Proof. Let $A = (a_{ij})_{n\times n}$, let $B = (b_{ij})_{n\times n}$ and let $C = AB$. Denote the columns of A, B, C by A_j, B_j, C_j, respectively, $1 \le j \le n$. Let $e_1, \ldots, e_n$ denote the standard basis of $k^{(n)}$. Our results on matrix multiplication imply that $C_j = AB_j = \sum_{i=1}^n b_{ij}A_i$, $1 \le j \le n$. Since $A_j = \sum_{i=1}^n a_{ij}e_i$, Corollary 6.11 implies that

$$\begin{aligned}\det(AB) &= \det C = \det(C_1, \ldots, C_n) \\ &= (\det B)\det(A_1, \ldots, A_n) \\ &= (\det B)(\det A)\det(e_1, \ldots, e_n) \\ &= (\det A)(\det B)\det I_n \\ &= (\det A)(\det B).\end{aligned}$$

□

Here is a second proof of Theorem 6.15: Define $f : \mathcal{M}_{n\times n}(k) \to k$ by $f(B) = \det(AB)$. Then $f(B_1, \ldots, B_n) = \det(AB_1, \ldots, AB_n)$, because the columns of AB are $AB_1, \ldots, AB_n$. It is straightforward to check that f is an n-alternating form on $k^{(n)}$. Let $B = (b_{ij})_{n\times n}$. Then $B_j = \sum_{i=1}^{n} b_{ij}e_i$, $1 \leq j \leq n$, where $e_1, \ldots, e_n$ denotes the standard basis of $k^{(n)}$. Corollary 6.11 implies that

$$\begin{aligned}
\det(AB) &= f(B) = f(B_1, \ldots, B_n) \\
&= (\det B)f(e_1, \ldots, e_n) \\
&= (\det B)f(I_n) \\
&= (\det B)\det(AI_n) \\
&= (\det A)(\det B).
\end{aligned}$$

Proposition 6.16. *Let $A \in \mathcal{M}_{n\times n}(k)$. Then A is invertible if and only if $\det A \neq 0$.*

Proof. If A is invertible, then there exists a matrix $B \in \mathcal{M}_{n\times n}(k)$ such that $AB = I_n$. Then $(\det A)(\det B) = \det(AB) = \det I_n = 1$. Therefore, $\det A \neq 0$.

Now suppose that A is not invertible. Then Theorem 3.37 implies that there is a linear dependence relation $b_1A_1 + \cdots + b_nA_n = 0$ among the columns of A with some $b_i \neq 0$. We now use Proposition 6.3(4) to see that

$$\begin{aligned}
b_i \det A &= b_i \det(A_1, \ldots, A_{i-1}, A_i, A_{i+1}, \ldots A_n) \\
&= \det(A_1, \ldots, A_{i-1}, b_iA_i, A_{i+1}, \ldots A_n) \\
&= \det(A_1, \ldots, A_{i-1}, b_1A_1 + \cdots + b_nA_n, A_{i+1}, \ldots A_n) \\
&= \det(A_1, \ldots, A_{i-1}, 0, A_{i+1}, \ldots A_n) \\
&= 0.
\end{aligned}$$

Since $b_i \neq 0$, it follows that $\det A = 0$. □

Proposition 6.17 (Cramer's Rule)**.** *Let $A \in \mathcal{M}_{n\times n}(k)$. Consider the system of equations represented by $Ax = b$, where $x = (x_1, \ldots x_n)^t$ and $b = (b_1, \ldots b_n)^t \in k^{(n)}$.*

1. *If the system of equations has a solution, then*

$$x_i \det A = \det(A_1, \ldots, A_{i-1}, b, A_{i+1}, \ldots, A_n).$$

2. *If* $\det A \neq 0$*, then the system of equations has a unique solution given by*
$$x_i = \frac{\det(A_1, \ldots, A_{i-1}, b, A_{i+1}, \ldots, A_n)}{\det A}, \qquad 1 \leq i \leq n.$$

Proof. (1) The system of linear equations represented by $Ax = b$ is equivalent to the matrix equation $x_1A_1 + \cdots + x_nA_n = b$. This gives

$$\begin{aligned}\det(A_1, \ldots, A_{i-1}, b, &A_{i+1}, \ldots, A_n)\\ &= \det(A_1, \ldots, A_{i-1}, \sum_{j=1}^{n} x_jA_j, A_{i+1}, \ldots, A_n)\\ &= \det(A_1, \ldots, A_{i-1}, x_iA_i, A_{i+1}, \ldots, A_n)\\ &= x_i \det(A_1, \ldots, A_n) = x_i \det A.\end{aligned}$$

(2) If $\det A \neq 0$, then A is invertible by Proposition 6.16. Then there is a unique solution given by $x = A^{-1}b$. Since $\det A \neq 0$, the result in the first part gives the desired formula for x_i. □

Definition 6.18. *Let* $T \in \mathcal{L}(V, V)$*,* $\dim V = n$*. Let* $\beta = \{v_1, \ldots, v_n\}$ *be an ordered basis of* V*. The* determinant *of* T*,* $\det T$*, is defined to be* $\det[T]_\beta^\beta$.

Proposition 6.19. *The determinant of a linear transformation is well defined. That is,* $\det T$ *does not depend on the choice of the ordered basis* β*.*

Proof. Let β, γ be two ordered bases of V. Then $[T]_\gamma^\gamma = [1_V]_\beta^\gamma [T]_\beta^\beta [1_V]_\gamma^\beta$. Let $P = [1_V]_\gamma^\beta$, $A = [T]_\beta^\beta$, and $B = [T_V]_\gamma^\gamma$. Then $[1_V]_\beta^\gamma = P^{-1}$ since $[1_V]_\gamma^\beta [1_V]_\beta^\gamma = [1_V]_\beta^\beta = I_n$.

Thus $B = P^{-1}AP$ and so

$$\det B = \det(P^{-1})(\det A)(\det P) = \det(P^{-1})(\det P)(\det A) = \det A.$$

Therefore, $\det T$ is well defined. □

6.3 The Adjoint of a matrix

Definition 6.20. *Let* $A \in \mathcal{M}_{n\times n}(k)$*. If* $n \geq 2$*, the adjoint matrix of* A*, written* $\operatorname{Adj}(A)$*, is defined as the matrix* $(b_{ij}) \in \mathcal{M}_{n\times n}(k)$ *where* $b_{ij} = (-1)^{i+j}\det(A_{ji})$*. If* $n = 1$*, we define* $\operatorname{Adj}(A) = (1)$ *(even if* $A = 0$*).*

Note that the entries of $\mathrm{Adj}(A)$ are polynomial expressions of the entries of A.

The adjoint matrix of A is sometimes called the *classical adjoint* matrix of A. It is natural to wonder if there is a connection between the adjoint of A just defined and the adjoint of a linear transformation T defined in Chapter 5 in relation to a given inner product. There is a connection, but it requires more background than has been developed so far.

Lemma 6.21. *Let $A \in \mathcal{M}_{n\times n}(k)$. Then* $(\mathrm{Adj}(A))^t = \mathrm{Adj}(A^t)$.

Proof. The (i,j)-entry of $(\mathrm{Adj}(A))^t$ is the (j,i)-entry of $\mathrm{Adj}(A)$, which is $(-1)^{i+j}\det A_{ij}$. The (i,j)-entry of $\mathrm{Adj}(A^t)$ equals

$$(-1)^{i+j}\det((A^t)_{ji}) = (-1)^{i+j}\det((A_{ij})^t) = (-1)^{i+j}\det A_{ij}.$$

Therefore, $(\mathrm{Adj}(A))^t = \mathrm{Adj}(A^t)$. □

Proposition 6.22. *Let $A \in \mathcal{M}_{n\times n}(k)$.*

1. $\mathrm{Adj}(A)A = A\,\mathrm{Adj}(A) = (\det A)I_n$.

2. *If* $\det(A) \neq 0$, *then A is invertible and*

$$\mathrm{Adj}(A) = (\det A)A^{-1}.$$

Proof. (1) The (i,l)-entry of $A\,\mathrm{Adj}(A)$ equals

$$\sum_{j=1}^{n} a_{ij}b_{jl} = \sum_{j=1}^{n} a_{ij}(-1)^{j+l}\det A_{lj}.$$

If $i = l$, this expression equals $\det A$, because this is the formula for $\det A$ given in the proof of Theorem 6.10 if we expand along the i^{th} row. If $i \neq l$, then let A' be the matrix obtained from A by replacing the l^{th} row of A with the i^{th} row of A. Then $\det A' = 0$ by Proposition 6.14(1) because A has two rows that are equal. Thus, if we compute $\det A'$ by expanding along the l^{th} row, we get $0 = \sum_{j=1}^{n}(-1)^{l+j}a_{ij}\det A_{lj}$, because the (l,j)-entry of A' is a_{ij} and $(A')_{lj} = A_{lj}$. Therefore, the (i,l)-entry of $A\,\mathrm{Adj}(A)$ equals $\det A$ if $i = l$ and equals 0 if $i \neq l$. This implies that $A\,\mathrm{Adj}(A) = (\det A)I_n$.

Now we prove that $\mathrm{Adj}(A)A = (\det A)I_n$. We have

$$(\mathrm{Adj}(A)A)^t = A^t(\mathrm{Adj}(A))^t = A^t\,\mathrm{Adj}(A^t) = \det(A^t)I_n = (\det A)I_n.$$

Thus, $\mathrm{Adj}(A)A = ((\det A)I_n)^t = (\det A)I_n$.

(2) This is clear from (1) and Proposition 6.16. □

The equality $\operatorname{Adj}(A)A = (\det A)I_n$ in Proposition 6.22 (1) could also have been proved in a way similar to the proof of the equality $A\operatorname{Adj}(A) = (\det A)I_n$. In this case, we could have computed $\det A$ by expanding along the l^{th} column of A and at the appropriate time replace the l^{th} column of A with the j^{th} column of A. See Exercise 8.

Note that if $\det A \neq 0$, then the formula in Proposition 6.22 (1) could be used to give a different proof of the implication in Proposition 6.16 stating that A is invertible when $\det A \neq 0$.

Lemma 6.23. *Let $A, B \in \mathcal{M}_{n\times n}(k)$. Then $\operatorname{Adj}(AB) = \operatorname{Adj}(B)\operatorname{Adj}(A)$.*

Proof. First assume that A, B are both invertible. Then AB is invertible, so Proposition 6.22 implies that

$$\begin{aligned}\operatorname{Adj}(AB) &= \det(AB)(AB)^{-1} = \det(A)\det(B)B^{-1}A^{-1}\\ &= (\det(B)B^{-1})(\det(A)A^{-1}) = \operatorname{Adj}(B)\operatorname{Adj}(A).\end{aligned}$$

Now assume that $A, B \in \mathcal{M}_{n\times n}(k)$ are arbitrary. Let $k(x)$ denote the rational function field over k. Then $A + xI_n, B + xI_n \in \mathcal{M}_{n\times n}(k(x))$ are both invertible (because $\det(A + xI_n) = x^n + \cdots \neq 0$). Thus

$$\operatorname{Adj}((A + xI_n)(B + xI_n)) = \operatorname{Adj}(B + xI_n)\operatorname{Adj}(A + xI_n).$$

Since each entry in these matrices lies in $k[x]$, it follows that we may set $x = 0$ to obtain the result. □

If $A \in \mathcal{M}_{n\times n}(k)$, let $\operatorname{rk}(A)$ denote the rank of A.

Proposition 6.24. *Let $A \in \mathcal{M}_{n\times n}(k)$.*

1. *If $\operatorname{rk}(A) = n$, then $\operatorname{rk}(\operatorname{Adj}(A)) = n$.*
2. *If $\operatorname{rk}(A) = n - 1$, then $\operatorname{rk}(\operatorname{Adj}(A)) = 1$.*
3. *If $\operatorname{rk}(A) \leq n - 2$, then $\operatorname{Adj}(A) = 0$, so $\operatorname{rk}(\operatorname{Adj}(A)) = 0$.*

Proof. If $n = 1$, then (1) and (2) hold, while (3) is vacuous. Now assume that $n \geq 2$. If $\operatorname{rk}(A) = n$, then A is invertible. Thus $\operatorname{Adj}(A) = \det(A)A^{-1}$ has rank n. This proves (1). If $\operatorname{rk}(A) \leq n - 2$, then $\det(A_{ji}) = 0$ for each $(n-1)\times(n-1)$ submatrix A_{ji}. The definition of $\operatorname{Adj}(A)$ implies that $\operatorname{Adj}(A) = 0$, and so $\operatorname{rk}(\operatorname{Adj}(A)) = 0$. This proves (3).

Now assume that $\mathrm{rk}(A) = n-1$. For any matrix $A \in \mathcal{M}_{n\times n}(k)$, there exist invertible matrices $B, C \in \mathcal{M}_{n\times n}(k)$ such that BAC is a diagonal matrix. Let $D = BAC$. Then $\mathrm{rk}(D) = \mathrm{rk}(A)$. Lemma 6.23 implies that

$$\mathrm{Adj}(D) = \mathrm{Adj}(BAC) = \mathrm{Adj}(C)\,\mathrm{Adj}(A)\,\mathrm{Adj}(B).$$

Since $\mathrm{Adj}(B)$ and $\mathrm{Adj}(C)$ are invertible by (1), it follows that

$$\mathrm{rk}(\mathrm{Adj}(D)) = \mathrm{rk}(\mathrm{Adj}(A)).$$

For a diagonal matrix D, it is easy to check that if $\mathrm{rk}(D) = n-1$, then $\mathrm{rk}(\mathrm{Adj}(D)) = 1$. Thus the same folds for A and this proves (2). □

Proposition 6.25. *Let $A \in \mathcal{M}_{n\times n}(k)$.*

1. $\mathrm{Adj}(cA) = c^{n-1}\,\mathrm{Adj}(A)$ *for all $c \in k$.*
2. *If A is invertible, then* $\mathrm{Adj}(A^{-1}) = (\mathrm{Adj}(A))^{-1}$.
3. $\mathrm{Adj}(A^t) = (\mathrm{Adj}(A))^t$.
4. $\mathrm{Adj}(\mathrm{Adj}(A)) = (\det(A))^{n-2}A$ *for $n \geq 3$. If $n = 1$, this result holds when $A \neq 0$. If $n = 2$, then the result holds if we set* $(\det(A))^{n-2} = 1$, *including when* $\det(A) = 0$.

Proof. (1) follows from the definition of the adjoint matrix. For (2), we have

$$I_n = \mathrm{Adj}(I_n) = \mathrm{Adj}(AA^{-1}) = \mathrm{Adj}(A^{-1})\,\mathrm{Adj}(A).$$

Thus, $\mathrm{Adj}(A^{-1}) = (\mathrm{Adj}(A))^{-1}$.

Although we already proved (3) in Lemma 6.21, here is a different proof. First assume that A is invertible. Then

$$\begin{aligned} A^t\,\mathrm{Adj}(A^t) &= \det(A^t)I_n = \det(A)I_n \\ &= \mathrm{Adj}(A)A = (\mathrm{Adj}(A)A)^t = A^t(\mathrm{Adj}(A))^t. \end{aligned}$$

Since A^t is also invertible, we have $\mathrm{Adj}(A^t) = (\mathrm{Adj}(A))^t$.

Now assume that A is arbitrary. Then consider $A + xI_n$ over the field $k(x)$. Since $A + xI_n$ is invertible over $k(x)$, we have

$$\mathrm{Adj}(A^t + xI_n) = \mathrm{Adj}((A + xI_n)^t) = (\mathrm{Adj}(A + xI_n))^t\,.$$

Since all entries lie in $k[x]$, we may set $x = 0$ to obtain the result.

(4) If $n = 1$ and $A \neq 0$, then

$$\mathrm{Adj}(\mathrm{Adj}(A)) = (1) \text{ and } (\det(A))^{n-2}A = A/\det(A) = (1).$$

Now assume that $n \geq 2$. First assume that A is invertible. Then $\mathrm{Adj}(A)$ is invertible and we have

$$\begin{aligned}\mathrm{Adj}(\mathrm{Adj}(A)) &= \det(\mathrm{Adj}(A))(\mathrm{Adj}(A))^{-1} \\ &= \det(\det(A)A^{-1})\left(\det(A)A^{-1}\right)^{-1} \\ &= (\det(A))^n(1/\det(A))(1/\det(A))A = (\det(A))^{n-2}A.\end{aligned}$$

Now assume that A is not invertible and $n \geq 3$. Then $\det(A) = 0$ and so $(\det(A))^{n-2}A = 0$. Since $\mathrm{rk}(A) \leq n-1$, we have $\mathrm{rk}(\mathrm{Adj}(A)) \leq 1 \leq n-2$, by Proposition 6.24. Thus $\mathrm{Adj}(\mathrm{Adj}(A)) = 0$ by Proposition 6.24 again.

Now assume that $n = 2$ and A is not invertible. One checks directly that $\mathrm{Adj}(\mathrm{Adj}(A)) = A$ when $n = 2$. We have $(\det(A))^{n-2}A = A$ (because we set $(\det(A))^{n-2} = 1$, including when $\det(A) = 0$). □

6.4 The vector space of alternating forms

Let V be a vector space over a field k and assume that $\dim V = n$. For an integer $m \geq 1$, let $\mathrm{Mul}^m(V)$ denote the set of m-multilinear forms $f : V \times \cdots \times V \to k$. Then $\mathrm{Mul}^m(V)$ is a vector space over k and

$$\dim(\mathrm{Mul}^m(V)) = (\dim V)^m = n^m,$$

To see this, consider a basis $\beta = \{v_1, \ldots, v_n\}$ of V and let $f : V \times \cdots \times V \to k$ be an m-multilinear map. Then f is uniquely determined by

$$\{f(v_{i_1}, \ldots, v_{i_m}) \mid i_1, \ldots, i_m \in \{1, 2, \ldots, n\}\}.$$

Thus a basis of $\mathrm{Mul}^m(V)$ consists of n^m elements.

Let $\mathrm{Alt}^m(V)$ denote the subset of $\mathrm{Mul}^m(V)$ consisting of m-alternating forms $f : V \times \cdots \times V \to k$. Then $\mathrm{Alt}^m(V)$ is a subspace of $\mathrm{Mul}^m(V)$ over k. We will compute the dimension of $\mathrm{Alt}^m(V)$ below.

Proposition 6.26. *If $m \geq 1$, then $\dim(\mathrm{Alt}^m(V)) = \binom{n}{m}$. In particular, $\dim(\mathrm{Alt}^m(V)) = 0$ if $m > n$, and $\dim(\mathrm{Alt}^n(V)) = 1$.*

Proof. Let $\{v_1, \dots, v_n\}$ be a basis of V and let $f \in \mathrm{Alt}^m(V)$. Since $f \in \mathrm{Mul}^m(V)$, we know that f is uniquely determined by

$$\{f(v_{i_1}, \dots, v_{i_m}) \,|\, i_1, \dots, i_m \in \{1, 2, \dots, n\}\}.$$

Since $f \in \mathrm{Alt}^m(V)$, Proposition 6.3 (2) and (3) implies that f is uniquely determined by

$$\{f(v_{i_1}, \dots, v_{i_m}) \,|\, 1 \leq i_1 < i_2 < \cdots < i_m \leq n\}.$$

Thus $\mathrm{Alt}^m(V) = (0)$ if $m > n$. If $1 \leq m \leq n$, it follows that $\mathrm{Alt}^m(V)$ is spanned by $\binom{n}{m}$ elements. To show that these elements are linearly independent over k, it is sufficient to construct for each $1 \leq i_1 < i_2 < \cdots < i_m$ an element $g \in \mathrm{Alt}^m(V)$ such that $g(v_{i_1}, \cdots, v_{i_m}) = 1$ and $g(v_{j_1}, \cdots, v_{j_m}) = 0$ for all $j_1, \dots, j_m$ satisfying $1 \leq j_1 < j_2 \cdots < j_m$ and $\{j_1, \dots, j_m\} \neq \{i_1, \dots, i_m\}$.

We define g by

$$g(v_{j_1}, \cdots, v_{j_m}) = \begin{cases} (-1)^{\epsilon(\sigma)} & \text{if } (j_1, \dots, j_m) = (i_{\sigma(1)}, \dots, i_{\sigma(m)}) \\ 0 & \text{otherwise.} \end{cases}$$

□

6.5 The adjoint as a matrix of a linear transformation

In this section, we identify the adjoint of a matrix as the matrix of a particular linear transformation.

Let V, W be vector spaces over k and let $T : V \to W$ be a linear transformation. Then for each $m \geq 1$, there exists a linear transformation $T'_m : \mathrm{Alt}^m(W) \to \mathrm{Alt}^m(V)$ defined by $T'_m(g) = g \circ T$ where $g \in \mathrm{Alt}^m(W)$ and $(g \circ T)(u_1, \dots, u_m) = g(T(u_1), \dots, T(u_m))$ for $u_1, \dots, u_m \in V$. It is straightforward to check that $g \circ T \in \mathrm{Alt}^m(V)$.

Now let $U \overset{S}{\to} V \overset{T}{\to} W$ be two linear transformations. Then for each $m \geq 1$, we obtain linear transformations

$$\mathrm{Alt}^m(W) \overset{T'_m}{\to} \mathrm{Alt}^m(V) \overset{S'_m}{\to} \mathrm{Alt}^m(U).$$

Lemma 6.27. *With the notation above, we have*

$$(T \circ S)'_m = S'_m \circ T'_m.$$

Proof. Let $g \in \mathrm{Alt}^m(W)$. Then

$$(S'_m \circ T'_m)(g) = S'_m(T'_m(g)) = S'_m(g \circ T) = (g \circ T) \circ S = g \circ (T \circ S) = (T \circ S)'_m(g).$$

□

Assume now that $\dim(V) = \dim(W) = n$ and $m = n-1$. Let $T' = T'_{n-1}$ where $T' : \mathrm{Alt}^{n-1}(W) \to \mathrm{Alt}^{n-1}(V)$ is the linear transformation from above.

Let $\beta = \{v_1, \dots, v_n\}$ be a basis of V and let $\gamma = \{w_1, \dots, w_n\}$ be a basis of W. Let $T(v_j) = \sum_{i=1}^n a_{ij} w_i$, $1 \le j \le n$. Then $[T]_\beta^\gamma = (a_{ij})_{n \times n}$.

Let $\beta' = \{f_1, \dots, f_n\}$ be the basis of $\mathrm{Alt}^{n-1}(V)$ where

$$f_j(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n) = (-1)^{j-1}$$

and $f_j(v_{i_1}, \dots, v_{i_{n-1}}) = 0$ when $1 \le i_1 < \cdots < i_{n-1} \le n$ and $\{i_1, \dots, i_{n-1}\} \neq \{1, 2, \dots, j-1, j+1, \dots, n\}$.

Define $\gamma' = \{g_1, \dots, g_n\}$ to be a basis of $\mathrm{Alt}^{n-1}(W)$ with similar properties with respect to the basis γ.

Let $T'(g_j) = \sum_{i=1}^n c_{ij} f_i$, $1 \le j \le n$. Then $[T']_{\gamma'}^{\beta'} = (c_{ij})_{n \times n}$.

Proposition 6.28. *Using the notation above, we have* $[T']_{\gamma'}^{\beta'} = \mathrm{Adj}([T]_\beta^\gamma)$.

Proof. Let $A = [T]_\beta^\gamma$ and let $C = [T']_{\gamma'}^{\beta'}$. We now make two computations. First,

$$\begin{aligned}\left(\sum_{i=1}^n c_{ij} f_i\right)(v_1, \dots, v_{l-1}, v_{l+1}, \dots, v_n) &= c_{lj} f_l(v_1, \dots, v_{l-1}, v_{l+1}, \dots, v_n)\\ &= (-1)^{l-1} c_{lj}.\end{aligned}$$

Second,

$$\begin{aligned}
T'(g_j)(v_1, \dots, v_{l-1}, v_{l+1}, \dots, v_n) &= g_j(T(v_1), \dots, T(v_{l-1}), T(v_{l+1}), \dots, T(v_n)) \\
&= g_j\left(\sum_{i=1}^{n} a_{i1}v_i, \dots, \sum_{i=1}^{n} a_{i,l-1}v_i, \sum_{i=1}^{n} a_{i,l+1}v_i, \dots, \sum_{i=1}^{n} a_{in}v_i\right) \\
&= g_j\left(\sum_{\substack{i=1\\ i\neq j}}^{n} a_{i1}v_i, \dots, \sum_{\substack{i=1\\ i\neq j}}^{n} a_{i,l-1}v_i, \sum_{\substack{i=1\\ i\neq j}}^{n} a_{i,l+1}v_i, \dots, \sum_{\substack{i=1\\ i\neq j}}^{n} a_{in}v_i\right) \\
&= \det(A_{jl})g_j(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n) \\
&= \det(A_{jl})(-1)^{j-1}.
\end{aligned}$$

Since $T'(g_j) = \sum_{i=1}^{n} c_{ij}f_i$, it follows that

$$(-1)^{l-1}c_{lj} = (-1)^{j-1}\det(A_{jl}).$$

Thus $c_{lj} = (-1)^{l+j}(\det(A_{jl}))$. Therefore, $C = \operatorname{Adj}(A)$ and so $[T']_{\gamma'}^{\beta'} = \operatorname{Adj}([T]_{\beta}^{\gamma})$. □

Let $U \xrightarrow{S} V \xrightarrow{T} W$ be two linear transformations and assume that

$$\dim(U) = \dim(V) = \dim(W) = n.$$

Let β, β', γ, γ' be as before. Let α be a basis of U and let α' be the corresponding basis of $\operatorname{Alt}^{n-1}(U)$.

Let $A = [T]_{\beta}^{\gamma}$ and $B = [S]_{\alpha}^{\beta}$.

Proposition 6.29. *Using the notations above, we have*

$$\operatorname{Adj}(AB) = \operatorname{Adj}(B)\operatorname{Adj}(A).$$

Proof.

$$\begin{aligned}
\operatorname{Adj}(AB) &= \operatorname{Adj}([T]_{\beta}^{\gamma}[S]_{\alpha}^{\beta}) = \operatorname{Adj}([T \circ S]_{\alpha}^{\gamma}) \\
&= [(T \circ S)']_{\gamma'}^{\alpha'} = [S' \circ T']_{\gamma'}^{\alpha'} = [S']_{\beta'}^{\alpha'}[T']_{\gamma'}^{\beta'} \\
&= \operatorname{Adj}([S]_{\alpha}^{\beta})\operatorname{Adj}([T]_{\beta}^{\gamma}) = \operatorname{Adj}(B)\operatorname{Adj}(A).
\end{aligned}$$

□

6.6 The Vandermonde Determinant and Applications

Definition 6.30. *Let* $a_1, \dots, a_n \in k$. *The* Vandermonde determinant *of* $a_1, \dots, a_n$, *written* $V(a_1, \dots, a_n)$, *is defined to be* $\det A$ *where*

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ & & \vdots & \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{pmatrix}_{n\times n}.$$

The matrix A *is called the Vandermonde matrix of* $a_1, \dots, a_n$.

Proposition 6.31.

$$V(a_1, \dots, a_n) = \prod_{1\le i<j\le n} (a_j - a_i).$$

Proof. We prove the result by induction on n. For $n = 1$ the result holds since an empty product is defined to be 1 and $\det(1) = 1$. For $n = 2$, the result holds since

$$\det \begin{pmatrix} 1 & 1 \\ a_1 & a_2 \end{pmatrix} = a_2 - a_1.$$

Now assume that $n \geq 3$.
replace row n by (row n - a_1·row $(n-1)$),
replace row $(n-1)$ by (row $(n-1)$ $-a_1$·row $(n-2)$),

$$\vdots$$

replace row 2 by (row 2 $-a_1$·row 1).

These row operations do not change $V(a_1, \dots, a_n)$ by Proposition 6.3(4) since det is an alternating function on the rows of A. Therefore,

$$V(a_1, \dots, a_n) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & \cdots & a_n - a_1 \\ 0 & a_2(a_2 - a_1) & \cdots & a_n(a_n - a_1) \\ & & \vdots & \\ 0 & a_2^{n-2}(a_2 - a_1) & \cdots & a_n^{n-2}(a_n - a_1) \end{pmatrix}_{n\times n}.$$

Now we expand the determinant along the first column to obtain

$$V(a_1, \ldots, a_n) = \det \begin{pmatrix} a_2 - a_1 & \cdots & a_n - a_1 \\ a_2(a_2 - a_1) & \cdots & a_n(a_n - a_1) \\ & \vdots & \\ a_2^{n-2}(a_2 - a_1) & \cdots & a_n^{n-2}(a_n - a_1) \end{pmatrix}_{(n-1)\times(n-1)}.$$

Since det is multilinear on the columns, we can factor out common factors from entries in each column to obtain

$$V(a_1, \ldots, a_n) = (a_2 - a_1) \cdots (a_n - a_1) \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_2 & a_3 & \cdots & a_n \\ a_2^2 & a_3^2 & \cdots & a_n^2 \\ & & \vdots & \\ a_2^{n-2} & a_3^{n-2} & \cdots & a_n^{n-2} \end{pmatrix}.$$

By induction on n, we conclude now that

$$V(a_1, \ldots, a_n) = \prod_{1 \leq i < j \leq n} (a_j - a_i).$$

□

We will now apply the Vandermonde determinant and related ideas to the problem of finding polynomials in one variable which pass through given points in the affine plane.

For each integer $n \geq 1$, let V_n denote the vector space of polynomials in $k[x]$ of degree $\leq n$, including the zero polynomial. Thus

$$V_n = \{g \in k[x] \mid g = c_0 + c_1 x + \cdots + c_n x^n\},$$

where $c_i \in k$. Then $\beta = \{1, x, x^2, \ldots, x^n\}$ is a basis of V_n. It is easy to check that $V_n \cong k^{(n+1)}$ by the isomorphism $\phi : V_n \to k^{(n+1)}$, where $\sum_{j=0}^{n} c_j x^j \mapsto (c_0, \ldots, c_n)$. Thus ϕ maps β to the standard basis of $k^{(n+1)}$.

We will shortly construct some other special bases of V_n.

Proposition 6.32. *Let $a_0, a_1, \ldots, a_n$ be distinct elements in k and suppose that $b_0, b_1, \ldots, b_n$ are arbitrary elements in k. Then there is a unique polynomial $g \in V_n$ such that $g(a_i) = b_i$, $0 \leq i \leq n$.*

Proof. We will give two proofs of this result.

(1) The conditions $g(a_i) = b_i$, $0 \leq i \leq n$, are equivalent to finding $c_0, c_1, \dots, c_n \in k$ such that $g(x) = c_0 + c_1 x + \cdots + c_n x^n$ and $\sum_{j=0}^{n} c_j a_i^j = b_i$, $0 \leq i \leq n$. Let

$$A = \begin{pmatrix} 1 & a_0 & a_0^2 & \cdots & a_0^n \\ 1 & a_1 & a_1^2 & \cdots & a_1^n \\ & & \vdots & & \\ 1 & a_n & a_n^2 & \cdots & a_n^n \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

The stated conditions are equivalent to finding c such that $Ac = b$. Propositions 6.31 and 6.13 imply that $\det A \neq 0$ because the a_i's are distinct. Proposition 6.16 implies that A is invertible and therefore $c = A^{-1}b$ is the unique solution to the system of equations. This implies that $g(x) = c_0 + c_1 x + \cdots + c_n x^n$ is the unique polynomial in V_n such that $g(a_i) = b_i$, $0 \leq i \leq n$.

(2) Let

$$f_{j,a_0,\dots,a_n}(x) = \frac{(x - a_0)\cdots(x - a_{j-1})(x - a_{j+1})\cdots(x - a_n)}{(a_j - a_0)\cdots(a_j - a_{j-1})(a_j - a_{j+1})\cdots(a_j - a_n)}, \quad 0 \leq j \leq n.$$

Then $f_{j,a_0,\dots,a_n}(x) \in V_n$ because $\deg f_{j,a_0,\dots,a_n}(x) = n$. We have

$$f_{j,a_0,\dots,a_n}(a_i) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

Let $g(x) = \sum_{j=0}^{n} b_j f_{j,a_0,\dots,a_n}(x)$. Then $g(a_i) = b_i$, $0 \leq i \leq n$, and $g \in V_n$.

Suppose that $h \in V_n$ and $h(a_i) = b_i$, $0 \leq i \leq n$. Then $(g - h)(a_i) = 0$, $0 \leq i \leq n$. This implies that $g - h \in V_n$ and $g - h$ has $n + 1$ distinct zeros. A theorem from Algebra implies that $g - h = 0$ and so $g = h$. This proves the existence and uniqueness of g. □

The first proof of Proposition 6.32 depended on knowledge of the Vandermonde determinant, while the second proof used no linear algebra, but used a theorem from Algebra on polynomials. We will now develop a third approach to Proposition 6.32 and the Vandermonde determinant using the polynomials $f_{j,a_0,\dots,a_n}(x)$ from above.

Proposition 6.33. *1. Let $a \in k$ and consider the function $a^* : V_n \to k$, defined by $a^*(h) = h(a)$. Then $a^* \in V_n^*$.*

2. *Let $a_0, \dots, a_n$ be distinct elements in k. Then*

$$\gamma = \{f_{j,a_0,\dots,a_n}(x) \,|0 \le j \le n\}$$

is a basis of V_n. The set $\gamma^* = \{a_0^*, a_1^*, \dots, a_n^*\}$ is a dual basis of V_n^* to γ.

Proof. (1) Let $g, h \in V_n$ and let $c \in k$. Then $a^*(g+h) = (g+h)(a) = g(a) + h(a) = a^*(g) + a^*(h)$, and $a^*(ch) = (ch)(a) = ch(a) = ca^*(h)$. Therefore $a^* \in V_n^*$.

(2) We first show that γ is a linearly independent set in V_n. Suppose that $\sum_{j=0}^n b_j f_{j,a_0,\dots,a_n}(x) = 0$. Since

$$a_i^*(f_{j,a_0,\dots,a_n}(x)) = f_{j,a_0,\dots,a_n}(a_i) = \begin{cases} 0 & \text{if } i \ne j \\ 1 & \text{if } i = j \end{cases},$$

we apply a_i^* to both sides to obtain $b_i f_{i,a_0,\dots,a_n}(a_i) = 0$. Thus, $b_i = 0$, $0 \le i \le n$, and so γ is a linearly independent set in V_n. Since $\dim V_n = n+1$, it follows that γ is a basis of V_n. Therefore, γ^* is a dual basis of V_n^* to γ by Proposition 4.2. □

Proposition 6.34. *Let $a_0, a_1, \dots, a_n$ be distinct elements of k.*

1. *If $g(x) \in V_n$, then $g(x) = \sum_{j=0}^n g(a_j) f_{j,a_0,\dots,a_n}(x)$.*

2. *Suppose that $b_0, b_1, \dots, b_n$ are arbitrary elements in k. Then there is a unique polynomial $g \in V_n$ such that $g(a_i) = b_i$, $0 \le i \le n$.*

3. *$V(a_0, a_1, \dots, a_n) \ne 0$.*

Proof. (1) Let $g(x) \in V_n$. Then $g(x) = \sum_{j=0}^n c_j f_{j,a_0,\dots,a_n}(x)$ where each $c_j \in k$, because γ is a basis of V_n. For $0 \le i \le n$, we have

$$g(a_i) = a_i^*(g(x)) = a_i^*\left(\sum_{j=0}^n c_j f_{j,a_0,\dots,a_n}(x)\right) = c_i f_{i,a_0,\dots,a_n}(a_i) = c_i.$$

(2) We saw above that $g(x) = \sum_{j=0}^n b_j f_{j,a_0,\dots,a_n}(x)$ has the property that $g(a_i) = b_i$, $0 \le i \le n$. Now suppose that $h \in V_n$ and $h(a_i) = b_i$, $0 \le i \le n$. Let $h(x) = \sum_{j=0}^n c_j f_{j,a_0,\dots,a_n}(x)$. Then the proof of (1) implies that $c_i =$

$h(a_i) = b_i$ for $0 \leq i \leq n$. Therefore the polynomial $h = g$, so g is uniquely determined.

(3) Apply (1) to the polynomials $g(x) = x^j$, $0 \leq j \leq n$. Then

$$x^j = \sum_{i=0}^{n} a_i^j f_{i,a_0,\ldots,a_n}(x), \;\; 0 \leq j \leq n.$$

Let

$$A = \begin{pmatrix} 1 & a_0 & a_0^2 & \cdots & a_0^n \\ 1 & a_1 & a_1^2 & \cdots & a_1^n \\ & & \vdots & & \\ 1 & a_n & a_n^2 & \cdots & a_n^n \end{pmatrix}.$$

Then $A = [1_{V_n}]_\beta^\gamma$. Therefore, A is an invertible matrix, being a change of basis matrix. In particular, $V(a_0, a_1, \ldots, a_n) = \det A \neq 0$, by Proposition 6.16. □

The next application of Vandermonde matrices is to the linear independence of exponential functions.

Proposition 6.35. *Let $a_1, \ldots, a_n$ be distinct complex numbers. Then the functions $e^{a_1x}, \ldots, e^{a_nx}$ are linearly independent over* $\mathbf{C}$.

Proof. Suppose that $c_1e^{a_1x} + \cdots + c_ne^{a_nx} = 0$, where $c_i \in \mathbf{C}$. Differentiate this equation $n - 1$ times to obtain the system of equations

$$\begin{array}{rcrl} c_1e^{a_1x} & + \cdots + & c_ne^{a_nx} & = 0 \\ a_1c_1e^{a_1x} & + \cdots + & a_nc_ne^{a_nx} & = 0 \\ a_1^2c_1e^{a_1x} & + \cdots + & a_n^2c_ne^{a_nx} & = 0 \\ & \vdots & & \\ a_1^{n-1}c_1e^{a_1x} & + \cdots + & a_n^{n-1}c_ne^{a_nx} & = 0. \end{array}$$

This is equivalent to

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ & & \cdots & \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{pmatrix} \begin{pmatrix} c_1e^{a_1x} \\ c_2e^{a_2x} \\ \\ \vdots \\ c_ne^{a_nx} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \\ \vdots \\ 0 \end{pmatrix}.$$

Since $a_1, \ldots, a_n$ are distinct, it follows that the Vandermonde determinant $V(a_1, \ldots, a_n) \neq 0$. Therefore, the coefficient matrix is invertible and so we must have $c_1 e^{a_1 x} = \cdots = c_n e^{a_n x} = 0$. Since $e^{a_i x} \neq 0$, this implies $c_1 = \cdots = c_n = 0$. Therefore $e^{a_1 x}, \ldots, e^{a_n x}$ are linearly independent functions over $\mathbf{C}$. □

Proposition 6.35 can be strengthened as follows.

Proposition 6.36. *Let $a_1, \ldots, a_n$ be distinct complex numbers. Then the functions $e^{a_1 x}, \ldots, e^{a_n x}$ are linearly independent over $\mathbf{C}(x)$, the field of rational functions in x. ($\mathbf{C}(x) = \{g(x)/h(x) \mid g, h \in \mathbf{C}[x]\}$).*

Proof. Suppose that $e^{a_1 x}, \ldots, e^{a_n x}$ are linearly dependent over $\mathbf{C}(x)$. Then there exists an equation

$$\sum_{i=1}^{n} f_i(x) e^{a_i x} = 0,$$

where each $f_i(x) \in \mathbf{C}(x)$ and at least one $f_i(x) \neq 0$. After multiplying by a common denominator of the $f_i(x)$'s, we may assume that each $f_i(x) \in k[x]$.

Of all such equations, choose one with the minimal number of nonzero coefficients. Then after relabeling, we may assume that $\sum_{i=1}^{n} f_i(x) e^{a_i x} = 0$, where each $f_i(x) \in k[x]$ is nonzero and no nontrivial linear dependence relation exists with fewer summands. Of all such linear dependence relations, choose one where $\sum_{i=1}^{n} \deg f_i(x)$ is minimal.

Clearly $n \geq 2$. Differentiate the equation to obtain

$$\sum_{i=1}^{n} (a_i f_i(x) + f_i'(x)) e^{a_i x} = 0.$$

From this equation, subtract the equation $a_1 \sum_{i=1}^{n} f_i(x) e^{a_i x} = 0$. This gives

$$f_1'(x) e^{a_1 x} + \sum_{i=2}^{n} \left((a_i - a_1) f_i(x) + f_i'(x)\right) e^{a_i x} = 0.$$

If this is a nontrivial linear dependence relation, then each coefficient must be nonzero from our assumptions. But we now show that the sum of the degrees of the coefficients has decreased, which is a contradiction. If $i \geq 2$, then we have $\deg((a_i - a_1) f_i(x) + f_i'(x)) = \deg f_i(x)$ because $a_i - a_1 \neq 0$ and $\deg f_i'(x) < f_i(x)$ (since $f_i(x) \neq 0$). For $i = 1$ we have $\deg f_1'(x) < f_1(x)$.

Therefore, we must have a trivial linear dependence relation. This implies that $(a_i - a_1)f_i(x) + f_i'(x) = 0$ for $i \geq 2$. This is also impossible because $(a_i - a_1)f_i(x)$ is nonzero and has degree greater than $f_i'(x)$.

Therefore $e^{a_1 x}, \ldots, e^{a_n x}$ are linearly independent over $\mathbf{C}(x)$. $\square$

Exercises

1. Let $A = (a_{ij})_{n\times n}$. Define
$$f(A) = f(A_1, \ldots, A_n) = \sum_{\sigma \in S_n} (-1)^{\epsilon(\sigma)} a_{\sigma(1)1} \cdots a_{\sigma(n)n}.$$
Prove directly (without use of Theorem 6.10, for example) that f is an n-alternating form on $k^{(n)}$ such that $f(e_1, \ldots, e_n) = 1$. (Assume the fact that $\epsilon(\sigma)$ is well defined modulo 2.)

2. Compute the determinant of an arbitrary 2×2 and 3×3 matrix.

3. Suppose that $A \in \mathcal{M}_{n\times n}(k)$ with $A = (a_{ij})_{n\times n}$. If A is either a diagonal matrix, an upper triangular matrix, or a lower triangular matrix, then $\det A = a_{11}a_{22}\cdots a_{nn}$. (Try to find several solutions. Methods could rely on Problem 3, or Proposition 6.9 or Theorem 6.10, for example.)

4. Let $A \in \mathcal{M}_{n\times n}(k)$. Suppose that A has the following block form.
$$A = \begin{pmatrix} B_{r\times r} & C_{r\times s} \\ 0_{s\times r} & D_{s\times s} \end{pmatrix}$$
Then $\det A = (\det B)(\det D)$.

5. Consider the situation in Proposition 6.17 and assume that $\det A = 0$ in each of the following parts.
   (a) If $Ax = b$ has a solution, then $\det(A_1, \ldots, A_{i-1}, b, A_{i+1}, \ldots, A_n) = 0$, $1 \leq i \leq n$.
   (b) Suppose that $\det(A_1, \ldots, A_{i-1}, b, A_{i+1}, \ldots, A_n) = 0$, $1 \leq i \leq n$, and dim(column space of A) $= n - 1$. Then the system $Ax = b$ has a solution.
   (c) If $Ax = b$ has a solution and k is infinite, then there are infinitely many solutions to the system.

6. Let $A \in \mathcal{M}_{n\times n}(k)$ and let $A = (a_{ij})_{n\times n}$. The *trace* of A is defined by $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$. Verify the following statements for $A, B \in \mathcal{M}_{n\times n}(k)$ and $c \in k$.

   (a) $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$

   (b) $\operatorname{tr}(cA) = c \operatorname{tr} A$

   (c) $\operatorname{tr} : \mathcal{M}_{n\times n}(k) \to k$ is a linear transformation.

   (d) $\dim(\ker(\operatorname{tr})) = n^2 - 1$.

   (e) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$

   (f) If B is invertible, then $\operatorname{tr}(B^{-1}AB) = \operatorname{tr} A$.

7. Let $T \in \mathcal{L}(V, V)$, $\dim V = n$. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of V. The *trace* of T, $\operatorname{tr} T$, is defined to be $\operatorname{tr}[T]_{\beta}^{\beta}$. Then $\operatorname{tr} T$ is well defined. That is, $\operatorname{tr} T$ does not depend on the choice of the ordered basis β.

8. Prove that $\operatorname{Adj}(A)A = (\det A)I_n$ using the ideas in the proof of Proposition 6.22, but in this case expand along the l^{th} column of A and at the appropriate time replace the l^{th} column of A with the j^{th} column of A.

Chapter 7

Canonical Forms of a Linear Transformation

7.1 Preliminaries

Let V be a vector space over k. We do not assume in this chapter that V is finite dimensional over k, unless it is explicitly stated. Let $T : V \to V$ be a linear transformation.

We let T^i denote the linear transformation $T \circ T \circ \cdots \circ T$, where T is composed with itself i times. Let $T^0 = 1_V$. If

$$f(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + x_0 \in k[x],$$

then $f(T)$ denotes the linear transformation

$$a_m T^m + a_{m-1}T^{m-1} + \cdots + a_1 T + a_0 1_V.$$

Let $A \in \mathcal{M}_{n\times n}(k)$. The *characteristic polynomial of* A is defined to be the polynomial $\det(xI_n - A)$ that lies in $k[x]$. The methods in Chapter 6 for computing a determinant show that $\det(xI_n - A)$ is a monic polynomial in $k[x]$ of degree n. (A polynomial in $k[x]$ of degree n is monic if the coefficient of x^n is 1). Thus the characteristic polynomial of A is a monic polynomial of degree n. We let $f_A(x)$ denote the characteristic polynomial of A and so $f_A(x) = \det(x1_n - A)$.

Suppose that $\dim V = n$. We define the characteristic polynomial of T to be $\det(x1_V - T)$ by setting $\det(x1_V - T) = \det(xI_n - A)$ where $A = [T]_\beta^\beta$ and β is a basis of V. The characteristic polynomial of T is independent of the choice of basis of V by Proposition 6.19. We let $f_T(x)$ denote the characteristic polynomial of T. Thus $f_T(x) = \det(x1_V - T)$.

7.2 The Cayley-Hamilton Theorem

Let V be a vector space over k and let $T : V \to V$ be a linear transformation.

Proposition 7.1. *If* $\dim V$ *is finite, then there exists a nonzero polynomial* $g \in k[x]$ *such that* $g(T) = 0$.

Proof. Assume that $\dim V = n$. Then the $n^2 + 1$ linear transformations $1_V, T, T^2, T^3, \ldots, T^{n^2}$ are linearly dependent over k since $\dim(\mathcal{L}(V, V)) = n^2$. Thus there exist $a_0, a_1, \ldots, a_{n^2} \in k$, not all zero, such that

$$a_{n^2}T^{n^2} + \cdots + a_1T + a_0 1_V = 0.$$

Then $g(T) = 0$ where $g(x) = a_{n^2}x^{n^2} + \cdots + a_1x + a_0$ is a nonzero polynomial. □

Proposition 7.1 shows that we can choose g such that $\deg g \leq n^2$. A stronger theorem, known as the Cayley-Hamilton Theorem, states that there exists a polynomial g with $\deg g \leq n$ such that $g(T) = 0$. We prove the Cayley-Hamilton Theorem in Theorem 7.7 below.

Definition 7.2. *Let* $v \in V$. *We define the* T*-cyclic subspace of* V *generated by* v *to be the subspace* $k[T]v = \{h(T)v \mid h \in k[x]\}$. *If* $V = k[T]v$, *then* v *is called a cyclic vector for* T.

Thus $k[T]v$ is the subspace of V spanned by $\{v, Tv, T^2v, \ldots\}$.

Lemma 7.3. *Suppose that* $\dim V = n \geq 1$ *and that* $V = k[T]v$ *for some* $v \in V$. *Then* $\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\}$ *is a basis of* V.

Proof. Suppose that $\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\}$ is a linearly dependent set over k. Then there is a dependence relation

$$a_iT^i(v) + a_{i-1}T^{i-1}(v) + \cdots + a_1T(v) + a_0v = 0$$

where $a_i \neq 0$ and $1 \leq i \leq n-1$. We can divide by a_i so that we may assume from the start that $a_i = 1$. Then

$$T^i(v) \in \text{Span}(\{v, T(v), T^2(v), \ldots, T^{i-1}(v)\}).$$

It follows that $T^j(v) \in \mathrm{Span}(\{v, T(v), T^2(v), \ldots, T^{i-1}(v)\})$ for all $j \geq i$. Thus $k[T]v = \mathrm{Span}(\{v, T(v), T^2(v), \ldots, T^{i-1}(v)\})$, and so $n = \dim k[T]v \leq i$. This is impossible because $i \leq n-1$. Thus $\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\}$ is a linearly independent set over k. Since $\dim V = n$, it follows that this set is a basis of V. □

A subspace $W \subseteq V$ is called a T-invariant subspace of V if $T(W) \subseteq W$. For example, it is easy to see that $k[T]v$ is a T-invariant subspace of V.

Let W be a T-invariant subspace of V and assume that $\dim V = n$. Then T induces a linear transformation $T|_W : W \to W$. Let $\beta_1 = \{v_1, \ldots, v_m\}$ be a basis of W. Extend β_1 to a basis $\beta = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$ of V. Let $[T|_W]_{\beta_1}^{\beta_1} = A$, where $A \in \mathcal{M}_{m\times m}(k)$. Then

$$[T]_\beta^\beta = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

where $B \in \mathcal{M}_{m\times(n-m)}(k)$, $C \in \mathcal{M}_{(n-m)\times(n-m)}(k)$, and $0 \in \mathcal{M}_{(n-m)\times m}(k)$. Let $f_{T|_W}$ denote the characteristic polynomial of $T|_W$. Thus $f_{T|_W} = \det(x1_W - T|_W)$.

Lemma 7.4. *Suppose that $V = V_1 \oplus V_2$ where V_1 and V_2 are each T-invariant subspaces of V. Let $T_1 = T|_{V_1}$ and $T_2 = T|_{V_2}$. Suppose that $\gamma = \{w_1, \ldots, w_l\}$ is a basis of V_1 and that $\delta = \{y_1, \ldots, y_m\}$ is a basis of V_2. Let $\beta = \gamma \cup \delta = \{w_1, \ldots, w_l, y_1, \ldots, y_m\}$. Then*

$$[T]_\beta^\beta = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

where $A = [T_1]_\gamma^\gamma$ and $B = [T_2]_\delta^\delta$.

Proof. We have that $T(w_j) = \sum_{i=1}^{l} a_{ij} w_i$ and $T(y_j) = \sum_{i=1}^{m} b_{ij} y_i$. Then $A = (a_{ij})_{l\times l} = [T_1]_\gamma^\gamma$ and $B = (b_{ij})_{m\times m} = [T_2]_\delta^\delta$. □

Lemma 7.5. *Let V be a finite dimensional vector space over k and let W be a T-invariant subspace of V. Then $f_{T|_W}$ divides f_T.*

Proof. As above, we select a basis β_1 of W and extend β_1 to a basis β of V so that $\beta_1 \subseteq \beta$. Then

$$\begin{aligned} f_T(x) &= \det\left(xI_n - \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}\right) = \det\begin{pmatrix} xI_m - A & -B \\ 0 & xI_{n-m} - C \end{pmatrix} \\ &= \det(xI_m - A)\det(xI_{n-m} - C) = f_{T|_W}\det(xI_{n-m} - C). \end{aligned}$$

Thus $f_{T|_W}$ divides f_T. □

For $n \geq 1$, let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in k[x]$. For $n \geq 2$, the *companion matrix* of f is defined to be the $n \times n$ matrix

$$A_f = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -a_2 \\ 0 & 0 & 1 & \cdots & 0 & 0 & -a_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix}.$$

For $n = 1$ and $f(x) = x + a_0$, we let $A_f = (-a_0)$.

Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V. For $n \geq 2$, let $T : V \to V$ be the linear transformation defined by $T(v_i) = v_{i+1}$ for $1 \leq i \leq n-1$, and

$$T(v_n) = -a_0v_1 - a_1v_2 - \cdots - a_{n-2}v_{n-1} - a_{n-1}v_n.$$

For $n = 1$, we define T by $T(v_1) = -a_0v_1$. Then $[T]_\beta^\beta = A_f$. Note that for $n \geq 2$, we have $T^i(v_1) = v_{i+1}$ for $1 \leq i \leq n-1$.

We now show that $f(T)(v_1) = 0$.

$$\begin{aligned} f(T)(v_1) &= T^n(v_1) + a_{n-1}T^{n-1}(v_1) + \cdots + a_1T(v_1) + a_0v_1 \\ &= T(T^{n-1}(v_1)) + a_{n-1}v_n + a_{n-2}v_{n-1} + \cdots + a_1v_2 + a_0v_1 \\ &= T(v_n) + a_{n-1}v_n + a_{n-2}v_{n-1} + \cdots + a_1v_2 + a_0v_1 \\ &= 0, \end{aligned}$$

because $T(v_n) = -a_0v_1 - a_1v_2 - \cdots - a_{n-2}v_{n-1} - a_{n-1}v_n$.

Proposition 7.6. *Let $T : V \to V$ be the linear transformation defined above. Then the characteristic polynomial of T is f.*

Proof. Since $[T]_\beta^\beta = A_f$, we must show that $\det(xI_n - A_f) = f$. The proof is by induction on n. If $n = 1$, then

$$\det(xI_1 - A_f) = \det(xI_1 + a_0) = x + a_0 = f.$$

If $n = 2$, then

$$\det(xI_2 - A_f) = \det \begin{pmatrix} x & a_0 \\ -1 & x + a_1 \end{pmatrix} = x^2 + a_1x + a_0 = f.$$

Now assume that $n \geq 3$. We have

$$xI_n - A_f = \begin{pmatrix} x & 0 & 0 & \cdots & 0 & 0 & a_0 \\ -1 & x & 0 & \cdots & 0 & 0 & a_1 \\ 0 & -1 & x & \cdots & 0 & 0 & a_2 \\ 0 & 0 & -1 & \cdots & 0 & 0 & a_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & x & a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & -1 & x + a_{n-1} \end{pmatrix}.$$

Expanding along the first row and using induction gives

$$\begin{aligned} \det(xI_n - A_f) &= x(x^{n-1} + a_{n-1}x^{n-2} + \cdots + a_2 x + a_1) + (-1)^{n+1}a_0 \cdot (-1)^{n-1} \\ &= x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = f. \end{aligned}$$

□

Theorem 7.7 (Cayley-Hamilton Theorem). *Let V be a finite dimensional vector space over k and let $T : V \to V$ be a linear transformation. Let $f_T(x)$ be the characteristic polynomial of T. Then $f_T(T) = 0$ as a linear transformation.*

Proof. It is sufficient to show that $f_T(T)(v) = 0$ for each $v \in V$. This is clearly true if $v = 0$. Now assume that $v \neq 0$. Let $W = k[T]v$ be the cyclic subspace generated by v. Since W is a T-invariant subspace of V, we know that $f_{T|_W}$ divides f_T. It is sufficient to show that $f_{T|_W}(T)(v) = 0$. Thus we can assume from the beginning that $V = k[T]v$. Let $\dim V = n$. Then $\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\}$ is a basis of V by Lemma 7.3.

Let $T^n(v) = -a_0 v - a_1 T(v) - \cdots - a_{n-1}T^{n-1}(v)$ and let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Then $f_T(x) = f$ and it follows that $f_T(T)(v) = f(T)(v) = 0$, as desired. □

7.3 Eigenvectors, eigenvalues, diagonalizability

Let $T : V \to V$ be a linear transformation.

Definition 7.8. *An element $a \in k$ is an eigenvalue of T if there exists a nonzero vector $v \in V$ such that $T(v) = av$. A nonzero vector $v \in V$ is an eigenvector of T with eigenvalue a, $a \in k$, if $T(v) = av$.*

Lemma 7.9. *An element $a \in k$ is an eigenvalue of T if and only if $\ker(T - aI_V)$ is nonzero. The nonzero elements of $\ker(T - aI_V)$ are the set of eigenvectors of T having eigenvalue a.*

Proof. A vector $v \in V$ satisfies $T(v) = av$ if and only if $(T - aI_V)(v) = 0$, which is equivalent to $v \in \ker(T - aI_V)$. Thus an element $a \in k$ is an eigenvalue of T if and only if $\ker(T - aI_V)$ is nonzero. The second statement follows easily from this. □

Proposition 7.10. *Assume that $\dim V$ is finite and let $f_T(x) = \det(x1_V - T)$ be the characteristic polynomial of T. An element $a \in k$ is an eigenvalue of T if and only if $f_T(a) = 0$.*

Proof. Assume first that $a \in k$ and $f_T(a) = 0$. Then $\det(a1_V - T) = 0$, so $T - a1_V$ is not an invertible linear transformation. (See Exercise 1.) Thus $\ker(T - a1_V)$ is nonzero, so a is an eigenvalue of T.

Now assume that $a \in k$ is an eigenvalue of T. Then $\ker(T - a1_V)$ is nonzero. Thus $T - a1_V$ is not an invertible linear transformation, and so $\det(a1_V - T) = 0$. Therefore, $f_T(a) = 0$. □

Definition 7.11. *A matrix $A \in \mathcal{M}_{n\times n}(k)$ is diagonalizable (or, can be diagonalized) if there exists $C \in \mathcal{M}_{n\times n}(k)$ such that $C^{-1}AC$ is a diagonal matrix.*

If $\dim V$ is finite, a linear transformation $T \in \mathcal{L}(V, V)$ is diagonalizable (or, can be diagonalized) if there exists a basis β of V such that $[T]_\beta^\beta$ is a diagonal matrix.

If β is an arbitrary basis of V, $\dim V$ finite, then T is diagonalizable if and only if $[T]_\beta^\beta$ is diagonalizable. (See Exercise 2.)

Proposition 7.12. *Let $T : V \to V$ be a linear transformation, $\dim V$ finite. The following statements are equivalent.*

1. *T is diagonalizable.*

2. *There exists a basis of V consisting of eigenvectors of T.*

Proof. Assume that T is diagonalizable. Then there exists a basis $\beta = \{v_1, \ldots, v_n\}$ of V such that $[T]_\beta^\beta$ is a diagonal matrix $A = (a_{ij})$. Then $T(v_i) = a_{ii}v_i$ for $1 \le i \le n$. Thus β is a basis of V consisting of eigenvectors of T.

Now assume that there exists a basis $\beta = \{v_1, \ldots, v_n\}$ of V consisting of eigenvectors of T. Let $T(v_i) = c_i v_i$ for $1 \leq i \leq n$. Then $[T]_\beta^\beta = (a_{ij})$ where $a_{ij} = 0$ for $i \neq j$ and $a_{ij} = c_i$ if $i = j$. Thus $[T]_\beta^\beta$ is a diagonal matrix, so T is diagonalizable. □

If $T : V \to V$ is a linear transformation with $\dim V$ finite, then the Cayley-Hamilton Theorem guarantees that there is a monic polynomial $g \in k[x]$ such that $g(T) = 0$. Namely, we can take $g = f_T$. In general, for a linear transformation $T : V \to V$, suppose that there exists a nonzero polynomial $g \in k[x]$ such that $g(T) = 0$. Then there is a nonzero polynomial $f \in k[x]$ of least degree such that $f(T) = 0$. We can multiply f by a nonzero scalar so that we may assume that f is also a monic polynomial. Suppose that h is another monic polynomial of least degree such that $h(T) = 0$ with $f \neq h$. Then $(f - h)(T) = f(T) - h(T) = 0$. Since $\deg(f) = \deg(h)$, both are f and h are monic, and $f - h \neq 0$, it follows that $\deg(f - h) < \deg f$. This is impossible because there is no nonzero polynomial h of degree less that $\deg(f)$ such that $h(T) = 0$. Thus $f = h$. Therefore, there exists a unique monic polynomial f of least degree such that $f(T) = 0$. This justifies the following definitions.

Definition 7.13. *Let $T : V \to V$ be a linear transformation. Suppose that there exists a nonzero polynomial $g \in k[x]$ such that $g(T) = 0$. (This always occurs if $\dim V$ is finite.) The minimal polynomial of T is the unique monic polynomial $p_T(x)$ of least degree such that $p_T(T) = 0$. Similarly, the minimal polynomial of a matrix $A \in \mathcal{M}_{n\times n}(k)$ is the unique monic polynomial $p_A(x)$ of least degree such that $p_A(A) = 0$.*

Lemma 7.14. *Let $T : V \to V$ be a linear transformation. Suppose that $f \in k[x]$ and $f(T) = 0$. Let $p_T(x) \in k[x]$ be the minimal polynomial of T. Then $p_T(x) \mid f(x)$. In particular, if $\dim V$ is finite, then $p_T(x) \mid f_T(x)$.*

A similar result holds for the minimal polynomial of a matrix $A \in \mathcal{M}_{n\times n}(k)$.

Proof. Use the division algorithm to write $f(x) = p_T(x)q(x) + r(x)$ where $q(x), r(x) \in k[x]$ and either $r(x) = 0$ or $\deg r(x) < \deg p_T(x)$. Then

$$0 = f(T) = p_T(T)q(T) + r(T) = r(T).$$

If $r(x) \neq 0$, then we can write $r(x) = cs(x)$ where $s(x)$ is a monic polynomial with $\deg(s(x)) = \deg(r(x))$ and $c \in k$ is nonzero. Then $r(T) = 0$

implies $s(T) = 0$. This is impossible because $s(x)$ is a monic polynomial with $\deg(s(x)) < \deg(p_T(x))$. Therefore $r(x) = 0$ and so $p_T(x) \mid f(x)$.

If $\dim V$ is finite, then $f_T(T) = 0$ by the Cayley-Hamilton Theorem. It follows that $p_T(x) \mid f_T(x)$.

An analogous proof holds for the minimal polynomial of a matrix $A \in \mathcal{M}_{n\times n}(k)$. □

7.4 More results on T-Invariant Subspaces

Let $T : V \to V$ be a linear transformation.

Lemma 7.15. *Let $g \in k[x]$ be a polynomial and suppose that $\deg(g) = m \geq 1$. If $\ker(g(T)) \neq 0$, then V contains a nonzero T-invariant subspace W of dimension $\leq m$. If g is irreducible in $k[x]$, then $\dim W = m$.*

Proof. Let $v \in \ker(g(T))$ and assume that $v \neq 0$. Let

$$W = \langle v, Tv, T^2v, \ldots, T^{m-1}v\rangle.$$

Let $g = a_mx^m + \cdots + a_1x + a_0$ with $a_m \neq 0$. Since $g(T)(v) = 0$, it follows that $T^m(v) = -a_m^{-1}(a_{m-1}T^{m-1} + \cdots + a_1T + a_0 1_V)(v)$. Then $T^m(v)$ lies in W and thus $T(T^iv) \in W$ for $0 \leq i \leq m-1$. Then W is a nonzero T-invariant subspace of V and $\dim(W) \leq m$.

Now assume that g is an irreducible polynomial. We will show that $\dim W = m$. Suppose that $v, Tv, T^2v, \ldots, T^{m-1}v$ are linearly dependent over k. Then there exist $b_0, \ldots, b_{m-1}$ in k, not all zero, such that

$$b_0v + b_1Tv + \cdots + b_{m-1}T^{m-1} = 0.$$

Then $h(T)(v) = 0$, where $h(x) = b_{m-1}x^{m-1} + \cdots + b_1x + b_0$ is a nonzero polynomial. Let $f = \gcd(g, h)$. Recall that there exist polynomials $r(x)$ and $s(x)$ in $k[x]$ such that $f = rg + sh$. Then $f(T)(v) = 0$ because $g(T)(v) = 0$ and $h(T)(v) = 0$. We have $f = 1$ because g is irreducible in $k[x]$ and $\deg(h) < \deg(g)$. Therefore, $v = 0$, a contradiction. It follows that $v, Tv, T^2v, \ldots, T^{m-1}v$ are linearly independent over k and that $\dim(W) = m$. □

Corollary 7.16. *Suppose that there exists a nonzero polynomial $f \in k[x]$ such that $f(T) = 0$. Write $f(x) = f_1(x)\cdots f_m(x)$ where each $f_i(x)$ is an irreducible polynomial in $k[x]$. Then V contains a nonzero T-invariant subspace of dimension equal to $\dim f_i(x)$ for some i.*

Proof. First suppose that $\ker(f_i(T)) = 0$ for $1 \le i \le m$. Then $f_i(T)$ is injective for each i and it follows that the composition $f(T) = f_1(T) \cdots f_m(T)$ is injective. This is a contradiction because $f(T) = 0$. Therefore $\ker(f_i(T)) \neq 0$ for some i. Then the result follows from Lemma 7.15. □

Proposition 7.17. *Suppose that there exists a nonzero polynomial $f \in k[x]$ such that $f(T) = 0$.*

1. *If $k = \mathbf{C}$, the field of complex numbers (or any algebraically closed field), then V contains a one-dimensional T-invariant subspace. In other words, V contains an eigenvector for T.*

2. *If $k = \mathbf{R}$, the field of real numbers (or any real closed field), then V contains a T-invariant subspace of dimension ≤ 2.*

Proof. Let $f(T) = 0$, where $f \in k[x]$ is a nonzero polynomial of degree m.

(1) Since k is algebraically closed, f factors as a product of linear polynomials over k. Thus we have $f(x) = b(x - c_1)(x - c_2) \cdots (x - c_m)$ where b and each c_i lie in k. The result now follows from Corollary 7.16.

(2) Since k is the field of real numbers (or is real closed), f factors as a product of linear and irreducible quadratic polynomials over k. Thus we have $f(x) = \prod_{i=1}^m g_i(x)$, where $\deg g_i \le 2$ and each g_i is an irreducible polynomial in $k[x]$. The result follows from Corollary 7.16. □

In the next two sections, we attempt to find bases of V so that the matrix $[T]_\beta^\beta$ is as simple as possible.

7.5 Primary Decomposition

Let $T : V \to V$ be a linear transformation.

Proposition 7.18. *Suppose that f is a nonzero polynomial in $k[x]$ such that $f(T) = 0$. Let $f(x) = f_1(x)f_2(x)$, where $f_1, f_2 \in k[x]$ and $\gcd(f_1, f_2) = 1$.*

Let $V_i = \ker(f_i(T))$, $i = 1, 2$. Then the following statements hold.

1. *$V = V_1 \oplus V_2$.*

2. *V_i is a T-invariant subspace of V for $i = 1, 2$.*

3. *Let $T_i = T|_{V_i} : V_i \to V_i$ be the induced linear transformation from (2). Let $p_{T_i}(x)$ denote the minimal polynomial of T_i. Then $p_{T_i}(x) \mid f_i(x)$.*

4. *If* $\dim V_i$ *is finite, let* $f_{T_i}(x)$ *denote the characteristic polynomial of* T_i*. Then* $p_{T_i}(x) \mid f_{T_i}(x)$*.*

5. *We have* $p_T(x) = p_{T_1}(x)p_{T_2}(x)$*. If* $\dim V_1$ *and* $\dim V_2$ *are both finite, then* $f_T(x) = f_{T_1}(x)f_{T_2}(x)$*.*

Proof. There exist polynomials $g_1, g_2 \in k[x]$ such that $f_1g_1 + f_2g_2 = 1$. Let $v \in V$. Then $v = 1_V(v) = f_1(T)g_1(T)(v) + f_2(T)g_2(T)(v)$. Let $v_1 = f_2(T)g_2(T)(v)$ and let $v_2 = f_1(T)g_1(T)(v)$, so that $v = v_1 + v_2$. Then $v_1 \in V_1$ because

$$f_1(T)(v_1) = f_1(T)f_2(T)g_2(T)(v) = f(T)(g_2(T)(v)) = 0.$$

Similarly, $v_2 \in V_2$. Thus $V = V_1 + V_2$.

Now suppose that $v_1 \in V_1$, $v_2 \in V_2$ and $v_1 + v_2 = 0$. Then

$$0 = f_2(T)(v_1 + v_2) = f_2(T)(v_1) + f_2(T)(v_2) = f_2(T)(v_1).$$

Since $f_1(T)(v_1) = 0$, it follows that

$$v_1 = 1_V(v_1) = g_1(T)f_1(T)(v_1) + g_2(T)f_2(T)(v_1) = 0.$$

Similarly, it follows that $v_2 = 0$. Therefore, $V = V_1 \oplus V_2$, which proves (1).

For (2), let $v_i \in V_i$. Then $f_i(T)(T(v_i)) = Tf_i(T)(v_i) = T(0) = 0$. Thus $T(v_i) \in \ker(f_i(T)) = V_i$, so (2) holds.

For (3), let $v \in V_i$. Then $f_i(T_i)(v) = f_i(T)(v) = 0$ because $v \in V_i = \ker(f_i(T))$. Thus $f_i(T_i) = 0$, so $p_{T_i}(x) \mid f_i(x)$ by Lemma 7.14.

If $\dim V_i$ is finite, then $p_{T_i}(x) \mid f_{T_i}(x)$ by Lemma 7.14 because $f_{T_i}(T_i) = 0$ by the Cayley-Hamilton Theorem. This proves (4).

Now we prove (5). Let $v \in V_i$. Then $p_T(T_i)(v) = p_T(T)(v) = 0$. Thus, for $i = 1, 2$, we have $p_{T_i}(x) \mid p_T(x)$ by Lemma 7.14. Since $p_{T_i}(x) \mid f_i(x)$ by (3) and $\gcd(f_1, f_2) = 1$, it follows that $\gcd(p_{T_1}, p_{T_2}) = 1$, so $p_{T_1}(x)p_{T_2}(x) \mid p_T(x)$.

Now let $v \in V$ and let $v = v_1 + v_2$ where $v_i \in V_i$. Then

$$\begin{aligned} p_{T_1}(T)p_{T_2}(T)(v) &= p_{T_1}(T)p_{T_2}(T)(v_1 + v_2) \\ &= p_{T_2}(T)p_{T_1}(T)(v_1) + p_{T_1}(T)p_{T_2}(T)(v_2) = 0 + 0 = 0. \end{aligned}$$

Thus $p_T(x) \mid p_{T_1}(x)p_{T_2}(x)$. Therefore, $p_T(x) = p_{T_1}(x)p_{T_2}(x)$.

Now assume that $\dim V_1$ and $\dim V_2$ are both finite. Then Lemma 7.4 (with its notation) implies that

$$\begin{aligned} f_T(x) &= \det(xI_n - T) = \det \begin{pmatrix} xI_l - A & 0 \\ 0 & xI_m - B \end{pmatrix} \\ &= \det(xI_l - A)\det(xI_m - B) = f_{T_1}(x) f_{T_2}(x). \end{aligned}$$

□

Proposition 7.19. *Suppose that $f \in k[x]$ is a nonzero polynomial such that $f(T) = 0$. Let $f(x) = f_1(x) \cdots f_m(x)$, where $f_1, \dots, f_m \in k[x]$ are pairwise relatively prime polynomials.*

Let $V_i = \ker(f_i(T))$, $1 \le i \le m$. Then the following statements hold.

1. *$V = V_1 \oplus \cdots \oplus V_m$.*
2. *V_i is a T-invariant subspace of V.*
3. *Let $T_i = T|_{V_i} : V_i \to V_i$ be the induced linear transformation from (2). Let $p_{T_i}(x)$ denote the minimal polynomial of T_i. Then $p_{T_i}(x) \mid f_i(x)$.*
4. *If $\dim V_i$ is finite, let $f_{T_i}(x)$ denote the characteristic polynomial of T_i. Then $p_{T_i}(x) \mid f_{T_i}(x)$.*
5. *We have $p_T(x) = p_{T_1}(x) \cdots p_{T_m}(x)$. If $\dim V_i$ is finite, $1 \le i \le m$, then $f_T(x) = f_{T_1}(x) \cdots f_{T_m}(x)$.*

Proof. We prove this by induction on m. The case $m = 1$ is trivial and the case $m = 2$ follows from Proposition 7.18. Now assume that $m \ge 3$. Let $g = f_1(x) \cdots f_{m-1}(x)$, and let $W = \ker(g(T))$. Then $\gcd(g, f_m) = 1$ and $V = W \oplus V_m$ by Proposition 7.18(1). By induction, we have that $W = V_1 \oplus \cdots \oplus V_{m-1}$ and thus (1) holds. The remaining statements are easy conclusions from the induction hypothesis and Proposition 7.18. (See Exercise 4.) □

The next result gives the so called Primary Decomposition of V respect to T.

Corollary 7.20. *Suppose that $f \in k[x]$ is a monic polynomial such that $f(T) = 0$. Let $f(x) = f_1(x)^{e_1} \cdots f_m(x)^{e_m}$, where $f_1, \dots, f_m \in k[x]$ are distinct monic irreducible polynomials and each $e_i > 0$.*

Let $V_i = \ker(f_i(T)^{e_i})$, $1 \le i \le m$. Then the following statements hold.

1. $V = V_1 \oplus \cdots \oplus V_m$.

2. *V_i is a T-invariant subspace of V.*

3. *Let $T_i = T|_{V_i} : V_i \to V_i$ be the induced linear transformation from (2). Let $p_{T_i}(x)$ denote the minimal polynomial of T_i. Then $p_{T_i}(x) \mid f_i^{e_i}(x)$.*

4. *If $\dim V_i$ is finite, let $f_{T_i}(x)$ denote the characteristic polynomial of T_i. Then $p_{T_i}(x) \mid f_{T_i}(x)$.*

5. *We have $p_T(x) = p_{T_1}(x) \cdots p_{T_m}(x)$. If $\dim V_i$ is finite, $1 \leq i \leq m$, then $f_T(x) = f_{T_1}(x) \cdots f_{T_m}(x)$.*

Proof. The result follows immediately from Proposition 7.19. □

We now use a slightly more abstract way to obtain the primary decomposition of V with respect to T.

Lemma 7.21. *Suppose that $E_i : V \to V$, $1 \leq i \leq m$, are linear transformations satisfying the following statements.*

1. $E_1 + \cdots + E_m = 1_V$,

2. $E_iE_j = 0$ *if* $i \neq j$.

Let $V_i = \operatorname{im}(E_i)$, $1 \leq i \leq m$. Then

i. $E_i^2 = E_i$, $1 \leq i \leq m$.

ii. $V = V_1 \oplus \cdots \oplus V_m$.

iii. $\ker(E_i) = \sum_{\substack{j=1 \\ j \neq i}}^{m} \operatorname{im}(E_j) = \sum_{\substack{j=1 \\ j \neq i}}^{m} V_j$.

Proof. Since $E_iE_j = 0$ when $i \neq j$, we have

$$E_i^2 = E_i(E_1 + E_2 + \cdots + E_i + \cdots + E_m) = E_i 1_V = E_i.$$

Let $v \in V$. Then $v = 1_V v = (E_1 + \cdots + E_m)v \in V_1 + \cdots + V_m$, since $E_j v \in \operatorname{im}(E_j) = V_j$. Thus $V = V_1 + \cdots + V_m$.

Suppose that $v \in V_1 \cap (V_2 + \cdots + V_m)$. Then there exist $y_i \in V_i$, $1 \leq i \leq m$, such that

$$v = E_1y_1 = E_2y_2 + \cdots + E_my_m.$$

Applying E_1 gives

$$E_1 v = E_1^2 y_1 = E_1 E_2 y_2 + \cdots + E_1 E_m y_m = 0,$$

because $E_1 E_j = 0$ for $j \geq 2$. Since $E_1^2 = E_1$, we have $0 = E_1^2 y_1 = E_1 y_1 = v$. Thus $V_1 \cap (V_2 + \cdots + V_m) = 0$. Similarly,

$$V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_m) = 0,$$

for $1 \leq i \leq m$. Therefore, $V = V_1 \oplus \cdots \oplus V_m$.

Since $E_i E_j = 0$ if $i \neq j$, it follows that $\mathrm{im}(E_j) \subseteq \ker(E_i)$, for $i \neq j$. Thus $\sum_{\substack{j=1 \\ j\neq i}}^{m} \mathrm{im}(E_j) \subseteq \ker(E_i)$. Let $v \in \ker(E_i)$. Then

$$v = (E_1 + \cdots + E_m)v = \sum_{j=1}^{m} E_j(v) = \sum_{\substack{j=1 \\ j\neq i}}^{m} E_j(v) \in \sum_{\substack{j=1 \\ j\neq i}}^{m} \mathrm{im}(E_j).$$

Therefore, $\ker(E_i) = \sum_{\substack{j=1 \\ j\neq i}}^{m} \mathrm{im}(E_j)$. □

We now use Lemma 7.21 to obtain a second proof of Proposition 7.19(1).

Proposition 7.22. *Suppose that f is a nonzero monic polynomial $f \in k[x]$ such that $f(T) = 0$. Let $f(x) = f_1(x) \cdots f_m(x)$, where $f_1, \ldots, f_m \in k[x]$ are pairwise relatively prime polynomials.*

Let $V_i = \ker(f_i(T))$, $1 \leq i \leq m$. Then $V = V_1 \oplus \cdots \oplus V_m$.

Proof. Let $g_j(x) = \dfrac{f(x)}{f_j(x)}$, $1 \leq j \leq m$. Then $\gcd(g_1(x), \ldots, g_m(x)) = 1$. Thus there exist $h_j(x) \in k[x]$, $1 \leq j \leq m$, such that $\sum_{j=1}^{m} g_j(x)h_j(x) = 1$. If $i \neq j$, then $f_i(x) \mid g_j(x)$ and so $f(x) \mid g_i(x)g_j(x)$. In other words, $\dfrac{g_i(x)g_j(x)}{f(x)} = \dfrac{g_i(x)g_j(x)}{g_i(x)f_i(x)} \in k[x]$.

Now we show that Lemma 7.21 can be applied. Let $E_j = g_j(T)h_j(T)$. Since $\sum_{j=1}^{m} g_j(x)h_j(x) = 1$, we have

$$\sum_{j=1}^{m} E_j = \sum_{j=1}^{m} g_j(T)h_j(T) = 1_V.$$

Since $f(x) \mid g_i(x)g_j(x)$ when $i \neq j$ and $f(T) = 0$, we have $g_i(T)g_j(T) = 0$, and thus

$$E_iE_j = g_i(T)h_i(T)g_j(T)h_j(T) = g_i(T)g_j(T)h_i(T)h_j(T) = 0.$$

Therefore, the hypotheses of Lemma 7.21 hold.

Now we show that $\operatorname{im}(E_i) = \ker(f_i(T))$. Since $f_i(T)g_i(T) = f(T) = 0$, we have $\operatorname{im}(g_i(T)) \subseteq \ker(f_i(T))$. Thus

$$\operatorname{im}(E_i) = \operatorname{im}(g_i(T)h_i(T)) \subseteq \operatorname{im}(g_i(T)) \subseteq \ker(f_i(T)).$$

Let $v \in \ker(f_i(T))$. If $i \neq j$, then $f_i(x) \mid g_j(x)$, and so $f_i(T)(v) = 0$ implies that $g_j(T)(v) = 0$. Thus $E_j(v) = g_j(T)h_j(T)(v) = h_j(T)g_j(T)(v) = 0$. Thus

$$v = (E_1 + \cdots + E_m)(v) = E_i(v) \in \operatorname{im}(E_i).$$

Therefore, $\operatorname{im}(E_i) = \ker(f_i(T)) = V_i$, $1 \leq i \leq m$. Now the result follows from Lemma 7.21. □

We can now extend Proposition 7.12 by giving another condition that characterizes when a linear transformation T is diagonalizable.

Proposition 7.23. *Let $T : V \to V$ be a linear transformation. The following statements are equivalent.*

1. *T is diagonalizable.*
2. *There exists a basis of V consisting of eigenvectors of T.*
3. *The minimal polynomial $p_T(x)$ of T is square-free and factors completely over k.*

Proof. We have already proved in Proposition 7.12 that (1) and (2) are equivalent. Assume that (2) holds and let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V consisting of eigenvectors of T. Let $A = [T]_\beta^\beta = (a_{ij})$. Then A is a diagonal matrix and $T(v_i) = a_{ii}v_i$ for $1 \leq i \leq m$. Let $c_1, \ldots, c_m$ be the distinct elements that lie in $\{a_{11}, \ldots, a_{nn}\}$. Thus $m \leq n$.

Let $g(x) = (x - c_1)\cdots(x - c_m)$. Let $v \in \beta$ and suppose that $T(v) = c_iv$. Since all linear transformations in $k[T]$ commute with one another, we have

$$g(T)(v) = (T - c_1 1_V)\cdots(T - c_{i-1}1_V)(T - c_{i+1}1_V)\cdots(T - c_m 1_V)(T - c_i 1_V)(v).$$

Since $(T - c_i 1_V)(v) = T(v) - c_i v = 0$, it follows that $g(T)(v) = 0$ for all $v \in \beta$. Then $g(T)(v) = 0$ for all $v \in V$. Therefore $g(T) = 0$. Then $p_T(x) \mid g(x)$ by Lemma 7.14. Since $g(x)$ is square-free and factors completely over k, it follows that the same holds for $p_T(x)$. Thus (3) holds.

Now assume that (3) holds, so that the minimal polynomial $p_T(x)$ of T is square-free and factors completely over k. Then $p_T(x) = (x - c_1) \cdots (x - c_m)$, where $c_1, \ldots, c_m \in k$ are distinct elements. Proposition 7.19 implies that $V = V_1 \oplus \cdots \oplus V_m$, where $V_i = \ker(T - c_i 1_V)$, $1 \leq i \leq m$. Each element of $\ker(T - c_i 1_V)$ is an eigenvector of T with eigenvalue c_i. Let β_i be a basis of $\ker(T - c_i 1_V)$. Then $\beta = \beta_1 \cup \cdots \cup \beta_m$ is a basis of V consisting of eigenvectors of T. Thus (2) holds. □

In the proof of (2) implies (3) in the proof of Proposition 7.23, we can prove even more. We now prove that $g(x) = p_T(x)$. We have already proved that $p_T(x) \mid g(x)$. If $g(x) \neq p_T(x)$, then we can relabel to assume that $p_T(x) \mid (x - c_1) \cdots (x - c_{m-1})$. There exists $v \in \beta$ such that $T(v) = c_m v$. We have

$$(T - c_1 1_V) \cdots (T - c_{m-1} 1_V)(v) = g(T)(v) = 0.$$

Since $(T - c_i 1_V)(v) = (c_m - c_i)v$, it follows that

$$(T - c_1 1_V) \cdots (T - c_{m-1} 1_V)(v) = \prod_{i=1}^{m-1} (c_m - c_i)v \neq 0.$$

This is impossible, and therefore $g(x) = p_T(x)$.

7.6 Further Decomposition

We begin with two definitions that are similar to Definition 7.13.

Definition 7.24. *Let $v \in V$ and let $W \subseteq V$ be a subspace. Suppose that there exists a nonzero polynomial $f \in k[x]$ such that $f(T)(v) \in W$.*

1. *We let $p(v, W) \in k[x]$ denote the unique monic polynomial $g \in k[x]$ of least degree such that $g(T)(v) \in W$.*

2. *In the special case of (1) when $W = (0)$ and $f(T)(v) = 0$, we let $p_v(x)$ denote the unique monic polynomial $g \in k[x]$ of least degree such that $p_v(T)(v) = 0$. We call $p_v(x)$ the T-order of v.*

The following lemma is proved as in Lemma 7.14.

Lemma 7.25. *Let $v \in V$ and let $W \subseteq V$ be a T-invariant subspace. If $f \in k[x]$ and $f(T)(v) \in W$, then $p(v, W) \mid f(x)$. If $f(T)(v) = 0$, then $p_v(x) \mid f(x)$.*

Proof. Use the division algorithm to write $f(x) = p(v, W)q(x) + r(x)$ where either $r(x) = 0$ or $\deg(r(x)) < \deg(p(v, W))$. We have that $f(T)(v) \in W$ and $p(v, W)(T)q(T)(v) = q(T)p(v, W)(T)(v) \in W$ because $p(v, W)(T)(v) \in W$ and W is a T-invariant subspace. Then $r(T)(v) \in W$. The proof finishes as in Lemma 7.14.

Since (0) is a T-invariant subspace of V, it follows that if $f(T)(v) = 0$, then $p_v(x) \mid f(x)$. □

Proposition 7.26. *Let $v \in V$ and suppose that $\deg p_v(x) = m$. Then $\{v, Tv, T^2v, \ldots, T^{m-1}v\}$ is a basis of $k[T]v$ and $\dim k[T]v = \deg p_v(x)$.*

Proof. Let $p_v(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0$. Since $p_v(T)(v) = 0$, we have $T^m v = -(a_{m-1}T^{m-1}v + \cdots + a_1Tv + a_0v)$. It follows easily that $k[T]v$ is spanned by $\{v, Tv, T^2v, \ldots, T^{m-1}v\}$.

Suppose that $\sum_{i=0}^{m-1} b_iT^iv = 0$ is a nontrivial linear dependence relation. Then $g(x) = b_{m-1}x^{m-1} + \cdots + b_1x + b_0$ is a nonzero polynomial with $\deg g(x) \le m - 1 < \deg p_v(x)$ such that $g(T)(v) = 0$. Since we may multiply by a scalar to obtain a monic polynomial with the same property, this contradicts the definition of $p_v(x)$. Therefore, $\{v, Tv, T^2v, \ldots, T^{m-1}v\}$ is a linearly independent set and thus forms a basis of $k[T]v$. It is now clear that $\dim k[T]v = \deg p_v(x)$. □

Definition 7.27. *Let W be a subspace of V. Then W is a T-admissible subspace of V if the following two statements hold.*

1. *W is a T-invariant subspace of V.*
2. *If $y \in V$, $g \in k[x]$, and $g(T)y \in W$, then there exists $z \in W$ such that $g(T)z = g(T)y$.*

For example, (0) and V are T-admissible subspaces of V. For if $W = (0)$, we may always take $z = 0$, while if $W = V$, then we may always take $z = y$.

The following lemma constitutes the key step in the proofs of Theorems 7.30 and 7.31 below.

Lemma 7.28. *Let $W \subsetneq V$ be a T-admissible subspace. Suppose that $f(x) \in k[x]$ is a polynomial such that $f(T) = 0$. Then there exists $v \in V$, $v \notin W$, such that*

1. $W \cap k[T]v = (0)$
2. $W + k[T]v$ *is a T-admissible subspace of V.*

Proof. Choose $y \in V$ such that $\deg p(y, W)$ is maximal. This is possible because $p(y, W) \mid f(x)$ and hence $\deg p(y, W) \le \deg f(x)$. We have $y \notin W$ because $W \subsetneq V$. Let $g(x) = p(y, W)$. Then $g(T)y \in W$. Since W is T-admissible, there exists $y' \in W$ such that $g(T)y = g(T)y'$.

Let $v = y - y'$. Thus $v \notin W$ because $y \notin W$ and $y' \in W$. We now show that $W \cap k[T]v = (0)$. Suppose that $h \in k[x]$ and $h(T)v \in W$. Then

$$h(T)y - h(T)y' = h(T)(y - y') = h(T)v \in W.$$

We have $h(T)y' \in W$ because $y' \in W$ and W is T-invariant. Thus $h(T)y \in W$. Therefore, $g(x) \mid h(x)$. We have

$$g(T)v = g(T)(y - y') = g(T)y - g(T)y' = 0.$$

Thus $h(T)v = 0$ because $g(x) \mid h(x)$. Therefore $W \cap k[T]v = (0)$. This proves (1).

For (2), we first note that $W + k[T]v$ is T-invariant because W and $k[T]$ are each T-invariant.

Suppose that $u \in V$ and $h(T)u \in W + k[T]v$. We must find $u' \in W + k[T]v$ such that $h(T)u = h(T)u'$. We can assume that $h(x) = p(u, W + k[T]v)$ because if $h(T)u = h(T)u'$, then $s(T)h(T)u = s(T)h(T)u'$ for all $s(x) \in k[x]$.

Let $h(T)u = w + j(T)v$, where $w \in W$ and $j(x) \in k[x]$. We now prove that $h(x) \mid j(x)$. Use the division algorithm to write $j(x) = h(x)q(x) + r(x)$ where either $r(x) = 0$ or $\deg r(x) < \deg h(x)$. Let $z = u - q(T)v$. Then $z - u \in k[T]v$, and so $p(z, W + k[T]v) = p(u, W + k[T]v) = h(x)$. We have

$$\begin{aligned} h(T)z &= h(T)(u - q(T)v) = h(T)u - h(T)q(T)v \\ &= w + j(T)v - h(T)q(T)v = w + (j(T) - h(T)q(T))v = w + r(T)v. \end{aligned}$$

Let $n(x) = p(z, W)$. Since $h(x) = p(z, W + k[T]v)$, it follows that $h(x) \mid n(x)$. Let $n(x) = h(x)m(x)$. Then

$$n(T)z = m(T)h(T)z = m(T)(w + r(T)v) = m(T)w + m(T)r(T)v.$$

Since $n(T)z \in W$ (because $n(x) = p(z, W)$) and $m(T)w \in W$, we have $m(T)r(T)v \in W$. If $r(x) \neq 0$, then

$$\begin{aligned}\deg(m(x)r(x)) &\geq \deg p(v, W) = \deg p(y, W) \geq \deg p(z, W)\\ &= \deg n(x) = \deg h(x)m(x).\end{aligned}$$

(The y above is the y from the proof of (1). We also used that $\deg(p(y, W))$ is maximal.)

Thus $\deg r(x) \geq \deg h(x)$, a contradiction. Therefore, $r(x) = 0$ and $h(x) \mid j(x)$. Thus $j(x) = h(x)q(x)$.

From above, $h(T)z = w + r(T)v = w$. Since $w = h(T)z$ and W is T-admissible, we have $w = h(T)z = h(T)z'$ for some $z' \in W$. Thus,

$$h(T)u = w + j(T)v = h(T)z' + h(T)q(T)v = h(T)(z' + q(T)v) = h(T)u',$$

where $u' = z' + q(T)v \in W + k[T]v$. Therefore, $W + k[T]v$ is a T-admissible subspace of V. □

Lemma 7.29. *Let W be a subspace of V and suppose that $W = W_1 \oplus \cdots \oplus W_m$. Assume that $W_1, \ldots, W_m$ are each T-invariant subspaces of V. If W is a T-admissible subspace of V, then $W_1, \ldots, W_m$ are each T-admissible subspaces of V.*

Proof. Let $y \in V$ and suppose that $g(T)y \in W_i$. Since $W_i \subseteq W$, there exists $z \in W$ such that $g(T)y = g(T)z$. Let $z = z_1 + \cdots + z_m$ where $z_j \in W_j$. Since each W_j is T-invariant, we have $g(T)y = g(T)z = g(T)z_1 + \cdots + g(T)z_m$ with $g(T)z_j \in W_j$, $1 \leq j \leq m$. Since $g(T)y \in W_i$, it follows that $g(T)y = g(T)z_i$ with $z_i \in W_i$. Thus each W_i is a T-admissible subspace of V. □

Theorem 7.30. *Assume that $f(T) = 0$ for some polynomial $f \in k[x]$. The following statements are equivalent for a subspace W of V.*

1. *W is a T-admissible subspace of V.*
2. *W is a T-invariant subspace of V and there exists a T-invariant subspace W' of V such that $V = W \oplus W'$.*

Proof. First assume that (2) holds. Since V is a T-admissible subspace, Lemma 7.29 implies that W is a T-admissible subspace of V and (1) holds.

Now assume that (1) holds. If $W = V$, then let $W' = 0$. We may now suppose that $W \subsetneq V$. Let $\mathcal{S}$ denote the set of subspaces Y of V satisfying the following three conditions.

1. Y is a T-invariant subspace of V.

2. $W \cap Y = (0)$.

3. $W + Y$ is a T-admissible subspace of V.

Then $\mathcal{S}$ is a nonempty set because $(0) \in \mathcal{S}$. The set S is partially ordered by inclusion. Let W' be a maximal element of S. If $\dim V$ is finite, then it is clear that W' exists. Otherwise the existence of W' follows from Zorn's Lemma.

We have $W + W' = W \oplus W'$ by (2). We now show that $V = W \oplus W'$. Suppose that $W \oplus W' \subsetneq V$. Then Lemma 7.28 implies that there exists $v \in V$, $v \notin W \oplus W'$, such that $(W \oplus W') \cap k[T]v = (0)$ and $(W \oplus W') + k[T]v$ is a T-admissible subspace of V. Then $(W \oplus W') + k[T]v = W \oplus W' \oplus k[T]v$. We now check that $W' \oplus k[T]v \in \mathcal{S}$. We have that $W' \oplus k[T]v$ is T-invariant because W' and $k[T]v$ are each T-invariant. We have $W \cap (W' \oplus k[T]v) = 0$ because of the direct sum decomposition. Finally, $W + (W' \oplus k[T]v)$ is T-admissible. This is a contradiction because $W' \subsetneq W' \oplus k[T]v$. Therefore, $V = W \oplus W'$ and (2) holds. □

Theorem 7.31. *Let V be a finite dimensional vector space over k. Then V is a direct sum of T-cyclic subspaces of V. One summand $k[T]v$ can be chosen such that $k[T]v$ is a maximal T-cyclic subspace of V.*

Proof. Let W be a maximal subspace of V with the property that W is a T-admissible subspace of V and is also a direct sum of T-cyclic subspaces of V. (The set of such subspaces is nonempty because (0) is a T-cyclic subspace of V that is also T-admissible.) If $W \subsetneq V$, then Lemma 7.28 and Proposition 7.1 imply that there exists $v \in V$, $v \notin W$, such that $W \cap k[T]v = (0)$ and $W + k[T]v$ is a T-admissible subspace of V. Then $W + k[T]v = W \oplus k[T]v$ and $W \subsetneq W \oplus k[T]v$. This is a contradiction, and thus $W = V$.

Now we show that one summand $k[T]v$ can be chosen such that $k[T]v$ is a maximal T-cyclic subspace of V. Suppose that $v \in V$ is such that $k[T]v$ is a maximal T-cyclic subspace of V. Let $W = k[T]v$. The proof of Lemma 7.28 shows that W is a T-admissible subspace of V. (In the proof of Lemma 7.28, let $W = (0)$, $y = v$, and $y' = 0$ so that $v = y - y'$.) Theorem 7.30 implies that $V = W \oplus W'$, where W' is a T-invariant subspace of V. The first part of this Theorem implies that W' is a direct sum of T-cyclic subspaces of V. Then $V = k[T]v \oplus W'$ is also a direct sum of T-cyclic subspaces of V. □

Exercises

1. Assume that $\dim V$ is finite and let $T : V \to V$ be a linear transformation. Then T is invertible if and only if $\det(T) \neq 0$.

2. If β is an arbitrary basis of V, then T is diagonalizable if and only if $[T]_\beta^\beta$ is diagonalizable.

3. Suppose $[T]_\beta^\beta = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, where $a \neq b$. Then $P_T(x) = (x-a)(x-b)$.

4. Fill in the details to the proof of Proposition 7.19.

5. Let V be the vector space over $\mathbf{R}$ consisting of all infinitely differentiable functions $f : \mathbf{R} \to \mathbf{R}$. Let $T : V \to V$ be the linear transformation given by differentiation. That is, $T(f) = f'$. Show that there is no nonzero polynomial $g \in \mathbf{R}[x]$ such that $g(T) = 0$.