An Advanced Course in Linear Algebra

Jim L. Brown

November 24, 2014
## Contents

1 Introduction 3

2 4
   2.1 Getting started ........................................ 4
   2.2 Bases and dimension .................................. 13
   2.3 Direct sums and quotient spaces .................... 23
   2.4 Dual spaces .......................................... 27
   2.5 Problems ............................................. 31

3 33
   3.1 Linear transformations and matrices ............... 33
   3.2 Transpose of a matrix via the dual map .......... 38
   3.3 Problems ............................................ 41

4 46
   4.1 Invariant subspaces .................................. 46
   4.2 $T$-invariant complements ............................ 56
   4.3 Rational canonical form ............................... 63
   4.4 Jordan canonical form ................................. 73
   4.5 Diagonalizable operators .............................. 81
   4.6 Canonical forms via modules ......................... 91
   4.7 Problems ............................................. 93

5 96
   5.1 Basic definitions and facts .......................... 96
   5.2 Symmetric, skew-symmetric, and Hermitian forms .. 107
   5.3 The adjoint map ..................................... 120
   5.4 Problems ............................................. 124

6 126
   6.1 Inner product spaces .................................. 126
   6.2 The spectral theorem ................................ 135
   6.3 Polar decomposition and the Singular Value Theorem . 144
   6.4 Problems ............................................. 152
Chapter 1

Introduction

These notes are the product of teaching Math 8530 several times at Clemson University. This course serves as the first year breadth course for the Algebra and Discrete Mathematics subfaculty and is the only algebra course many of our graduate students will see. As such, this course was designed to be a serious introduction to linear algebra at the graduate level giving students a flavor of what modern abstract algebra consists of. While undergraduate abstract algebra is not a formal prerequisite, it certainly helps. To aid with this an appendix reviewing essential undergraduate concepts is included.

It is assumed students have had an undergraduate course in basic linear algebra and are comfortable with concepts such as multiplying matrices, Gaussian elimination, finding the null and column space of matrices, etc. In this course we work with abstract vector spaces over an arbitrary field and linear transformations. We do not limit ourselves to $\mathbb{R}^n$ and matrices, but often do rephrase results in terms of matrices for the convenience of the reader. Once a student has completed and mastered the material in these notes s/he should have no trouble translating these results into the results typically presented in a first or second semester undergraduate linear algebra or matrix analysis course.

While it certainly would be helpful to use modules at several places in these notes, they are purposefully avoided so as not to stray too far from the original intent of the course. Notes using modules may be forthcoming at some point in the future.

Please report any errors you find in these notes to me so that they may be corrected.

The motivation for typing these notes was provided by Kara Stasikelis. She provided her typed class notes from this course Summer 2013. These were greatly appreciated and gave me the starting point from which I typed this version.
Chapter 2

Vector spaces

In this chapter we give the basic definitions and facts having to do with vector spaces that will be used throughout the rest of the course. Many of the results in this chapter are covered in an undergraduate linear algebra class in terms of matrices. We often convert the language back to that of matrices, but the focus is on abstract vector spaces and linear transformations.

Throughout this chapter $F$ is a field.

2.1 Getting started

We begin with the definition of a vector space.

**Definition 2.1.1.** Let $V$ be a non-empty set with an operation

$$V \times V \rightarrow V$$

$$(v, w) \mapsto v + w$$

referred to as *addition* and an operation

$$F \times V \rightarrow V$$

$$(c, v) \mapsto cv$$

referred to as *scalar multiplication* so that $V$ satisfies the following properties:

1. $(V, +)$ is an abelian group;
2. $c(v + w) = cv + cw$ for all $c \in F$, $v, w \in V$;
3. $(c + d)v = cv + dv$ for all $c, d \in F$, $v \in V$;
4. $(cd)v = c(dv)$ for all $c, d \in F$, $v \in V$;
5. $1_F \cdot v = v$ for all $v \in V$. 


We say $V$ is a vector space (or an $F$-vector space). If we need to emphasize the addition and scalar multiplication we write $(V, +, \cdot)$ instead of just $V$. We call elements of $V$ vectors and elements of $F$ scalars.

We now recall some familiar examples from undergraduate linear algebra. The verification that these are actually vector spaces is left as an exercise.

**Example 2.1.2.** Set $F^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} : a_i \in F \right\}$. Then $F^n$ is a vector space with addition given by

\[
\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}
\]

and scalar multiplication given by

\[
c \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix}.
\]

**Example 2.1.3.** Let $F$ and $K$ be fields and suppose $F \subseteq K$. Then $K$ is an $F$-vector space. The typical example of this one sees in undergraduate linear algebra is $\mathbb{C}$ as an $\mathbb{R}$-vector space.

**Example 2.1.4.** Let $m,n \in \mathbb{Z}_{>0}$. Set $\text{Mat}_{m,n}(F)$ to be the $m$ by $n$ matrices with entries in $F$. This forms an $F$-vector space with addition being matrix addition and scalar multiplication given by multiplying each entry of the matrix by the scalar. In the case $m = n$ we write $\text{Mat}_n(F)$ for $\text{Mat}_{n,n}(F)$.

**Example 2.1.5.** Let $n \in \mathbb{Z}_{\geq 0}$. Define

\[ P_n(F) = \{ f \in F[x] \mid \deg(f) \leq n \} \]

This is an $F$-vector space. We also have that

\[ F[x] = \bigcup_{n \geq 0} P_n(F) \]

is an $F$-vector space.

**Example 2.1.6.** Let $U$ and $V$ be subsets of $\mathbb{R}$. Define $C^0(U,V)$ to be the set of continuous functions from $U$ to $V$. This forms an $\mathbb{R}$-vector space. Let $f, g \in C^0(U,V)$ and $c \in \mathbb{R}$. Addition is defined point-wise, namely, $(f + g)(x) = f(x) + g(x)$ for all $x \in U$. Scalar multiplication is given point-wise as well: $(cf)(x) = cf(x)$ for all $x \in U$. 
More generally, for \( k \in \mathbb{Z}_{\geq 0} \) we let \( C^k(U,V) \) be the functions from \( U \) to \( V \) so that \( f^{(j)} \) is continuous for all \( 0 \leq j \leq k \) where \( f^{(j)} \) denotes the \( j \)th derivative of \( f \). This is an \( \mathbb{R} \)-vector space. If \( U = V \) we write \( C^k(U) \) for \( C^k(U,U) \).

We set \( C^\infty(U,V) \) to be the set of smooth functions, i.e., \( f \in C^\infty(U,V) \) if \( f \in C^k(U,V) \) for all \( k \geq 0 \). This is an \( \mathbb{R} \)-vector space. If \( U = V \) we write \( C^\infty(U) \) for \( C^\infty(U,U) \).

**Example 2.1.7.** Set

\[
F^\infty = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} : a_i \in F, a_i = 0 \text{ for all but finitely many } i \right\}
\]

is an \( F \)-vector space.

Set

\[
F^N = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} : a_i \in F \right\}.
\]

This is an \( F \)-vector space.

**Example 2.1.8.** Consider the sphere

\[
S^{n-1} = \{ (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \}.
\]

Let \( p = (a_1, \ldots, a_n) \) be a point on the sphere. We can realize the sphere as the \( w = 1 \) level surface of the function \( w = f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2 \). The gradient of \( f \) is \( \nabla f = 2x_1e_1 + 2x_2e_2 + \cdots + 2x_ne_n \) where \( e_1 = (1,0,\ldots,0), e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,\ldots,0,1) \). This gives that the tangent plane at the point \( p \) is given by

\[
2a_1(x_1 - a_1) + 2a_2(x_2 - a_2) + \cdots + 2a_n(x_n - a_n) = 0.
\]

This is pictured in the \( n = 2 \) case here:
Note the tangent plane is not a vector space because it does not contain a zero vector, but if we shift it to the origin we have a vector space called the tangent space of $S^{n-1}$ at $p$:

$$T_p(S^{n-1}) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : 2a_1x_1 + \cdots + 2a_nx_n = 0\}.$$ 

The shifted tangent space from the graph pictured above is given here:

### Lemma 2.1.9

Let $V$ be an $F$-vector space.

1. The zero element $0_V \in V$ is unique.
2. We have $0_F \cdot v = 0_V$ for all $v \in V$.
3. We have $(-1_F) \cdot v = -v$ for all $v \in V$.

**Proof.** Suppose that $0, 0'$ both satisfy the conditions necessary to be $0_V$, i.e.,

- $0 + v = v = v + 0$ for all $v \in V$
- $0' + v = v = v + 0'$ for all $v \in V$.

We apply this with $v = 0$ to obtain $0 = 0 + 0'$, and now use $v = 0'$ to obtain $0 + 0' = 0'$. Thus, we have that $0 = 0'$.

Observe that $0_F \cdot v = (0_F + 0_F) \cdot v = 0_F \cdot v + 0_F \cdot v$. So if we subtract $0_F \cdot v$ from both sides we get that $0_v = 0_F \cdot v$.

Finally, we have $(-1_F) \cdot v + v = (-1_F) \cdot v + (1_F) \cdot v = (-1_F + 1_F) \cdot v = 0_F \cdot v = 0_v$, i.e., $(-1_F) \cdot v + v = 0_v$. So $(-1_F) \cdot v$ is the unique additive identity of $v$, i.e., $(-1_F) \cdot v = -v$. 

Note we will often drop the subscripts on 0 and 1 when they are clear from context.

### Definition 2.1.10

Let $V$ be an $F$-vector space and let $W$ be a nonempty subset of $V$. If $W$ is an $F$-vector space with the same operations as $V$ we call $W$ a *subspace* (or an $F$-*subspace*) of $V$. 

---

7
Example 2.1.11. We saw above that $V = \mathbb{C}$ is an $\mathbb{R}$-vector space. Set $W = \{x + 0 \cdot i : x \in \mathbb{R}\}$. Then $W$ is an $\mathbb{R}$-subspace of $V$. We have that $V$ is also a $\mathbb{C}$-vector space. However, $W$ is not a $\mathbb{C}$-subspace of $V$ as it is not closed under scalar multiplication by $i$.

Example 2.1.12. Let $V = \mathbb{R}^2$. In the graph below $W_1$ is a subspace but $W_2$ is not. It is easy to check that any line passing through the origin is a subspace, but a line not passing through the origin cannot be a subspace because it does not contain the zero element.

Example 2.1.13. Let $n \in \mathbb{Z}_{\geq 0}$. We have $P_j(F)$ is a subspace of $P_n(F)$ for all $0 \leq j \leq n$.

Example 2.1.14. The $F$-vector space $F^\infty$ is a subspace of $F^\mathbb{N}$.

Example 2.1.15. Let $V = \text{Mat}_n(F)$. The set of diagonal matrices is a subspace of $V$.

The following lemma gives easy to check criterion for when one has a subspace. The proof is left as an exercise.

Lemma 2.1.16. Let $V$ an $F$-vector space and $W \subseteq V$. Then $W$ is a subspace of $V$ if

1. $W$ is nonempty;
2. $W$ is closed under addition;
3. $W$ is closed under scalar multiplication.

As is customary in algebra, in order to study objects we first introduce the subobjects (subspaces in our case) and then the appropriate maps between the objects of interest. In our case these are linear transformations.
Definition 2.1.17. Let \( V \) and \( W \) be \( F \)-vector spaces and let \( T : V \to W \) be a map. We say \( T \) is a linear transformation (or \( F \)-linear) if

1. \( T(v_1 + v_2) = T(v_1) + T(v_2) \) for all \( v_1, v_2 \in V \);
2. \( T(cv) = cT(v) \) for all \( c \in F, v \in V \).

The collection of all \( F \)-linear maps from \( V \) to \( W \) is denoted \( \text{Hom}_F(V,W) \).

Example 2.1.18. Define \( \text{id}_V : V \to V \) by \( \text{id}_V(v) = v \) for all \( v \in V \). Then \( \text{id}_V \in \text{Hom}_F(V,V) \). We refer to this as the identity transformation.

Example 2.1.19. Let \( T : \mathbb{C} \to \mathbb{C} \) be defined by \( T(z) = \overline{z} \). Since \( \mathbb{C} \) is a \( \mathbb{C} \) and an \( \mathbb{R} \)-vector space, it is natural to ask if \( T \) is \( \mathbb{C} \)-linear or \( \mathbb{R} \)-linear. Observe we have

\[
T(z + w) = \overline{z + w} = \overline{z} + \overline{w} = T(z) + T(w)
\]

for all \( z, w \in \mathbb{C} \). However, if we let \( c \in \mathbb{C} \) we have

\[
T(cv) = c\overline{v} = \overline{c}T(v).
\]

Note that if \( T(v) \neq 0 \), then \( T(cv) = cT(v) \) if and only if \( c = \overline{c} \). Thus, \( T \) is not \( \mathbb{C} \)-linear but is \( \mathbb{R} \)-linear.

Example 2.1.20. Let \( m, n \in \mathbb{Z}_{\geq 1} \). Let \( A \in \text{Mat}_{m,n}(F) \). Define \( T_A : F^n \to F^m \) by \( T_A(x) = Ax \). This is an \( F \)-linear map.

Example 2.1.21. Set \( V = C^\infty(\mathbb{R}) \). In this example we give several linear maps that arise in calculus.

Given any \( a \in \mathbb{R} \), define

\[
E_a : V \to \mathbb{R} \\
f \mapsto f(a).
\]

We have \( E_a \in \text{Hom}_\mathbb{R}(V,\mathbb{R}) \) for every \( a \in \mathbb{R} \).

Define

\[
D : V \to V \\
f \mapsto f'.
\]

We have \( D \in \text{Hom}_\mathbb{R}(V,V) \).
Let $a \in \mathbb{R}$. Define

$$I_a : V \to V \quad f \mapsto \int_a^x f(t) dt.$$ 

We have $I_a \in \text{Hom}_\mathbb{R}(V,V)$ for every $a \in \mathbb{R}$.

Let $a \in \mathbb{R}$. Define

$$\tilde{E}_a : V \to V \quad f \mapsto f(a)$$

where here we view $f(a)$ as the constant function. We have $\tilde{E}_a \in \text{Hom}_\mathbb{R}(V,V)$.

We can use these linear maps to rephrase the fundamental theorems of calculus as follows:

1. $D \circ I_a = \text{id}_V$;
2. $I_a \circ D = \text{id}_V - \tilde{E}_a$.

**Exercise 2.1.22.** Let $V$ and $W$ be $F$-vector spaces. Show that $\text{Hom}_F(V,W)$ is an $F$-vector space.

**Lemma 2.1.23.** Let $T \in \text{Hom}_F(V,W)$. Then $T(0_V) = 0_W$.

**Proof.** This is proved using the properties of the additive identity element:

$$T(0_V) = T(0_V + 0_V) = T(0_V) + T(0_V),$$

i.e., $T(0_V) = T(0_V) + T(0_V)$. Now, subtract $T(0_V)$ from both sides to obtain $0_W = T(0_V)$ as desired. \(\square\)

**Exercise 2.1.24.** Let $U, V, W$ be $F$-vector spaces. Let $S \in \text{Hom}_F(U,V)$ and $T \in \text{Hom}_F(V,W)$. Then $T \circ S \in \text{Hom}_F(U,W)$.

In the next chapter we will focus on matrices and their relation with linear maps much more closely, but we have the following elementary result that we can prove immediately.

**Lemma 2.1.25.** Let $m, n \in \mathbb{Z}_{\geq 1}$.

1. Let $A, B \in \text{Mat}_{m,n}(F)$. Then $A = B$ if and only if $T_A = T_B$.
2. Every $T \in \text{Hom}_F(F^n, F^m)$ is given by $T_A$ for some $A \in \text{Mat}_{m,n}(F)$.

**Proof.** If $A = B$ then clearly we must have $T_A = T_B$ from the definition. Conversely, suppose $T_A = T_B$. Consider the standard vectors $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$. We have $T_A(e_j) = T_B(e_j)$ for $j = 1, \ldots, n$. However, it is easy to see that $T_A(e_j)$ is the $j$th column of $A$. Thus, $A = B$. 

10
Let \( T \in \text{Hom}_F(F^n, F^m) \). Set

\[
A = \left( T(e_1) \ldots T(e_n) \right).
\]

It is now easy to check that \( T = T_A \).

In general, we do not multiply vectors. However, if \( V = \text{Mat}_n(F) \), we can multiply vectors in here! So \( V \) is a vector space, but also a ring. In this case, \( V \) is an example of an \( F \)-algebra. Though we will not be concerned with algebras in these notes, we give the definition here for the sake of completeness.

**Definition 2.1.26.** An \( F \)-algebra is a ring \( A \) with a multiplicative identity together with a ring homomorphism \( f : F \to A \) mapping \( 1_F \) to \( 1_A \) so that \( f(F) \) is contained in the center of \( A \), i.e., if \( a \in A \) and \( f(c) \in f(F) \), then \( af(c) = f(c)a \).

**Exercise 2.1.27.** Show that \( F[x] \) is an \( F \)-algebra.

The fact that we can multiply in \( \text{Mat}_n(F) \) is due to the fact that we can compose linear maps. In fact, it is natural to define matrix multiplication to be the matrix associated to the composition of the associated linear maps. This definition explains the “bizarre” multiplication rule defined on matrices.

**Definition 2.1.28.** Let \( m, n \) and \( p \) be positive integers. Let \( A \in \text{Mat}_{m,n}(F) \), \( B \in \text{Mat}_{n,p}(F) \). Then \( AB \in \text{Mat}_{m,p}(F) \) is the matrix corresponding to \( T_A \circ T_B \).

**Exercise 2.1.29.** Show this definition of matrix multiplication agrees with that given in undergraduate linear algebra class.

**Definition 2.1.30.** Let \( T \in \text{Hom}_F(V,W) \) be invertible, i.e., there exists \( T^{-1} \in \text{Hom}_F(W,V) \) so that \( T \circ T^{-1} = \text{id}_W \) and \( T^{-1} \circ T = \text{id}_V \). We say \( T \) is an isomorphism and we say \( V \) and \( W \) are isomorphic and right \( V \cong W \).

**Exercise 2.1.31.** Let \( T \in \text{Hom}_F(V,W) \). Show that \( T \) is an isomorphism if and only if \( T \) is bijective.

**Example 2.1.32.** Let \( V = \mathbb{R}^2 \), \( W = \mathbb{C} \). These are both \( \mathbb{R} \)-vector spaces. Define

\[
T : \mathbb{R}^2 \to \mathbb{C}
\]

\[
(x, y) \mapsto x + iy.
\]

We have \( T \in \text{Hom}_\mathbb{R}(V,W) \). It is easy to see this is an isomorphism with inverse given by \( T^{-1}(x + iy) = (x, y) \). Thus, \( \mathbb{C} \cong \mathbb{R}^2 \) as \( \mathbb{R} \)-vector spaces.

**Example 2.1.33.** Let \( V = P_n(F) \) and \( W = F^{n+1} \). Define a map \( T : V \to W \) by sending \( a_0 + a_1 x + \cdots + a_n x^n \) to \( (a_0, \ldots, a_n) \). It is elementary to check this is an isomorphism.

**Example 2.1.34.** Let \( V = \text{Mat}_n(F) \) and \( W = F^{n^2} \). Define \( T : V \to W \) by sending \( A = (a_{i,j}) \) to \( (a_{1,1}, a_{1,2}, \ldots, a_{n,n}) \). This gives an isomorphism between \( V \) and \( W \).
2.1. **GETTING STARTED**

**Definition 2.1.35.** Let \( T \in \text{Hom}_F(V,W) \). The *kernel of \( T \) is given by*

\[
\text{ker} \, T = \{ v \in V : T(v) = 0_W \}.
\]

The *image of \( T \) is given by*

\[
\text{Im} \, T = \{ w \in W : \text{there exists } v \in V \text{ with } T(v) = w \} = \text{Im}(T).
\]

One should note that in undergraduate linear algebra one often refers to \( \text{ker} \, T \) as the null space and \( \text{Im} \, T \) the column space when \( T = T_A \) for some \( A \in \text{Mat}_{m,n}(F) \).

**Lemma 2.1.36.** Let \( T \in \text{Hom}_F(V,W) \). Then \( \text{ker} \, T \) is a subspace of \( V \) and \( \text{Im} \, T \) is a subspace of \( W \).

**Proof.** First, observe that \( 0_V \in \text{ker} \, T \) so \( \text{ker} \, T \) is nonempty. Now let \( v_1, v_2 \in \text{ker} \, T \) and \( c \in F \). Then we have

\[
T(v_1 + v_2) = T(v_1) + T(v_2) = 0_W + 0_W = 0_W.
\]

and

\[
T(cv_1) = cT(v_1) = c \cdot 0_W = 0_W.
\]

Thus, \( \text{ker} \, T \) is a subspace of \( V \).

Next we show that \( \text{Im} \, T \) is a subspace. We have \( \text{Im} \, T \) is nonempty because \( T(0_V) = 0_W \in \text{Im} \, T \). Let \( w_1, w_2 \in \text{Im} \, T \) and \( c \in F \). There exists \( v_1, v_2 \in V \) such that \( T(v_1) = w_1 \) and \( T(v_2) = w_2 \). Thus,

\[
w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2).
\]

and

\[
cw_1 = cT(v_1) = T(cv_1).
\]

Thus, \( w_1 + w_2 \) and \( cw_1 \) are both in \( \text{Im} \, T \), i.e., \( \text{Im} \, T \) is a subspace of \( W \). \( \square \)

**Example 2.1.37.** Let \( m, n \in \mathbb{Z}_{>0} \) with \( m > n \). Define \( T : F^m \rightarrow F^n \) by

\[
T \left( \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \right) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.
\]
The image of this map is $F^n$ and the kernel is given by

$$\ker T = \left\{ \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{n+1} \\ \vdots \\ a_m \end{pmatrix} : a_i \in F \right\}.$$ 

It is easy to see that $\ker T \cong F^{m-n}$.

Define $S : F^m \to F^n$ by

$$S \left( \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \right) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where there are $m - n$ zeroes. The image of this map is isomorphic to $F^{m-n}$ and the kernel is trivial.

**Example 2.1.38.** Define $T : \mathbb{Q}[x] \to \mathbb{R}$ by $T(f(x)) = f(\sqrt{3})$. This is a $\mathbb{Q}$-linear map. The kernel of this map consists of those polynomials $f \in \mathbb{Q}[x]$ satisfying $f(\sqrt{3}) = 0$, i.e., those polynomials $f$ for which $(x^2 - 3) \mid f$. Thus, $\ker T = (x^2 - 3)\mathbb{Q}[x]$. The image of this map clearly contains $\mathbb{Q}(\sqrt{3}) = \{a+b\sqrt{3} : a,b \in \mathbb{Q}\}$. Let $\alpha \in \text{Im}(T)$. Then there exists $f \in \mathbb{Q}[x]$ so that $f(\sqrt{3}) = \alpha$. Write $f = q(x^2 - 3) + r$ with $\deg r < 2$. Then

$$\alpha = f(\sqrt{3}) = q(\sqrt{3})(\sqrt{3})^2 - 3) + r(\sqrt{3}) = r(\sqrt{3}).$$

Since $\deg r < 2$, we have $\alpha = r(\sqrt{3}) = a + b\sqrt{3}$ for some $a,b \in \mathbb{Q}$. Thus, $\text{Im}(T) = \mathbb{Q}(\sqrt{3})$.

### 2.2 Bases and dimension

One of the most important features of vector spaces is the fact that they have bases. This essentially means that one can always find a nice subset (in many interesting cases a finite set) that will completely describe the space. Many of the concepts of this section are likely familiar from undergraduate linear algebra.

Throughout this section $V$ denotes an $F$-vector space unless indicated otherwise.
Definition 2.2.1. Let $\mathcal{B} = \{v_i\}_{i \in I}$ be a subset of $V$. We say $v \in V$ is an $F$-linear combination of $\mathcal{B}$ and write $v \in \text{span}_F \mathcal{B}$ if there exists a finite collection $\{c_1, \ldots, c_n\} \subset F$ such that

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

Definition 2.2.2. Let $\mathcal{B} = \{v_i\}_{i \in I}$ be a subset of $V$. We say $\mathcal{B}$ is $F$-linearly independent if whenever we have a finite linear combination $\sum c_i v_i = 0$ for some $c_i \in F$ we must have $c_i = 0$ for all $i$.

Definition 2.2.3. Let $\mathcal{B} = \{v_i\}$ be a subset of $V$. We say $\mathcal{B}$ is an $F$-basis of $V$ if

1. $\text{span}_F \mathcal{B} = V$;
2. $\mathcal{B}$ is linearly independent.

As is usual, if $F$ is clear from context we drop it from the notation and simply say “linear combination”, “linearly independent”, and “basis”.

The first goal of this section is to show every vector space has a basis. We will need Zorn’s lemma to prove this. We recall Zorn’s lemma for the convenience of the reader.

Theorem 2.2.4 (Zorn’s Lemma). Let $X$ be any partially ordered set with the property that every chain in $X$ has an upper bound. Then $X$ contains at least one maximal element.

Theorem 2.2.5. Let $V$ be a vector space. Let $\mathcal{A} \subseteq \mathcal{C}$ be subsets of $V$. Furthermore, assume $\mathcal{A}$ linearly independent and $\mathcal{C}$ spans $V$. Then there exists a basis $\mathcal{B}$ of $V$ satisfying $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$. In particular, if we set $\mathcal{A} = \emptyset$, $\mathcal{C} = V$, then this says that $V$ has a basis.

Proof. Let

$$X = \{\mathcal{B}' \subseteq V : \mathcal{A} \subseteq \mathcal{B}' \subseteq \mathcal{C}, \mathcal{B}' \text{ is linearly independent}\}.$$ 

We have $X$ is a partially ordered set under inclusion and is nonempty because $\mathcal{A} \in X$. We also have that $\mathcal{C}$ provides an upper bound on $X$. This allows us to apply Zorn’s Lemma to the chain $X$ to conclude it has a maximal element, say $\mathcal{B}$. If $\text{span}_F \mathcal{B} = V$, we are done. Suppose $\text{span}_F \mathcal{B} \neq V$. Then there exists $v \in \mathcal{C}$ such that $v \notin \text{span}_F \mathcal{B}$. However, this gives $\mathcal{B}' = \mathcal{B} \cup \{v\}$ is an element of $X$ that properly contains $\mathcal{B}$, a contradiction. Thus, $\mathcal{B}$ is the desired basis. 

One should note that the above proof gives that every vector space has a basis, but it does not provide an algorithm for constructing a basis. We will come back to this issue momentarily. We first deal with the easier case that there is a finite collection of vectors $\mathcal{B}$ so that $\text{span}_F \mathcal{B} = V$. In this case we say $V$ is a finite dimensional $F$-vector space. We will make use of the following fact from undergraduate linear algebra.
2.2. BASES AND DIMENSION

Lemma 2.2.6. ([3, Theorem 1.1]) A homogeneous system of \( m \) linear equations in \( n \) unknowns with \( m < n \) always has nontrivial solutions.

Corollary 2.2.7. Let \( B \subseteq V \) be such that \( \text{span}_F B = V \) and \( \#B = m \). Then any set with more than \( m \) elements cannot be linearly independent.

Proof. Let \( C = \{w_1, \ldots, w_n\} \) with \( n > m \). Let \( B = \{v_1, \ldots, v_m\} \) be a spanning set for \( V \). For each \( i \) write

\[
  w_i = \sum_{j=1}^{m} a_{ji}v_j.
\]

Consider the equation

\[
  \sum_{i=1}^{n} a_{ji}x_i = 0.
\]

The previous lemma gives a solution \((c_1, \ldots, c_n)\) to this equation with \((c_1, \ldots, c_n) \neq (0, \ldots, 0)\). We have

\[
  0 = \sum_{j=1}^{m} \left( \sum_{i=1}^{n} a_{ji}c_i \right) v_j
  = \sum_{i=1}^{n} c_i \left( \sum_{j=1}^{m} a_{ji}v_j \right)
  = \sum_{i=1}^{n} c_i w_i.
\]

This shows \( C \) is not linearly independent.

Theorem 2.2.8. Let \( V \) be a finite dimensional \( F \)-vector space. Any two bases of \( V \) have the same number of elements.

Proof. Let \( B \) and \( C \) be bases of \( V \). Suppose that \( \#B = m \) and \( \#C = n \). Since \( B \) is a basis, it is a spanning set and since \( C \) is a basis it is linearly independent. The previous result now gives so \( n \leq m \). We now reverse the roles of \( B \) and \( C \) to get the other direction.

Now that we have shown any two bases of a finite dimensional vector space have the same number of elements we can make the following definition.

Definition 2.2.9. Let \( V \) be a finite dimensional vector space. The number of elements in a \( F \)-basis of \( V \) is called the \( F \)-dimension of \( V \) and written \( \dim_F V \).

Note that if \( V \) is not finite dimensional we will write \( \dim_F V = \infty \). The notion of basis is not very useful in this context, as we will see below, so we do not spend much time on it here.

The following theorem does not contain anything new, but it is useful to summarize some of the important facts to this point.
Theorem 2.2.10. Let $V$ be a finite dimensional $F$-vector space with $\dim_F V = n$. Let $C \subseteq V$ with $\# C = m$.

1. If $m > n$, then $C$ is not linearly independent.
2. If $m < n$, then $\text{span}_F C \neq V$.
3. If $m = n$, then the following are equivalent:
   - $C$ is a basis;
   - $C$ is linearly independent;
   - $C$ spans $V$.

This theorem immediately gives the following corollary.

Corollary 2.2.11. Let $W \subseteq V$ be a subspace. Then $\dim_F W \leq \dim_F V$. If $\dim_F V < \infty$, then $V = W$ if and only if $\dim_F V = \dim_F W$.

We now give several examples of bases of some familiar vector spaces. It is left as an exercise to check these are actually bases.

Example 2.2.12. Let $V = F^n$. Set $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$. Then $E = \{e_1, \ldots, e_n\}$ is a basis for $V$ and is often referred to as the standard basis. We have $\dim_F F^n = n$.

Example 2.2.13. Let $V = \mathbb{C}$. From the previous example we have $B = \{1\}$ is a basis of $V$ as a $\mathbb{C}$-vector space and $\dim_{\mathbb{C}} \mathbb{C} = 1$. We saw before that $\mathbb{C}$ is also a $\mathbb{R}$-vector space. In this case a basis is given by $C = \{1, i\}$ and $\dim_{\mathbb{R}} \mathbb{C} = 2$.

Example 2.2.14. Let $V = \text{Mat}_{m,n}(F)$. Let $e_{i,j}$ to be the matrix with 1 in $(i,j)$th position and zeros elsewhere. Then $B = \{e_{i,j}\}$ is a basis for $V$ over $F$ and $\dim_F \text{Mat}_{m,n}(F) = mn$.

Example 2.2.15. Set $V = \mathfrak{sl}_2(\mathbb{C})$ where

$$
\mathfrak{sl}_2(\mathbb{C}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_2(\mathbb{C}) : a + d = 0 \right\}.
$$

One can check this is a $\mathbb{C}$-vector space. Moreover, it is a proper subspace because $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is not in $\mathfrak{sl}_2(\mathbb{C})$. Set

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

It is easy to see that $B$ is linearly independent. We know that $\dim_{\mathbb{C}} \mathfrak{sl}_2(\mathbb{C}) < 4$ because it is a proper subspace of $\text{Mat}_2(\mathbb{C})$. This gives $\dim_{\mathbb{C}} \mathfrak{sl}_2(\mathbb{C}) = 3$ and $B$ is a basis. The vector space $\mathfrak{sl}_2(\mathbb{C})$ is actually the Lie algebra associated to the Lie group $\text{SL}_2(\mathbb{C})$. There is a very rich theory of Lie algebras and Lie groups, but this is too far afield to delve into here. (Note the algebra multiplication on $\mathfrak{sl}_2(\mathbb{C})$ is not matrix multiplication, it is given by $X \cdot Y = [X, Y] = XY - YX$.)
Example 2.2.16. Let \( f(x) \in F[x] \) be an polynomial of degree \( n \). We can use this polynomial to split \( F[x] \) into equivalence classes analogously to how one creates the field \( \mathbb{F}_p \). For details of this construction, please see Example A.0.25 found in Appendix A. This is an \( F \)-vector space under addition given by \([g(x)] + [h(x)] := [g(x) + h(x)]\) and scalar multiplication given by \([c]g(x) := cg(x)\).

Note that given any \( g(x) \in F[x] \) we can use the Euclidean algorithm on \( F[x] \) (polynomial long division) to find unique polynomials \( q(x) \) and \( r(x) \) satisfying \( g(x) = f(x)q(x) + r(x) \) where \( r(x) = 0 \) or \( \deg r < \deg f \). This shows that each nonzero equivalence class has a representative \( r(x) \) with \( \deg r < \deg f \). Thus, we can write

\[
F[x]/(f(x)) = \{ [r(x)] : r(x) \in F[x] \text{ with } r = 0 \text{ or } \deg r < \deg f \}.
\]

From this one can see that a spanning set for \( F[x]/(f(x)) \) is given by \{\([1],[x],...,[x^{n-1}]\)\}.

This is also seen to be linearly independent by observing if there exists \( a_0,...,a_{n-1} \in F \) so that \([a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}] = 0\), this means \( f(x) \) divides \( a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} \). However, this is impossible unless \( a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} = 0 \) because \( \deg f = n \). Thus \( a_0 = a_1 = \cdots = a_{n-1} = 0 \). This gives that \( F[x]/(f(x)) \) is an \( F \)-vector space of degree \( n \).

Exercise 2.2.17. Show that \( \mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C} \) as \( \mathbb{R} \)-vector spaces.

The following lemma gives an alternate way to check if a set is a basis.

Lemma 2.2.18. Let \( V \) be an \( F \)-vector space and let \( C = \{ v_j \} \) be a subset of \( V \). We have \( C \) is a basis of \( V \) if and only if every vector in \( V \) can be written uniquely as a linear combination of elements in \( C \).

Proof. First suppose \( C \) is a basis. Let \( v \in V \). Since \( \text{span}_F C = V \), we can write any vector as a linear combination of elements in \( C \). Suppose we can write \( v = \sum a_i v_i = \sum b_i v_i \) for some \( a_i, b_i \in F \). Subtracting these equations gives \( 0 = \sum (a_i - b_i) v_i \). We now use that the \( v_i \) are linearly independent to conclude \( a_i - b_i = 0 \), i.e., \( a_i = b_i \) for every \( i \).

Now suppose every \( v \in V \) can be written uniquely as a linear combination of the elements in \( C \). This immediately gives \( \text{span}_F C = V \). It remains to show \( C \) is linearly independent. Suppose there exists \( a_i \in F \) with \( \sum a_i v_i = 0 \). We also have \( \sum 0 v_i = 0 \), so the uniqueness gives \( a_i = 0 \) for every \( i \), i.e., \( C \) is linearly independent.

Before we go further, we briefly delve into these concepts for infinite dimensional spaces. We begin with two familiar examples. The first works out exactly as one would hope; the second shows things are not as nice in general.

Example 2.2.19. Consider the vector space \( F[x] \). This cannot be a finite dimensional vector space. For instance, if \( \{ f_1, \ldots, f_n \} \) were a basis, the element \( x^{M+1} \) for \( M = \max_{1 \leq j \leq n} \deg f_j \) would not be in the span of these vectors. We can find a basis for this space though. Consider the collection \( B = \{ 1, x, x^2, \ldots \} \). It is clear this set is linearly independent and spans \( F[x] \), thus it forms a basis.
Example 2.2.20. Recall the vector space $V = \mathbb{R}^N$ defined earlier. This can be identified with sequences $\{a_n\}$ of real numbers. One might be interested in a basis for this vector space. At first glance the most obvious choice would be $\mathcal{E} = \{e_1, e_2, \ldots\}$. However, it is immediate that this set does not span $V$ as $v = (1, 1, \ldots)$ cannot be represented as a finite linear combination of these elements. Now we know since $v$ is not in span$_\mathbb{R} \mathcal{E}$, that $\mathcal{E} \cup \{v\}$ is a linearly independent set. However, it is clear this does not span either as $(1, 2, 3, 4, \ldots)$ is not in the span of this set. We know that $V$ has a basis, but it can be shown that no countable collection of vectors forms a basis for this space. Thus, one cannot construct a basis of this space by adding one vector at a time. The next thing one might try to do is to allow oneself to add infinitely many vectors. However, without some notion of convergence this does not make sense. For instance, how would one define $(1, 1, \ldots) + (2, 2, \ldots) + (3, 3, \ldots) + \cdots$?

The previous example shows that while we know every vector space has a basis, it may not be practical to construct such a basis. In fact, this definition of basis is not very useful for infinite dimensional spaces and is given the name Hamel basis since other more useful concepts are often referred to as bases in this setting. We will not say more about other notions here as these are more appropriate for a functional analysis course. We will deal mostly with finite dimensional vector spaces in this course.

The following proposition will be used repeatedly throughout the course. It says that a linear transformation between vector spaces is completely determined by what it does to a basis. One way we will often use this is to define a linear transformation by only specifying what it does to a basis. This provides another important application of the fact that vector spaces are completely determined by their bases.

Proposition 2.2.21. Let $V, W$ be vector spaces.

1. Let $T \in \text{Hom}_F(V, W)$. Then $T$ is determined by its values on a basis of $V$.

2. Let $B = \{v_i\}$ be a basis of $V$ and $C = \{w_i\}$ be any collection of vectors in $W$ so that $\#B = \#C$. There is a unique linear transformation $T \in \text{Hom}_F(V, W)$ satisfying $T(v_i) = w_i$.

Proof. Let $B = \{v_i\}$ be a basis of $V$. Given any $v \in V$ there are elements $a_i \in F$ so that $v = \sum a_i v_i$. We have

\[
T(v) = T \left( \sum a_i v_i \right) \\
= \sum T(a_i v_i) \\
= \sum a_i T(v_i).
\]

Thus, if one knows the elements $T(v_i)$, one knows $T(v)$ for any $v \in V$. This gives the first claim.
The second part follows immediately from the first. Set \( T(v_i) = w_i \) for each \( i \). For \( v = \sum a_i v_i \in V \), define \( T(v) \) by
\[
T(v) = \sum a_i w_i.
\]
It is now easy to check this is a linear map from \( V \) to \( W \) and is unique. \( \square \)

**Example 2.2.22.** Let \( V = W = \mathbb{R}^2 \). It is easy to check that \( \mathcal{B} = \left\{ v_1 = \left( \begin{array}{c} 1 \\ 1 \end{array} \right), v_2 = \left( \begin{array}{c} -1 \\ 1 \end{array} \right) \right\} \) is a basis of \( V \). The previous results says to define a map from \( V \) to \( W \) it is enough to say where to send \( v_1 \) and \( v_2 \). For instance, let \( \left\{ (2,3), (5,-10) \right\} \) be a subset of \( W \). Then we have a unique linear map \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) given by \( T(v_1) = w_1 \), \( T(v_2) = w_2 \). Note it is exactly this property that allows us to represent a linear map \( T : \mathbb{F}^n \to \mathbb{F}^m \) as a matrix.

**Corollary 2.2.23.** Let \( T \in \text{Hom}_F(V,W) \), \( \mathcal{B} = \{ v_i \} \) be a basis of \( V \), and \( \mathcal{C} = \{ w_i = T(v_i) \} \subseteq W \). Then \( \mathcal{C} \) is a basis for \( W \) if and only if \( T \) is an isomorphism.

**Proof.** Suppose \( \mathcal{C} \) is a basis for \( W \). The previous result allows us to define \( S : W \to V \) such that \( S(w_i) = v_i \). This is an inverse for \( T \), so \( T \) is an isomorphism.

Now suppose \( T \) is an isomorphism. We need to show that \( \mathcal{C} \) spans \( W \) and it is linearly independent. Let \( w \in W \). Since \( T \) is an isomorphism there exists a \( v \in V \) with \( T(v) = w \). Using that \( \mathcal{B} \) is a basis of \( V \) we can write \( v = \sum a_i v_i \) for some \( a_i \in \mathbb{F} \). Applying \( T \) to \( v \) we have
\[
w = T(v) = T \left( \sum a_i v_i \right) = \sum a_i T(v_i) = \sum a_i w_i.
\]
Thus, \( w \in \text{span}_F \mathcal{C} \) and since \( w \) was arbitrary we have \( \text{span}_F \mathcal{C} = W \). Suppose there exists \( a_i \in \mathbb{F} \) with \( \sum a_i w_i = 0 \). We have
\[
T(0) = 0 = \sum a_i w_i = \sum a_i T(v_i) = \sum T(a_i v_i) = T \left( \sum a_i v_i \right).
\]
Applying that \( T \) is injective we have \( 0 = \sum a_i v_i \). However, \( \mathcal{B} \) is a basis so it must be the case that \( a_i = 0 \) for all \( i \). This gives \( \mathcal{C} \) is linearly independent and so a basis. \( \square \)
2.2. BASES AND DIMENSION

The following result ties the dimension of the kernel of a linear transformation, the dimensions of the image, and the dimension of the domain vector space. This is extremely useful as one often knows (or can bound) two of the three.

**Theorem 2.2.24.** Let \( V \) be a finite dimensional vector space and let \( T \in \text{Hom}_F(V,W) \). Then

\[
\dim_F \ker T + \dim_F \text{Im} T = \dim_F V.
\]

**Proof.** Let \( \dim_F \ker T = k \) and \( \dim_F V = n \). Let \( \mathcal{A} = \{v_1, \ldots, v_k\} \) be a basis of \( \ker T \) and extend this to a basis \( \mathcal{B} = \{v_1, \ldots, v_n\} \) of \( V \). It is enough to show that \( \mathcal{C} = \{T(v_{k+1}), \ldots, T(v_n)\} \) is a basis for \( \text{Im} T \).

Let \( w \in \text{Im} T \). There exists \( v \in V \) so that \( T(v) = w \). Since \( \mathcal{B} \) is a basis for \( V \) there exists \( a_i \) such that \( v = \sum_{i=1}^{n} a_i v_i \). This gives

\[
w = T(v) = T \left( \sum_{i=1}^{n} a_i v_i \right) = \sum_{i=1}^{n} a_i T(v_i) = \sum_{i=k+1}^{n} a_i T(v_i)
\]

where we have used \( T(v_i) = 0 \) for \( i = 1, \ldots, k \) because \( \mathcal{A} \) is a basis for \( \ker T \). Thus, \( \text{span}_F \mathcal{C} = \text{Im} T \). It remains to show \( \mathcal{C} \) is linearly independent. Suppose there exists \( a_i \in F \) such that \( \sum_{i=k+1}^{n} a_i T(v_i) = 0 \). Note

\[
0 = \sum_{i=k+1}^{n} T(a_i v_i) = T \left( \sum_{i=k+1}^{n} a_i v_i \right).
\]

Thus, \( \sum_{i=k+1}^{n} a_i v_i \in \ker T \). However, \( \mathcal{A} \) spans \( \ker T \) so there exists \( a_1, \ldots, a_k \) in \( F \) such that

\[
\sum_{i=k+1}^{n} a_i v_i = \sum_{i=1}^{k} a_i v_i,
\]

20
2.2. BASES AND DIMENSION

i.e.,
\[ \sum_{i=k+1}^{n} a_i v_i + \sum_{i=1}^{k} (-a_i)v_i = 0. \]

Since \( B = \{v_1, \ldots, v_n\} \) is a basis of \( V \) we must have \( a_1 = \cdots = a_k = -a_{k+1} = \cdots = -a_n = 0 \). In particular, \( a_{k+1} = \cdots = a_n = 0 \) as desired.

This theorem allows to prove the following important results.

**Corollary 2.2.25.** Let \( V, W \) be \( F \)-vector spaces with \( \dim_F V = n \). Let \( V_1 \subseteq V \) be a subspace of dimension \( k \) and \( W_1 \subseteq W \) be a subspace of dimension \( n - k \). Then there exists \( T \in \text{Hom}_F(V,W) \) such that \( V_1 = \ker T \) and \( W_1 = \text{Im} T \).

**Proof.** Let \( B = \{v_1, \ldots, v_k\} \) be a basis of \( V_1 \). Extend this to a basis \( \{v_1, \ldots, v_k, \ldots, v_n\} \) of \( V \). Let \( C = \{w_{k+1}, \ldots, w_n\} \) be a basis of \( W_1 \). Define \( T \) by \( T(v_1) = \cdots = T(v_k) = 0 \) and \( T(v_{k+1}) = w_{k+1}, \ldots, T(v_n) = w_n \). This is the required linear map.

The previous corollary says the only limitation on a subspace being the kernel or image of a linear transformation is that the dimensions add up properly. One should contrast this to the case of homomorphisms in group theory for example. There in order to be a kernel one requires the subgroup to satisfy the further property of being a normal subgroup. This is another way in which vector spaces are very nice to work with.

The following corollary follows immediately from Theorem 2.2.24. This corollary makes checking a map between two vector spaces of the same dimension is an isomorphism much easier as one only needs to check the map is injective or surjective, not both.

**Corollary 2.2.26.** Let \( T \in \text{Hom}_F(V,W) \) and \( \dim_F V = \dim_F W \). Then the following are equivalent:

1. \( T \) is an isomorphism;
2. \( T \) is surjective;
3. \( T \) is injective.

We can also rephrase this result in terms of matrices.

**Corollary 2.2.27.** Let \( A \in \text{Mat}_n(F) \). Then the following are equivalent:

1. \( A \) is invertible;
2. There exists \( B \in \text{Mat}_n(F) \) with \( BA = 1_n \);
3. There exists \( B \in \text{Mat}_n(F) \) with \( AB = 1_n \).

One should prove the previous two corollaries as well as the following corollary as exercises.
Corollary 2.2.28. Let $V, W$ be $F$-vector spaces and let $\dim_F V = m$, $\dim_F W = n$.

1. If $m < n$ and $T \in \text{Hom}_F(V, W)$, then $T$ is not surjective.

2. If $m > n$ and $T \in \text{Hom}_F(V, W)$, then $T$ is not injective.

3. We have $m = n$ if and only if $V \cong W$. In particular, $V \cong F^n$.

Note that while it is true that if $\dim_F V = \dim_F W = n < \infty$ then $V \cong W$, it is not the case that every linear map from $V$ to $W$ is an isomorphism. This result is only saying there is a map $T : V \to W$ that is an isomorphism. Clearly we can define a linear map $T : V \to W$ by $T(v) = 0_W$ for all $v \in V$ and this is not an isomorphism.

The previous corollary gives a very important fact: if $V$ is an $n$-dimensional $F$-vector space, then $V \cong F^n$. This result gives that for any positive integer $n$, all $F$-vector spaces of dimension $n$ are isomorphic. One obtain this isomorphism by choosing a basis. This is why in undergraduate linear algebra one often focuses almost exclusively on the vector spaces $F^n$ and matrices as the linear transformations.

The following example gives a nice application of what we have studied thus far.

Example 2.2.29. Recall $P_n(\mathbb{R})$ is the vector space of polynomials of degree less than or equal to $n$ and $\dim_\mathbb{R} P_n(\mathbb{R}) = n + 1$. Set $V = P_{n-1}(\mathbb{R})$. Let $a_1, \ldots, a_k \in \mathbb{R}$ be distinct and pick $m_1, \ldots, m_k \in \mathbb{Z}_{\geq 0}$ such that $\sum_{j=1}^{k} (m_j + 1) = n$.

Our goal is to show given any real numbers $b_{1,0}, \ldots, b_{1,m_1}, \ldots, b_{k,m_k}$ there is a unique polynomial $f \in P_{n-1}(\mathbb{R})$ satisfying $f^{(j)}(a_i) = b_{i,j}$.

Define

$$T : P_{n-1}(\mathbb{R}) \to \mathbb{R}^n$$

$$f(x) \mapsto \begin{pmatrix}
  f(a_1) \\
  \vdots \\
  f^{(m_1)}(a_1) \\
  \vdots \\
  f(a_k) \\
  \vdots \\
  f^{(m_k)}(a_k)
\end{pmatrix}.$$ 

If $f \in \ker T$, then for each $i = 1, \ldots, k$ we have $f^{(j)}(a_i) = 0$ for $j = 1, \ldots, m_i$. Thus, for each $i$ we have $(x - a_i)^{m_i + 1} | f(x)$. Since these polynomials are relatively prime, this gives their product divides $f$ and thus $f$ is divisible by a polynomial of degree $n$. Since $f \in P_{n-1}(F)$ this implies $f = 0$. Hence $\ker T = 0$.

Applying Theorem 2.2.24 we have $\text{Im} T$ must have dimension $n$, i.e., $\text{Im} T = \mathbb{R}^n$.

This gives the result.
2.3 Direct sums and quotient spaces

In this section we cover two of the basic constructions in vector space theory, namely the direct sum and the quotient space. The direct sum is a way to split a vector space into subspaces that are "independent." The quotient space construction is a way to identify some vectors to be 0. We begin with the direct sum.

**Definition 2.3.1.** Let $V$ be an $F$-vector space and $V_1, \ldots, V_k$ be subspaces of $V$. The sum of $V_1, \ldots, V_k$ is defined by

$$V_1 + \cdots + V_k = \{v_1 + \cdots + v_k : v_i \in V_i\}.$$  

**Definition 2.3.2.** Let $V_1, \ldots, V_k$ be subspaces of $V$. We say $V_1, \ldots, V_k$ are independent if whenever $v_1 + \cdots + v_k = 0$ with $v_i \in V_i$, then $v_i = 0$ for all $i$.

**Definition 2.3.3.** Let $V_1, \ldots, V_k$ be subspaces of $V$. We say $V$ is the direct sum of $V_1, \ldots, V_k$ and write

$$V = V_1 \oplus \cdots \oplus V_k$$

if the following two conditions are satisfied

1. $V = V_1 + \cdots + V_k$;
2. $V_1, \ldots, V_k$ are independent.

**Example 2.3.4.** Set $V = F^2$, $V_1 = \{(x,0) : x \in F\}$, $V_2 = \{(0,y) : y \in F\}$. Then we clearly have $V_1 + V_2 = \{(x,y) : x, y \in F\} = V$. Moreover, if $(x,0) + (0,y) = (0,0)$, then $(x,y) = (0,0)$ and so $x = y = 0$. This gives $(x,0) = (0,0) = (0,y)$ and so $V_1$ and $V_2$ are independent. Thus $F^2 = V_1 \oplus V_2$.

**Example 2.3.5.** Let $B = \{v_1, \ldots, v_n\}$ be a basis of $V$. Set $V_i = F \cdot v_i = \{av_i : a \in F\}$. We have $V_1, \ldots, V_n$ are clearly subspaces of $V$, and by the definition of a basis we obtain $V = V_1 \oplus \cdots \oplus V_n$.

**Lemma 2.3.6.** Let $V$ be a vector space, $V_1, \ldots, V_k$ be subspaces of $V$. We have $V = V_1 \oplus \cdots \oplus V_k$ if and only if each $v \in V$ can be written uniquely in the form $v = v_1 + \cdots + v_k$ with $v_i \in V_i$.

**Proof.** Suppose $V = V_1 \oplus \cdots \oplus V_k$. Then certainly we have $V = V_1 + \cdots + V_k$ and so given any $v \in V$ there are elements $v_i \in V_i$ so that $v = v_1 + \cdots + v_k$. The only thing to show is this expression is unique. Suppose $v = v_1 + \cdots + v_k = w_1 + \cdots + w_k$ with $v_i, w_i \in V_i$. Then $0 = (v_1 - w_1) + \cdots + (v_k - w_k)$. Since the $V_i$ are independent we have $v_i - w_i = 0$, i.e., $v_i = w_i$ for all $i$.

Suppose each $v \in V$ can be written uniquely in the form $v = v_1 + \cdots + v_k$ with $v_i \in V_i$. Immediately we get $V = V_1 + \cdots + V_k$. Suppose $0 = v_1 + \cdots + v_k$ with $v_i \in V_i$. We have $0 = 0 + \cdots + 0$ as well, so by uniqueness, we get $v_i = 0$ for all $i$. This gives $V_1, \ldots, V_k$ are independent and so $V = V_1 \oplus \cdots \oplus V_k$.  

\[\Box\]
Exercise 2.3.7. Let \( V_1, \ldots, V_k \) be subspaces of \( V \). For each \( i \) let \( B_i \) be a basis of \( V_i \). Set \( B = B_1 \cup \cdots \cup B_k \). Then

1. \( \mathcal{B} \) spans \( V \) if and only if \( V = V_1 + \cdots + V_k \);
2. \( \mathcal{B} \) is linearly independent if and only if \( V_1, \ldots, V_k \) are independent;
3. \( \mathcal{B} \) is a basis if and only if \( V = V_1 \oplus \cdots \oplus V_k \).

Given a subspace \( U \subset V \), it is natural to ask if there is a subspace \( W \) so that \( V = U \oplus W \). This leads us to the following definition.

Definition 2.3.8. Let \( V \) be a vector space and \( U \subseteq V \) a subspace. We say \( W \subseteq V \) is a complement of \( U \) in \( V \) if \( V = U \oplus W \).

Lemma 2.3.9. Let \( U \subseteq V \) be a subspace. Then \( U \) has a complement.

Proof. Let \( \mathcal{A} \) be a basis of \( U \), and extend \( \mathcal{A} \) to a basis of \( \mathcal{B} \) of \( V \). Set \( W = \text{span}_F(\mathcal{B} - \mathcal{A}) \). One checks immediately that \( V = U \oplus W \).

Exercise 2.3.10. Let \( U \subset V \) be a subspace. Is the complement of \( U \) unique? If so, prove it. If not, give a counterexample.

We now turn our attention to quotient spaces. We have already seen an example of a quotient space. Namely, recall Example 2.2.16. Consider \( V = F[x] \) and \( W = (f(x)) := \{ g \in F[x] : f|g = 0 \} \). One can check that \( W \) is a subspace of \( V \). In that example we defined a vector space \( V/W = F[x]/(f(x)) \) and saw that the elements in \( W \) become 0 in the new space. This construction is the one we generalize to form quotient spaces.

Let \( V \) be a vector space and \( W \subseteq V \) be a subspace. Define an equivalence relation on \( V \) as follows \( v_1 \sim_W v_2 \) if and only if \( v_1 - v_2 \in W \). We write the equivalence classes as \( [v_1] = \{ v_2 \in V : v_1 - v_2 \in W \} = v_1 + W \).

Set \( V/W = \{ v + W : v \in V \} \). Addition and scalar multiplication on \( V/W \) are defined as follows. Let \( v_1, v_2 \in V \) and \( c \in F \). Define

\[
(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W;
\]  
\[
c(v_1 + W) = cv_1 + W.
\]

Exercise 2.3.11. Show that \( V/W \) is an \( F \)-vector space.

We call \( V/W \) the quotient space of \( V \) by \( W \).

Example 2.3.12. Let \( V = \mathbb{R}^2 \) and \( W = \{(x,0) : x \in \mathbb{R}\} \). Clearly we have \( W \subseteq V \) is a subspace. Let \( (x_0, y_0) \in V \). To find \( (x_0, y_0) + W \), we want all \( (x, y) \) such that \( (x_0, y_0) - (x, y) \in W \). However, it is clear that \( (x_0 - x, y_0 - y) \in W \) if and only if \( y = y_0 \). Thus, \( (x_0, y_0) + W = \{(x, y) : x \in \mathbb{R}\} \). The graph below gives the elements \((0,0) + W \) and \((0,1) + W \).
One immediately sees from this that \((x, y) + W\) is not a subspace of \(V\) unless 
\(x, y) + W = (0, 0) + W\).

Define

\[ \pi : \mathbb{R} \rightarrow V/W \]

\[ y_0 \mapsto (x_0, y_0). \]

It is straightforward to check this is an isomorphism, so \(V/W \cong \mathbb{R}\).

**Example 2.3.13.** More generally, let \(m, n \in \mathbb{Z}_{>0}\) with \(m > n\). Consider \(V = \mathbb{F}^m\) and let \(W\) be the subspace of \(V\) spanned by \(e_1, \ldots, e_n\) with \(\{e_1, \ldots, e_m\}\) the standard basis. We can form the quotient space \(V/W\). This space is isomorphic to \(\mathbb{F}^{m-n}\) with a basis given by \(\{e_{n+1} + W, \ldots, e_{m} + W\}\).

**Example 2.3.14.** Let \(V = \mathbb{F}[x]\) and let \(W = (f(x))\). We saw before that the quotient space \(V/W = \mathbb{F}[x]/(f(x))\) has as a basis \(\{[1], [x], \ldots, [x^{n-1}]\}\). Define \(T : V/W \rightarrow P_{n-1}(F)\) by \(T([x^j]) = x^j\). One can check this is an isomorphism, and so \(\mathbb{F}[x]/(x^n) \cong P_{n-1}(F)\) as \(\mathbb{F}\)-vector spaces.

**Definition 2.3.15.** Let \(W \subseteq V\) be a subspace. The **canonical projection map** is given by

\[ \pi_W : V \rightarrow V/W \]

\[ v \mapsto v + W. \]

It is immediate that \(\pi_W \in \text{Hom}_F (V, V/W)\).

One important point to note is that when working with quotient spaces, if one defines a map from \(V/W\) to another vector space, one must always check the map is well-defined as defining the map generally involves a choice of representative for \(v + W\). In other words, one must show if \(v_1 + W = v_2 + W\) then \(T(v_1 + W) = T(v_2 + W)\). Consider the following example.

**Example 2.3.16.** Let \(V = \mathbb{R}^2\) and \(W = \{(x, 0) : x \in \mathbb{R}\}\). We saw above the elements of the quotient space \(V/W\) are of the form \((x, y) + W\) and \((x_1, y_1) + W = (x_2, y_2) + W\) if and only if \(y_1 = y_2\). Suppose we with to define a linear map \(T : V/W \rightarrow \mathbb{R}\). We could try to define such a map by specifying \(T((x, y) + W) = x\).
2.3. DIRECT SUMS AND QUOTIENT SPACES

However, we know that \((x, y) + W = (x + 1, y) + W\), so \(x = T((x, y) + W) = T((x + 1, y) + W) = x + 1\). This doesn’t make sense, so our map is not well-defined. The “correct” map in this situation is to send \((x, y) + W\) to \(y\) since \(y\) is fixed across the equivalence class.

The following result allows us to avoid checking the map is well-defined when it is induced from another linear map. One should look back at the examples of quotient spaces above to see how this theorem applies.

**Theorem 2.3.17.** Let \(T \in \text{Hom}_F(V, W)\). Define

\[
\overline{T} : V/\ker T \to W \\
v + \ker T \mapsto T(v).
\]

Then \(\overline{T} \in \text{Hom}_F(V/\ker T, W)\). Moreover, \(\overline{T}\) gives an isomorphism

\[
V/\ker T \xrightarrow{\approx} \text{Im} T.
\]

**Proof.** The first step is to show \(\overline{T}\) is well-defined. Suppose \(v_1 + \ker T = v_2 + \ker T\), i.e., \(v_1 - v_2 \in \ker T\). So there exists \(x \in \ker T\) such that \(v_1 - v_2 = x\). We have

\[
\overline{T}(v_1 + \ker T) = T(v_1) \\
= T(v_2 + x) \\
= T(v_2) + T(x) \\
= T(v_2) \\
= \overline{T}(v_2 + \ker T).
\]

Thus, \(\overline{T}\) is well-defined.

The next step is to show \(\overline{T}\) is linear. Let \(v_1 + W, v_2 + W \in V/W\) and \(c \in F\). We have

\[
\overline{T}(v_1 + \ker T + v_2 + \ker T) = \overline{T}(v_1 + v_2 + \ker T) \\
= T(v_1 + v_2) \\
= T(v_1) + T(v_2) \\
= \overline{T}(v_1 + \ker T) + \overline{T}(v_2 + \ker T)
\]

and

\[
\overline{T}(c(v_1 + \ker T)) = \overline{T}(cv_1 + \ker T) \\
= T(cv_1) \\
= cT(v_1) \\
= c\overline{T}(v_1 + \ker T).
\]

Thus, \(\overline{T} \in \text{Hom}_F(V/\ker T, W)\).

It only remains to show that \(\overline{T}\) is a bijection. Let \(w \in \text{Im}(T)\). There exists \(v \in V\) so that \(T(v) = w\). Thus, \(\overline{T}(v + \ker T) = T(v) = w\), so \(\overline{T}\) is surjective. Now suppose \(v + \ker T \in \ker \overline{T}\). Then \(0 = \overline{T}(v + \ker T) = T(v)\). Thus, \(v \in \ker T\) which means \(v + \ker T = 0 + \ker T\) and so \(\overline{T}\) is injective. \(\square\)
Example 2.3.18. Let $V = F[x]$. Define a map $T : V \to P_2(F)$ by sending $f(x) = a_0 + a_1x + \cdots + a_nx^n$ to $a_0 + a_1x + a_2x^2$. This is a surjective linear map. The kernel of this map is exactly $(x^3) = \{g(x) : x^3 | g(x)\}$. Thus, the previous result gives an isomorphism $F[x]/(x^3) \cong P_2(F)$ as $F$-vector spaces.

Example 2.3.19. Let $V = \text{Mat}_2(F)$. Define a map $T : \text{Mat}_2(F) \to F^2$ by $T \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}$. One can check this is a surjective linear map. The kernel of this map is given by $\ker T = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} : a,d \in F \right\}$.

Thus, $\text{Mat}_2(F)/\ker T$ is isomorphic to $F^2$ as $F$-vector spaces.

Theorem 2.3.20. Let $W \subseteq V$ be a subspace. Let $B = \{b_i\}$ be a basis for $W$ and extend to a basis $B$ for $V$. Set $B_U = B - B_W = \{z_i\}$. Let $U = \text{span}_F B_U$, i.e., $U$ is a complement of $W$ in $V$. Then the linear map $p : U \to V/W$ $z_i \mapsto z_i + W$

is an isomorphism. Thus, $B_U = \{z_i + W : z_i \in B_U\}$ is a basis of $V/W$.

Proof. Note that $p$ is linear because we defined it on a basis. It only remains to show $p$ is an isomorphism. We will do this by showing $B_U = \{z_i + W : z_i \in B_U\}$ is a basis for $V/W$.

Let $v + W \in V/W$. Since $v \in V$, there exists $a_i, b_j \in F$ such that $v = \sum a_i v_i + \sum b_j z_j$. We have $\sum a_i v_i \in W$, so $v - \sum b_j z_j \in W$. Thus

$$v + W = \sum b_j z_j + W = \sum b_j (z_j + W).$$

This shows that $B_U$ spans $V/W$.

Suppose there exists $b_j \in F$ such that $\sum b_j (z_j + W) = 0 + W$, i.e., $\sum b_j (z_j + W) \in W$. So there exists $a_i \in F$ such that $\sum b_j z_j = \sum a_i v_i$. Thus $\sum a_i v_i + \sum -b_j z_j = 0$. Since $\{v_i, z_j\}$ is a basis of $V$ we obtain $a_i = 0 = b_j$ for all $i, j$. This gives $B_U$ is linearly independent and so completes the proof.

2.4 Dual spaces

We conclude this chapter by discussing dual spaces. Throughout this section $V$ is an $F$-vector space.

Definition 2.4.1. The dual space of $V$, denoted $V^\vee$, is given by $V^\vee = \text{Hom}_F(V, F)$. Elements of the dual space are called linear functionals.
2.4. DUAL SPACES

Theorem 2.4.2. The vector space $V$ is isomorphic to a subspace of $V'$. If $\dim_F V < \infty$, then $V \cong V'$.

Proof. Let $B = \{v_i\}$ be a basis of $V$. For each $v_i$, define an element $v^*_i$ by setting

$$v^*_i(v_j) = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{otherwise}. \end{cases}$$

One sees immediately that $v^*_i \in V'$ as it is defined on a basis. Define $T \in \text{Hom}_F(V, V')$ by $T(v_i) = v^*_i$. We claim $T$ is an injective linear map. Let $v \in V$ and suppose $T(v) = 0$. Write $v = \sum a_i v_i$, so that $\sum a_i v^*_i$ is then the 0 map, i.e., for any $v' \in V$ this gives $\sum a_i v^*_i(v') = 0$. In particular,

$$0 = \sum a_i v^*_i(v_j) = a_j v^*_j(v_j) = a_j.$$

Since $a_j = 0$ for all $j$, we have $v = 0$ and so $T$ is injective. We now use the fact that $V / \ker T \cong \text{Im} T$ and the fact that $\ker T = 0$ to conclude that $V$ is isomorphic to a subspace of $W$, which gives the first statement of the theorem.

Assume $\dim_F V < \infty$ so we can write $B = \{v_1, \ldots, v_n\}$. Given $v^* \in V'$, define $a_j = v^*(v_j)$. Set $v = \sum a_i v_i \in V$. Define $S : V' \to V$ by $v^* \to v$. This map defines an inverse to the map $T$ given above and thus $V \cong V'$.

Note it is not always the case that $V$ is isomorphic to its dual. In fact, if $V$ is infinite dimensional it is never the case that $V$ is isomorphic to its dual. In functional analysis or topology this problem is often overcome by requiring the linear functionals to be continuous, i.e. the dual space consists of continuous linear maps from $V$ to $F$. In that case one would refer to our dual as the “algebraic dual.” However, we will only consider the algebraic setting here so we don’t bother with saying algebraic dual.

Example 2.4.3. Let $V$ be a vector space over a field $F$ and let the dimension be denoted by $\alpha$. Note that $\alpha$ is not finite by assumption, but we need to work with different cardinalities here so we must keep track of this. The cardinality of $V$ as a set is given by $\alpha \cdot \#F = \max\{\alpha, \#F\}$. Moreover, we have $V$ is naturally isomorphic to the set of functions from a set of cardinality $\alpha$ to $F$ with finite support. We denote this space by $F^{(\alpha)}$.

The dual space of $V$ is the set of all functions from a set of cardinality $\alpha$ to $F$, i.e., to $F^{\alpha}$. If we set $\alpha' = \dim_F V'$, we wish to show $\alpha' > \alpha$. As above, the cardinality of $V'$ as a set is $\max\{\alpha', \#F\}$.

Let $A = \{v_i\}$ be a countable linear independent subset of $V$ and extend it to a basis of $V$. For each nonzero $c \in F$ define $f_c : V \to F$ by $f_c(v_i) = c^i$ for $v_i \in A$ and 0 for the other elements in the basis. One can show that $\{f_c\}$ is linearly independent, so $\alpha' \geq \#F$. Thus, we have $\#V' = \alpha' \#F = \max\{\alpha', \#F\} = \alpha'$. However, we also have $\#V' = \#F^{\alpha}$. Since $\alpha < \#F^{\alpha}$ because $\#F \geq 2$, we have $\alpha' = \#F^{\alpha} > \alpha$ as desired.
One should note that in the finite dimensional case when we have $V \cong V^\vee$, the isomorphism depends upon the choice of a basis. This means that while $V$ is isomorphic to its dual, the isomorphism is non-canonical. In particular, there is no “preferred” isomorphism between $V$ and its dual. Studying the possible isomorphisms that arise between $V$ and its dual is interesting in its own right. We will return to this problem in Chapter 6.

**Definition 2.4.4.** Let $V$ be a finite dimensional $F$-vector space and let $B = \{v_1, \ldots, v_n\}$ be a basis of $V$. The dual basis of $V^\vee$ with respect to $B$ is given by $B^\vee = \{v_1^\vee, \ldots, v_n^\vee\}$.

If $V$ is finite dimensional then we have $V \cong V^\vee \cong (V^\vee)^\vee$, i.e., $V \cong (V^\vee)^\vee$. The major difference is that while there is no canonical isomorphism between $V$ and $V^\vee$, there is a canonical isomorphism $V \cong (V^\vee)^\vee$! Note that in proof below we construct the map from $V$ to $(V^\vee)^\vee$ without choosing any basis. This is what makes the map canonical. We do use a basis in proving injectivity, but the map does not depend on this basis so it does not matter.

**Proposition 2.4.5.** There is a canonical injective linear map from $V$ to $(V^\vee)^\vee$. If $\dim_F V < \infty$, then this is an isomorphism.

**Proof.** Let $v \in V$. Define $\text{eval}_v : V^\vee \to F$ by sending $\varphi \in \text{Hom}_F(V, F) = V^\vee$ to $\text{eval}_v(\varphi) = \varphi(v)$. One must check that $\text{eval}_v$ is a linear map. To see this, let $\varphi, \psi \in V^\vee$ and $c \in F$. We have

$$
\text{eval}_v(c\varphi + \psi) = (c\varphi + \psi)(v) = c\varphi(v) + \psi(v) = c\text{eval}_v(\varphi) + \text{eval}_v(\psi).
$$

Thus, for each $v \in V$ we obtain a map $\text{eval}_v \in \text{Hom}_F(V^\vee, F)$. This allows us to define a well-defined map

$$
\Phi : V \to \text{Hom}_F(V^\vee, F) = (V^\vee)^\vee
$$

$$
v \mapsto \text{eval}_v : \varphi \mapsto \varphi(v).
$$

We claim that $\Phi$ is a linear map. Since $\Phi$ maps into a space of maps, we check equality by checking the maps agree on each element, i.e., for $v, w \in V$ and $c \in F$ we want to show that for each $\varphi \in \text{Hom}_F(V^\vee, F)$ that $\Phi(cv + w)(\varphi) = c\Phi(v)(\varphi) + \Phi(w)(\varphi)$. Observe we have

$$
\Phi(cv + w)(\varphi) = \text{eval}_{cv+w}(\varphi) = \varphi(cv + w) = c\varphi(v) + \varphi(w) = c\Phi(v)(\varphi) + \Phi(w)(\varphi).
$$

Thus, we have $\Phi \in \text{Hom}_F(V, (V^\vee)^\vee)$.

It remains to show that $\Phi$ is injective. Let $v \in V$, $v \neq 0$. Let $B$ be a basis containing $v$. This is possible because we can start with the set $\{v\}$ and
complete it to a basis. Note $v^\vee \in V^\vee$ and $\text{eval}_v(v^\vee) = v^\vee(v) = 1$. Moreover, for any $w \in \mathcal{B}$ with $w \neq v$ we have $\text{eval}_w(v^\vee) = v^\vee(w) = 0$. Thus, we have $\Phi(v) = \text{eval}_v$ is not the 0 map, so $\ker \Phi = \{0\}$, i.e., $\Phi$ is injective.

If $\dim_F V < \infty$, we have $\Phi$ is an isomorphism because $\dim_F V = \dim_F V^\vee$ since they are isomorphic, so $\dim_F V = \dim_F V^\vee = \dim_F (V^\vee)^\vee$. 

Let $T \in \text{Hom}_F(V,W)$. We obtain a natural map $T^\vee \in \text{Hom}_F(W^\vee,V^\vee)$ as follows. Let $\varphi \in W^\vee$, i.e., $\varphi : W \to F$. To obtain a map $T^\vee(\varphi) : V \to F$, we simply compose the maps, namely, $T^\vee(\varphi) = \varphi \circ T$. It is easy to check this is a linear map. In particular, we have the following result.

**Proposition 2.4.6.** Let $T \in \text{Hom}_F(V,W)$. The map $T^\vee$ defined by $T^\vee(\varphi) = \varphi \circ T$ is a linear map from $W^\vee$ to $V^\vee$.

We will see in the next chapter how the dual map gives the proper definition of a transpose.
2.5 Problems

For these problems $F$ is assumed to be a field.

1. Define

$$\mathfrak{sl}_n(\mathbb{Q}) = \left\{ X = (x_{i,j}) \in \text{Mat}_n(\mathbb{Q}) : \text{Tr}(X) = \sum_{i=1}^{n} x_{i,i} = 0 \right\}.$$ 

Show that $\mathfrak{sl}_n(\mathbb{Q})$ is a $\mathbb{Q}$-vector space.

2. Consider the vector space $F^3$. Determine, and justify your answer, whether each of the following are subspaces of $F^3$:

   (a) $W_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + 2x_2 + 3x_3 = 0 \right\}$

   (b) $W_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1x_2x_3 = 0 \right\}$

   (c) $W_3 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 = 5x_3 \right\}$.

3. Let $V$ be an $F$-vector space.

   (a) Prove that an arbitrary intersection of subspaces of $V$ is again a subspace of $V$.

   (b) Prove that the union of two subspaces of $V$ is a subspace of $V$ if and only if one of the subspaces is contained in the other.

4. Let $T \in \text{Hom}_F(F,F)$. Prove there exists $\alpha \in F$ so that $T(v) = \alpha v$ for every $v \in F$.

5. Let $U, V$, and $W$ be $F$-vector spaces. Let $S \in \text{Hom}_F(U,V)$ and $T \in \text{Hom}_F(V,W)$. Prove that $T \circ S \in \text{Hom}_F(U,W)$.

6. Let $V$ be an $F$-vector space. Prove that if $\{v_1, \ldots, v_n\}$ is linearly independent in $V$, then so is the set $\{v_1 - v_2, v_2 - v_3, \ldots, v_{n-1} - v_n, v_n\}$.

7. Let $V$ be the subspace of $\mathbb{R}^5$ defined by

$$V = \{ (x_1, x_2, \ldots, x_5) \in \mathbb{R}^5 : x_1 = 4x_4, x_2 = 5x_3 \}.$$ 

Find a basis for $V$.

8. Prove that there does not exist a $T \in \text{Hom}_F(F^5,F^2)$ so that

$$\ker(T) = \{ (x_1, x_2, \ldots, x_5) \in F^5 : x_1 = x_2 \text{ and } x_3 = x_4 = x_5 \}.$$
2.5. PROBLEMS

9. Let $V$ be a finite dimensional vector space and $T \in \text{Hom}_F(V,V)$ with $T^2 = T$.
   (a) Prove that $\text{Im}(T) \cap \ker(T) = \{0\}$.
   (b) Prove that $V = \text{Im}(T) \oplus \ker(T)$.
   (c) Let $V = F^n$. Prove that there is a basis of $V$ such that the matrix of $T$ with respect to this basis is a diagonal matrix whose entries are all 0 or 1.

10. Let $T \in \text{Hom}_F(V,F)$. Prove that if $v \in V$ is not in $\ker(T)$, then
    $$V = \ker(T) \oplus \{cv : c \in F\}.$$ 

11. Let $V_1, V_2$ be subspaces of the finite dimensional vector space $V$. Prove
    $$\dim_F(V_1 + V_2) = \dim_F(V_1) + \dim_F(V_2) - \dim_F(V_1 \cap V_2).$$

12. Suppose that $V$ and $W$ are both 5-dimensional $\mathbb{R}$-subspaces of $\mathbb{R}^9$. Prove that $V \cap W \neq \{0\}$.

13. Let $p$ be a prime and $V$ a dimension $n$ vector space over $\mathbb{F}_p$. Show there are
    $$(p^n - 1)(p^n - p)(p^n - p^2) \cdots (p^n - p^{n-1})$$
    distinct bases of $V$.

14. Let $V$ be an $F$-vector space of dimension $n$. Let $T \in \text{Hom}_F(V,V)$ so that $T^2 = 0$. Prove that the image of $T$ is contained in the kernel of $T$ and hence the dimension of the image of $T$ is at most $n/2$.

15. Let $W$ be a subspace of a finite dimensional vector space $V$. Let $T \in \text{Hom}_F(V,V)$ so that $T(W) \subset W$. Show that $T$ induces a linear transformation $\overline{T} \in \text{Hom}_F(V/W,V/W)$. Prove that $T$ is nonsingular (i.e., injective) on $V$ if and only if $T$ restricted to $W$ and $\overline{T}$ on $V/W$ are both nonsingular.
Chapter 3

Choosing coordinates

This chapter will make the connection between the more abstract version of vector space and linear transformation given in Chapter 2 and the material given in an undergraduate linear algebra class. Throughout this chapter all vector spaces are assumed to be finite dimensional.

3.1 Linear transformations and matrices

Let \( B = \{v_1, \ldots, v_n\} \) be a basis for \( V \). This choice of basis gives an isomorphism between \( V \) and \( F^n \). Namely, for \( v \in V \) if we write \( v = \sum a_i v_i \) for some \( a_i \in F \), we have an isomorphism \( T_B : V \to F^n \) given by \( v \mapsto \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \). When we identify \( V \) with \( F^n \) via this map, we will write \([v]_B = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}\). We refer to this as choosing coordinates on \( V \).

Example 3.1.1. Let \( V = \mathfrak{sl}_2(\mathbb{C}) \). Recall a basis for this vector space is

\[
\mathcal{B} = \left\{ v_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, v_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}.
\]

Let \( \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \) be an element of \( V \). Observe we have \( \begin{pmatrix} a & b \\ c & -a \end{pmatrix} = bv_1 + cv_2 + av_3 \).

Thus,

\[
\left[ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \right]_{\mathcal{B}} = \begin{pmatrix} b \\ c \\ a \end{pmatrix}.
\]
Example 3.1.2. Let \( V = P_2(\mathbb{R}) \). Recall a basis for \( V \) is given by \( B = \{1, x, x^2\} \). Let \( f = a + bx + cx^2 \). Then

\[
[f]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}.
\]

Example 3.1.3. Let \( V = P_2(\mathbb{R}) \). One can easily check that \( C = \{1, (x-1), (x-1)^2\} \) is a basis for \( V \). Let \( f = a + bx + cx^2 \). We can write \( f \) in terms of \( C \) as

\[
f = (a + b + c) + (b + 2c)(x - 1) + c(x - 1)^2.
\]

Thus, we have

\[
[f]_C = \begin{pmatrix} a + b + c \\ b + 2c \\ c \end{pmatrix}.
\]

Let \( T \in \text{Hom}_F(V, W) \). Let \( B = \{v_1, \ldots, v_n\} \) be a basis for \( V \) and \( C = \{w_1, \ldots, w_m\} \) be a basis for \( W \). Recall that we have \( W \cong F^m \) via the map \( Q(w) = [w]_C \) and \( V \cong F^n \) via the map \( P(v) = [v]_B \). Furthermore, recall that any linear transformation from \( F^n \) to \( F^m \) is given by a matrix \( A \in \text{Mat}_{m,n}(F) \). Thus, we have the following diagram:

\[
\begin{array}{ccc}
V & \xrightarrow{T} & W \\
P & \downarrow & Q \\
F^n & \xrightarrow{A} & F^m
\end{array}
\]

Thus, we have a unique matrix \( A \in \text{Mat}_{m,n}(F) \) given by \( A = Q \circ T \circ P^{-1} \). Write \( A = [T]_B^C \), i.e., \( A \) is the matrix that gives the map \( T \) when one chooses \( B \) as the coordinates on \( V \) and \( C \) as the coordinates on \( W \). In particular, \( [T]_B^C \) is the unique matrix that satisfies \( [T]_B^C[v]_B = [T(v)]_C \).

One can easily compute \( [T]_B^C \). Since \( C \) is a basis for \( W \), there are scalars \( a_{ij} \in F \) so that

\[
T(v_j) = \sum_{i=1}^m a_{ij} w_i.
\]

Observe that

\[
[T(v_j)]_C = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.
\]

We also have \( [v_j]_B = e_j \), so \( [T]_B^C[v_j]_B \) is exactly the \( j \)th column of \( [T]_B^C \). Thus, the matrix \( [T]_B^C \) is given by

\[
[T]_B^C = (a_{ij}) = ([T(v_1)]_C \cdots [T(v_n)]_C).
\]

34
Example 3.1.4. Let \( V = P_3(\mathbb{R}) \). Define \( T \in \text{Hom}_\mathbb{R}(V, V) \) by \( T(f(x)) = f'(x) \).

Let \( B = \{1, x, x^2, x^3\} \). For \( f(x) = a + bx + cx^2 + dx^3 \), we have \( T(f(x)) = b + 2cx + 3dx^2 \). In particular, \( T(1) = 0, T(x) = 1, T(x^2) = 2x, \) and \( T(x^3) = 3x^2 \).

The matrix for \( T \) with respect to \( B \) is given by

\[
[T]_B^B = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 3 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

Example 3.1.5. Let \( V = \mathfrak{sl}_2(\mathbb{C}) \) and \( W = \mathbb{C}^4 \). We pick the standard basis \( B = \{v_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, v_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\} \) for \( \mathfrak{sl}_2(\mathbb{C}) \).

Let

\[
C = \left\{ w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, w_2 = \begin{pmatrix} 0 \\ i \\ 0 \\ 0 \end{pmatrix}, w_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix}, w_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right\}.
\]

It is easy to check that \( C \) is a basis for \( W \). Define \( T \in \text{Hom}_F(V, W) \) by

\[
T(v_1) = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix},
T(v_2) = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 1 \end{pmatrix},
T(v_3) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.
\]

From this it is easy to check that

\[
T(v_1) = 2w_1 - w_3,
T(v_2) = -3iw_2 + w_4,
T(v_3) = w_4.
\]

Thus, the matrix for \( T \) with respect to \( B \) and \( C \) is

\[
[T]_B^C = \begin{pmatrix}
2 & 0 & 0 & 0 \\
0 & -3i & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 1 & 1
\end{pmatrix}.
\]
3.1. LINEAR TRANSFORMATIONS AND MATRICES  \[ \text{CHAPTER 3.} \]

**Exercise 3.1.6.** Let \( A \) be a basis of \( U \), \( B \) a basis of \( V \), and \( C \) a basis of \( W \). If \( S \in \text{Hom}_F(U,V) \) and \( T \in \text{Hom}_F(V,W) \), show that

\[
[T \circ S]_A^C = [T]_B^C[S]_A^B.
\]

Let \( T \in \text{Hom}_F(V,V) \) and \( B \) a basis of \( V \). To save notation we will write \([T]_B\) for the matrix \([T]_B^B\).

As one learns in undergraduate linear algebra, it can often be the case that one has information about \( V \) given in terms of a basis \( B \), but it would be more useful if it were given in terms of a basis \( B' \). We now recall how to change from \( B \) to \( B' \). We can recover this by specializing the situation we just studied. Let \( B = \{v_1, \ldots, v_n\} \) and \( B' = \{v'_1, \ldots, v'_n\} \). Define \( T : V \to F^n \) by \( T(v) = [v]_B \) and \( S : V \to F^n \) by \( S(v) = [v]_{B'} \). We have the following diagram:

\[
\begin{array}{ccc}
V & \xrightarrow{id} & W \\
\downarrow T & & \downarrow S \\
F^n & \xrightarrow{A} & F^n
\end{array}
\]

Applying our previous results we see \([v]_{B'} = (S \circ \text{id} \circ T^{-1})([v]_B) = (S \circ T^{-1})([v]_B)\). The change of basis matrix is \([id]_{B'}^B\).

**Exercise 3.1.7.** Let \( B = \{v_1, \ldots, v_n\} \). Prove the change of basis matrix \([id]_{B'}^B\) is given by \(([v_1]_{B'} | \cdots | [v_n]_{B'})\)

**Example 3.1.8.** Let \( V = \mathbb{Q}^2 \). It is elementary to check that \( B = \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \) and \( B' = \left\{ \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix}, \begin{pmatrix} 5 \\ 7 \end{pmatrix} \right\} \) both give bases of \( V \). To compute the change of basis matrix from \( B \) to \( B' \), we expand the elements of \( B \) in terms of the basis \( B' \). For example, we want to find \( a, b \in \mathbb{Q} \) so that

\[
\begin{pmatrix} 1 \\ -1 \end{pmatrix} = a \begin{pmatrix} 2 \\ 3 \end{pmatrix} + b \begin{pmatrix} 5 \\ 7 \end{pmatrix}.
\]

This leads to the system of linear equations

\[
\begin{align*}
1 &= 2a + 5b \\
-1 &= 3a + 7b.
\end{align*}
\]

One solves these by writing expressing them as the matrix \( \begin{pmatrix} 2 & 5 & 1 \\ 3 & 7 & -1 \end{pmatrix} \) and using Gaussian elimination to reduce this matrix to \( \begin{pmatrix} 1 & 0 & -12 \\ 0 & 1 & 5 \end{pmatrix} \). Thus, \( a = \)
3.1. LINEAR TRANSFORMATIONS AND MATRICES

−12 and \( b = 5 \). One now performs the same operation on the vector \( \begin{pmatrix} 1 \\ 1 \end{pmatrix} \) to obtain
\[
\begin{pmatrix} 1 \\ 1 \end{pmatrix} = -2 \begin{pmatrix} 2 \\ 3 \end{pmatrix} + 1 \begin{pmatrix} 5 \\ 7 \end{pmatrix}.
\]
Thus, the change of basis matrix is given by \( \begin{pmatrix} -12 & -2 \\ 5 & 1 \end{pmatrix} \).

**Example 3.1.9.** Consider \( V = P_2(F) \). Let \( B = \{1, x, x^2\} \) and \( B' = \{1, x - 2, (x - 2)^2\} \) be bases of \( V \). We calculate the change of basis matrix. We have
\[
B' \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} = B' \begin{pmatrix} 1 \\ x - 2 \\ (x - 2)^2 \end{pmatrix}.
\]
Thus, the change of basis matrix is given by \( A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \).

**Exercise 3.1.10.** Let \( V = P_3(F) \). Define
\[
x^{(i)} = x(x - 1)(x - 2) \cdots (x - i + 1).
\]
In particular, \( x^{(0)} = 1, x^{(1)} = x, x^{(2)} = x(x - 1) \) and \( x^{(3)} = x(x - 1)(x - 2) \). Set \( B = \{1, x, x^{(2)}, x^{(3)}\} \) and \( B' = \{1, x, x^2, x^3\} \).

1. Show that \( B \) is a basis for \( V \).
2. Find the change of basis matrix from \( B \) to \( B' \).

This gives the language to allow one to translate the results we prove in these notes from the language of vector spaces and linear transformations to the language of \( F^n \) and matrices. Many of the familiar results from undergraduate linear algebra will be proven in the problems at the end of the chapter. We conclude this section with a familiar example from undergraduate linear algebra. One should work the exercises to gain a better understanding of the theory behind the calculations.

**Example 3.1.11.** Consider the matrix
\[
A = \begin{pmatrix} 4 & -4 & 2 \\ -4 & 4 & -2 \\ 2 & -1 & 1 \end{pmatrix}.
\]
We wish to find a basis for the kernel and image of this matrix. To find this, we will compute the reduced row echelon form of the matrix. The first thing to recall is that doing row operations corresponds to changing the basis of the
domain and doing column operations corresponds to changing the basis of the range space. The reduced row echelon form of this matrix is
\[
B = \begin{pmatrix}
1 & 0 & 1/2 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

To find a basis for the kernel, we consider the augmented matrix
\[
(B|0) = \begin{pmatrix}
1 & 0 & 1/2 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

In terms of equations, if the variables are \(x_1, x_2, x_3\), this gives the equations \(x_1 + 1/2x_3 = 0\) and \(x_2 = 0\). Thus, a basis element for the kernel is given by \(\begin{pmatrix}
-1/2 \\
0 \\
1
\end{pmatrix}\).

We know from the exercises that the basis of the image consists of the columns of \(A\) that correspond to the columns of \(B\) containing the pivots, so the first and second columns. Thus, a basis for the image is given by
\[
\begin{align*}
\begin{pmatrix}
4 \\
-4 \\
2
\end{pmatrix},
\begin{pmatrix}
-4 \\
4 \\
-1
\end{pmatrix}
\end{align*}
\]

### 3.2 Transpose of a matrix via the dual map

Recall at the end of last chapter we introduced a dual map. Namely, given a linear map \(T \in \text{Hom}_F(V,W)\), we defined a map \(T^\vee : W^\vee \to V^\vee\) given by \(T(\varphi) = \varphi \circ T\). We also remarked at that point that this map could be used to properly define a transpose. This section deals with this construction.

Given a basis \(B = \{v_1, \ldots, v_n\}\) of \(V\) and a basis \(C = \{w_1, \ldots, w_m\}\) of \(W\) we have dual bases \(B^\vee\) of \(V^\vee\) and \(C^\vee\) of \(W^\vee\). The previous section gave an associated matrix to any linear transformation, so we have a matrix \([T^\vee]_{B^\vee}^{C^\vee} \in \text{Mat}_{n,m}(F)\).

**Definition 3.2.1.** Let \(T \in \text{Hom}_F(V,W)\), \(B\) a basis of \(V\), \(C\) a basis of \(W\), and set \(A = [T]_B^C\). The transpose of \(A\), denoted \(\trans{A}\), is given by \(\trans{A} = [T^\vee]_{B^\vee}^{C^\vee}\).

For this definition to be of interest we need to show it agrees with the definition of a transpose given in undergraduate linear algebra.

**Lemma 3.2.2.** Let \(A = (a_{ij}) \in \text{Mat}_{m,n}(F)\). Then \(\trans{A} = (b_{ij}) \in \text{Mat}_{n,m}(F)\) with \(b_{ij} = a_{ji}\).

**Proof.** Let \(\mathcal{E}_n = \{e_1, \ldots, e_n\}\) be the standard basis of \(F^n\) and \(\mathcal{F}_m = \{f_1, \ldots, f_m\}\) the standard basis for \(F^m\). Let \(\mathcal{E}_n^\vee\) and \(\mathcal{F}_m^\vee\) be the dual bases. Let \(T\) be the linear map associated to \(A\), i.e., \([T]_{\mathcal{E}_n}^{\mathcal{F}_m} = A\). In particular, we have
\[
T(e_i) = \sum_{k=1}^m a_{ki} f_k.
\]
We also have that $[T^\vee]_{E_m^\vee}^{E_n^\vee}$ is a matrix $B = (b_{ij}) \in \text{Mat}_{n,m}(F)$ where the entries of $B$ are given by
\[
T^\vee(f_j^\vee) = \sum_{k=1}^n b_{kj} e_k^\vee.
\]
If we apply $f_j^\vee$ to the first sum we see
\[
f_j^\vee(T(e_i)) = \sum_{k=1}^m a_{ki} f_j^\vee(f_k) = a_{ji}.
\]
If we evaluate the second sum at $e_i$ we have
\[
T^\vee(f_j^\vee)(e_i) = \sum_{k=1}^n b_{kj} e_k^\vee(e_i) = b_{ij}.
\]
We now use the definition of the map $T^\vee$ to see that $f_j^\vee T(e_i) = T^\vee(f_j^\vee)(e_i)$, and so $a_{ji} = b_{ij}$, as desired.

**Exercise 3.2.3.** Let $A_1, A_2 \in \text{Mat}_{m,n}(F)$. Use the definition given above to show that $\tau(A_1 + A_2) = \tau A_1 + \tau A_2$.

**Exercise 3.2.4.** Let $A \in \text{Mat}_{m,n}(F)$ and $c \in F$. Use the definition given above to show that $\tau(cA) = c \tau A$.

We can use our definition of the transpose to give a very simple proof of the following fact.

**Lemma 3.2.5.** Let $A \in \text{Mat}_{m,n}(F)$ and $B \in \text{Mat}_{p,m}(F)$. Then $\tau(BA) = \tau B \tau A$.

**Proof.** Write $E_n$ be the standard basis on $F^n$, and likewise for $E_n$ and $E_p$. Let $S$ be the multiplication by $A$ map and $T$ the multiplication by $B$ map so that $A = [S]_{E_n}^{E_m}$ and $B = [T]_{E_p}^{E_m}$. We also have $BA = [T \circ S]_{E_n}^{E_p}$. We now have
\[
\tau(BA) = [(T \circ S)^\vee]_{E_p}^{E_n} = [S^\vee \circ T^\vee]_{E_p}^{E_n} = [S^\vee]_{E_m}^{E_p} [T^\vee]_{E_p}^{E_n} = \tau B \tau A,
\]
as claimed.

We also get the transpose of the inverse of a matrix very easily.

**Lemma 3.2.6.** Let $A \in \text{GL}_n(F)$. Then $\tau(A^{-1}) = (\tau A)^{-1}$.
3.2. TRANPOSE OF A MATRIX VIA THE DUAL MAP  CHAPTER 3.

Proof. The strategy of proof is to show that \( t(A^{-1}) \) satisfies the conditions of being an inverse of \( tA \), i.e., \( tA t(A^{-1}) = 1_n = t(A^{-1}) tA \). We then use that inverses in a group are unique, so it must be that \( t(A^{-1}) = (tA)^{-1} \).

Let \( E_n \) be the standard basis of \( F^n \) and let \( T \) be the multiplication by \( A \) map, i.e., \( A = [T]_{E_n}^{E_n} \). By assumption we have \( A \) is invertible, so \( T^{-1} \) exists and \( A^{-1} = [T^{-1}]_{E_n}^{E_n} \). Write \( \text{id} \) for the identity map and we continue to denote the identity matrix by \( 1_n \). Observe we have

\[
1_n = [\text{id} \land]_{E_n}^{E_n} = [(T^{-1} \circ T) \land]_{E_n}^{E_n} = [T \land (T^{-1}) \land]_{E_n}^{E_n} = [T \land]_{E_n}^{E_n} ((T^{-1}) \land]_{E_n}^{E_n} = tA t(A^{-1}).
\]

Similarly one has \( 1_n = t(A^{-1}) tA \). As noted above, the uniqueness of inverses in \( \text{GL}_2(F) \) completes the proof. \( \square \)
3.3 Problems

For all of these problems $V$ is a finite dimensional $F$-vector space.

1. Let $V = P_n(F)$. Let $B = \{1, x, \ldots, x^n\}$ be a basis of $P_n(F)$. Let $\lambda \in F$ and set $C = \{1, x - \lambda, \ldots, (x - \lambda)^{n-1}, (x - \lambda)^n\}$. Define a linear transformation $T \in \text{Hom}_F(V, V)$ by defining $T(x^j) = (x - \lambda)^j$. Determine the matrix of this linear transformation. Use this to conclude that $C$ is also a basis of $P_n(F)$.

2. Let $V = P_5(\mathbb{Q})$ and let $B = \{1, x, \ldots, x^5\}$. Prove that the following are elements of $V^*$ and express them as linear combinations of the dual basis:
   (a) $\phi : V \to \mathbb{Q}$ defined by $\phi(p(x)) = \int_0^1 t^2 p(t) dt$.
   (b) $\phi : V \to \mathbb{Q}$ defined by $\phi(p(x)) = p'(5)$ where $p'(x)$ denotes the derivative of $p(x)$.

3. Let $V$ be a vector space over $F$ and let $T \in \text{Hom}_F(V, V)$. A nonzero element $v \in V$ satisfying $T(v) = \lambda v$ for some $\lambda \in F$ is called an eigenvector of $T$ with eigenvalue $\lambda$.
   (a) Prove that for any fixed $\lambda \in F$ the collection of eigenvectors of $T$ with eigenvalue $\lambda$ together with 0 forms a subspace of $V$.
   (b) Prove that if $V$ has a basis $B$ consisting of eigenvectors for $T$ then $[T]_B^B$ is a diagonal matrix with the eigenvalues of $T$ as diagonal entries.

4. Let $A, B \in \text{Mat}_n(F)$. We say $A$ and $B$ are similar if there exists $T \in \text{Hom}_F(V, V)$ for some $n$-dimensional $F$-vector space $V$ so that $A = [T]_B$ and $B = [T]_C$ for bases $B$ and $C$ of $V$.
   (a) Show that if $A$ and $B$ are similar, there exists $P \in \text{GL}_n(F)$ so that $A = PBP^{-1}$. Conversely, if there exists $P \in \text{GL}_n(F)$ so that $A = PBP^{-1}$, show that $A$ is similar to $B$.
   (b) Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined so that $T = T_A$ where $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 3 \\ 1 & -2 & 4 \end{pmatrix}$.

   Let $B = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}$ be a basis of $\mathbb{R}^3$. First, calculate $[T]_B$ directly. Then find $P$ so that $[T]_B = PAP^{-1}$.
5. Let $T \in \text{Hom}_\mathbb{R}(\mathbb{R}^4, \mathbb{R}^4)$ be the linear transformation given by the matrix

$$A = \begin{pmatrix} 1 & -1 & 0 & 3 \\ -1 & 2 & 1 & -1 \\ -1 & 1 & 0 & -3 \\ 1 & -2 & -1 & 1 \end{pmatrix}$$

with respect to the standard basis $\mathcal{E}_4$. Determine a basis for the image and kernel of $T$.

6. Let $T \in \text{Hom}_F(P_7(F), P_7(F))$ be defined by $T(f) = f'$ where $f'$ denotes the usual derivative of a polynomial $f \in P_7(F)$. For each of the fields below, determine a basis for the image and kernel of $T$:

(a) $F = \mathbb{R}$
(b) $F = \mathbb{F}_3$.

7. Let $V$ and $W$ be $F$ vector spaces of dimensions $n$ and $m$ respectively. Let $A \in \text{Mat}_{m,n}(F)$ be a matrix representing a linear transformation $T$ from $V$ to $W$ with respect to bases $\mathcal{B}_1$ for $V$ and $\mathcal{C}_1$ for $W$. Suppose $B$ is the matrix for $T$ with respect to the bases $\mathcal{B}_2$ for $V$ and $\mathcal{C}_2$ for $W$. Let $\text{id}_V$ denote the identity map on $V$ and $\text{id}_W$ denote the identity map on $W$. Set $P = [\text{id}_V]_{\mathcal{B}_1}^{\mathcal{B}_2}$ and $Q = [\text{id}_W]_{\mathcal{C}_1}^{\mathcal{C}_2}$. Prove that $Q^{-1}AP = B$.

The next problems recall Gaussian elimination. First we recall the set-up from undergraduate linear algebra.

Consider a system of equations

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

for unknowns $x_1, \ldots, x_n$ and scalars $a_{ij}, c_i$. We have a coefficient matrix $A = (a_{ij}) \in \text{Mat}_{m,n}(F)$, and an augmented matrix $(A|C) \in \text{Mat}_{m,(n+1)}(F)$ where we add the column vector given by the $c_i$’s on the right side. Note the solutions to the equations above are not altered if we perform the following operations:

(i) interchange any two equations
(ii) add a multiple of one equation to another
(iii) multiply any equation by a nonzero element of $F$.

In terms a the matrix these correspond to the elementary row operations given by

(r1) interchange any two rows
(r2) add a multiple of one row to another
(r3) multiply any row by a nonzero element of $F$.

A matrix $A$ that can be transformed into a matrix $B$ by a series of elementary row operations is said to be row reduced to $B$.

8. Describe the elementary row operations in terms of matrices. In particular, explain what it is doing on a basis. You can do this separately for each elementary operation.

We say $A \sim B$ if $A$ can be row reduced to $B$.

9. Prove that $\sim$ is an equivalence relation. Prove that if $A \sim B$ then the row rank of $A$ is the same as the row rank of $B$.

An $m$ by $n$ matrix is said to be in reduced row echelon form if

(a) the first nonzero entry $a_{i,j_i}$ in row $i$ is 1 and all other entries in the corresponding $j_i$th column are 0 and

(b) $j_1 < j_2 < \cdots < j_r$ where $r$ is the number of nonzero rows.

An augmented matrix $(A|C)$ is said to be in reduced row echelon form if the coefficient matrix $A$ is in reduced row echelon form. For example, the following matrix is in reduced row echelon form:

$$A = \begin{pmatrix}
1 & 0 & 5 & 7 & 0 & 3 \\
0 & 1 & -1 & 1 & 0 & -4 \\
0 & 0 & 1 & 6 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

The first nonzero entry in any given row of a reduced row echelon matrix is referred to as a pivotal element. The columns containing pivotal elements are referred to as pivotal columns.

10. Prove by induction that any augmented matrix can be put in reduced row echelon form by a series of elementary row operations.
11. Let $A$ and $B$ be two matrices in reduced row echelon form. Prove that if $A$ and $B$ are row equivalent, then $A = B$.

12. Prove that the row rank of a matrix in reduced row echelon form is the number of nonzero rows.

13. Find the reduced row echelon form of the matrix

$$A = \begin{pmatrix}
1 & 1 & 4 & 8 & 0 & -1 \\
1 & 2 & 3 & 9 & 0 & -5 \\
0 & -2 & 2 & -2 & 1 & 14 \\
1 & 4 & 1 & 11 & 0 & -13
\end{pmatrix}.$$

14. Use what you have done above to find solutions of the system of equations

$$\begin{align*}
x - 2y + z &= 5 \\
x - 4y + 6z &= 10 \\
4x - 11y + 11z &= 12.
\end{align*}$$

14. Let $V$ be an $n$-dimensional $F$ vector space with basis $\mathcal{E} = \{e_1, \ldots, e_n\}$ and let $W$ be an $m$-dimensional $F$ vector space with basis $\mathcal{F} = \{f_1, \ldots, f_m\}$. Let $T \in \text{Hom}_F(V, W)$ with $[T]_{\mathcal{E}}^{\mathcal{F}} = (a_{ij})$. Let $A'$ be the reduced row echelon form of $A$.

(a) Prove that the image $T(V)$ has dimension $r$ where $r$ is the number of nonzero rows of $A'$ and that a basis for $T(V)$ is given by the vectors $T(e_i)$, i.e., the columns of $A$ corresponding to the pivotal columns of $A'$ give the coordinates of a basis for the image of $T$. The elements in the kernel of $T$ are the vectors in $V$ whose coordinates $(x_1, \ldots, x_n)$ with respect to $\mathcal{E}$ satisfy the equation

$$A' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = 0$$

and the solutions $x_1, \ldots, x_n$ to this system of linear equations are determined by the matrix $A'$.

(b) Prove that $T$ is injective if and only if $A'$ has $n$ nonzero rows.

(c) By (a) the kernel of $T$ is nontrivial if and only if $A'$ has nonpivotal columns. Show that each of the variables $x_1, \ldots, x_n$ above corresponding to the nonpivotal columns of $A'$ can be prescribed arbitrarily and the values of the remaining variables are then uniquely
determined to give an element $x_1 e_1 + \cdots + x_n e_n$ in the kernel of $T$.
In particular, show that the coordinates of a basis for the kernel are
obtained by successively setting one nonpivotal variable equal to 1
and all other nonpivotal variables to 0 and solving for the remaining
pivotal variables. Conclude that the kernel of $T$ has dimension $n - r$
where $r$ is the rank of $A$.

(d) Give a basis for the image and kernel of $T$ if the matrix associated
to $T$ with respect to the standard bases is

$$
\begin{pmatrix}
1 & 2 & 3 & 4 & 4 \\
-2 & -4 & 0 & 0 & 2 \\
1 & 2 & 0 & 1 & -2 \\
1 & 2 & 0 & 0 & -1
\end{pmatrix}
$$
Chapter 4

Structure Theorems for Linear Transformations

Let $T \in \text{Hom}_F(V, V)$. Let $B$ be a basis for $V$, where $\dim_F V < \infty$, then we have a matrix $[T]_B$. Our goal is to pick this basis so that $[T]_B$ is as nice as possible.

Throughout this chapter we will always take $V$ to be a finite dimensional vector space.

Though we do not formally introduce determinants until Chapter 7, we use determinants throughout this chapter. Here one should just treat the determinant as it was used in undergraduate linear algebra, namely, in terms of the cofactor expansion.

4.1 Invariant subspaces

In this section we will define some of the basic objects needed throughout the remainder of the chapter. However, we begin with an example.

Example 4.1.1. Let $V$ be a 2-dimensional vector space with basis $\{v_1, v_2\}$. Let $T \in \text{Hom}_F(V, V)$ such that $T(v_1) = v_1 + 3v_2$ and $T(v_2) = 2v_1 + 4v_2$. Thus,

$$[T]_B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

It is natural to ask if there is a basis $C$ so that $[T]_C$ is a diagonal matrix. We will see that if $F = \mathbb{Q}$, there is no such basis $C$. However if $\mathbb{Q}(\sqrt{33}) \subseteq F$, then there is a basis $C$ so that

$$[T]_C = \begin{pmatrix} \frac{5+\sqrt{33}}{2} & 0 \\ 0 & \frac{5-\sqrt{33}}{2} \end{pmatrix}.$$

We can also consider the question over finite fields. For instance, $F = \mathbb{F}_3$, there is a basis $C$ so that

$$[T]_C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
but if $F = \mathbb{F}_5$ the matrix cannot be diagonalized. The results of this chapter will make answering such a question routine.

Let $V$ be an $F$-vector space with $\dim_F V = n$ and let $T \in \text{Hom}_F(V,V)$. Let $f(x) \in F[x]$. We will use throughout this chapter the notation $f(T)$. If $f(x) = a_m x^m + \cdots + a_1 x + a_0$, we will view $f(T)$ as the linear map $a_m T^m + \cdots + a_1 T + a_0$ where $T^m = T \circ T \circ \cdots \circ T$ with $m$ copies of $T$.

**Theorem 4.1.2.** Let $v \in V$ with $v \neq 0$. There is a unique nonzero monic polynomial $m_{T,v}(x) \in F[x]$ of lowest degree so that $m_{T,v}(T)(v) = 0$. Moreover, $\deg m_{T,v}(x) \leq \dim_F V$.

**Proof.** Consider the set $\{v, T(v), \ldots, T^n(v)\}$. There are $n + 1$ vectors in this set so they must be linearly dependent, i.e., there exists $a_0, \ldots, a_n \in F$ such that $a_n T^n(v) + \cdots + a_1 T(v) + a_0 v = 0$. Set $p(x) = a_n x^n + \cdots + a_1 x + a_0$. We have that $p(T)(v) = 0$.

Consider the subset $I_v \subseteq F[x]$ given by $I_v = \{f(x) : f(T)(v) = 0\}$. Since $p(x) \in I_v$ we have $I_v \neq \emptyset$. Pick $f(x) \in I_v$ nonzero of minimal degree and write $f(x) = a_m x^m + \cdots + a_1 x + a_0$ with $a_m \neq 0$, $m \leq n$. Set $\tilde{f}(x) = \frac{1}{a_m} f(x) \in I_v$. This is a monic polynomial of minimum degree in $I_v$.

It remains to show that $f(x)$ is unique. Suppose $g(x) \in I_v$ with $\deg g = m$ and $g$ monic. Write $f(x) = q(x)g(x) + r(x)$ for $q(x), r(x) \in F[x]$ with $r(x) = 0$ or $\deg r(x) < m$. Rewriting this gives $r(x) = f(x) - q(x)g(x)$. Observe $r(T)(v) = f(T)(v) - q(T)(v)g(T)(v) = 0$ and so $r(x) \in I_v$. However, this contradicts $f$ having minimal degree in $I_v$ unless $r(x) = 0$. Thus, $f(x) = q(x)g(x)$. Now observe that since $\deg f = m = \deg g$, we must have $\deg q = 0$. This gives $q(x) \in F$. Now we use that $f$ and $g$ are monic to see $q = 1$ and so $f = g$. This gives the result. \qed

One should note the previous proof amounted to showing that $F[x]$ is a principal ideal domain. Since we are not assuming that level of abstract algebra, the proof is necessary. We will use this type of argument repeatedly so it is important to understand it at this point.

**Definition 4.1.3.** The polynomial $m_{T,v}(x)$ is referred to as the $T$-annihilator of $v$.

**Corollary 4.1.4.** Let $v \in V$ and $T \in \text{Hom}_F(V,V)$. If $f(x) \in F[x]$ satisfies $f(T)(v) = 0$, then $m_{T,v}(x) \mid f(x)$.

**Proof.** We showed this in the course of proving the previous theorem, but we repeat the argument here as this fact will be extremely important in what follows. Let $f(x) \in F[x]$ so that $f(T)(v) = 0$. Using the division algorithm we can write

$$f(x) = q(x)m_{T,v}(x) + r(x)$$
with \( q, r \in F[x] \) and \( r(x) = 0 \) or \( \deg r < \deg m_{T,v} \). We have

\[
0 = f(T)(v) = q(T)m_{T,v}(T)(v) + r(T)(v) = r(T)(v).
\]

This contradicts the minimality of the degree of \( m_{T,v}(x) \) unless \( r(x) = 0 \). Thus, we have the result. \( \square \)

**Example 4.1.5.** Let \( V = \mathbb{R}^n \) and \( e_n = \{ e_1, \ldots, e_n \} \) be the standard basis of \( V \). Define \( T : V \to V \) by \( T(e_j) = e_{j-1} \) for \( 2 \leq j \leq m \) and \( T(e_1) = 0 \). We calculate \( m_{T,e_j}(x) \) for each \( j \).

Observe that if we set \( f_1(x) = x \) then \( f_1(T)(e_1) = T(e_1) = 0 \). Moreover, if \( g(x) = a \) a constant, then \( g(T)(e_1) \neq 0 \). Thus, using the above corollary we have that \( m_{T,e_1}(x) \) is a monic polynomial of degree at least one that divides \( f_1(x) = x \), i.e., \( m_{T,e_1}(x) = x \).

Next we calculate \( m_{T,e_2}(x) \). Observe that if we set \( f_2(x) = x^2 \), then \( f_2(T)(e_2) = T^2(e_2) = T(e_1) = 0 \). Thus, \( m_{T,e_2} \mid f_2 \). This gives that \( m_{T,e_2}(x) = 1, x, \) or \( x^2 \) since it must be monic. Since \( T(e_2) = e_1 \), it must be that \( m_{T,e_2}(x) = x^2 \). Similarly, one obtains that \( m_{T,e_j}(x) = x^j \) for \( 2 \leq j \leq m \).

**Example 4.1.6.** We now return to Example 4.1.1 and adopt the notation used there. We find \( m_{T,v_1}(x) \) and leave the calculation of \( m_{T,v_2}(x) \) as an exercise. We have \( T(v_1) = v_1 + 3v_2 \) and \( T^2(v_1) = T(v_1) + 3T(v_2) = v_1 + 3v_2 + 3(2v_1 + 4v_2) = 7v_1 + 15v_2 \). Since \( V \) has dimension 2 we know there exists \( b, c \in F \) such that \( T^2(v_1) + bv_1 + cv_1 = 0 \). Finding \( b \) and \( c \) amounts to solving a system of linear equations: we obtain \( b = -5, c = -2 \). So \( f_1(x) = x^2 - 5x - 2 \) satisfies \( f(T)(v_1) = 0 \). Whether this is \( m_{T,v_1}(x) \) depends on the field \( F \) and whether we can factor over \( F \). For example, over \( \mathbb{Q} \) this is an irreducible polynomial and so it must be that \( m_{T,v_1}(x) = x^2 - 5x - 2 \) over \( \mathbb{Q} \). In fact, we will later see this means that \( m_{T,v_1}(x) = x^2 - 5x - 2 \) over any field \( F \) that contains \( \mathbb{Q} \). One other thing to observe is that the roots of this polynomial are \( \frac{5}{2} \pm \frac{\sqrt{29}}{2} \).

**Exercise 4.1.7.** Redo the previous example with \( F = \mathbb{F}_3 \) this time.

Though the annihilating polynomials are useful, they only tell us about the linear map element by element. What we would really like is it a polynomial that tells us about the overall linear map. The minimal polynomial provides just such a polynomial.

**Theorem 4.1.8.** Let \( \dim_F V = n \). There is a unique monic polynomial \( m_T(x) \in F[x] \) of lowest degree so that \( m_T(T)(v) = 0 \) for all \( v \in V \). Furthermore, \( \deg m_T(x) \leq n^2 \).

**Proof.** Let \( B = \{ v_1, \ldots, v_n \} \) be a basis for \( V \). Let \( m_{T,v_i}(x) \) be the annihilating polynomial for each \( i \). Set \( m_T(x) = \text{lcm}_i m_{T,v_i}(x) \). Note that \( m_T(x) \) is monic and \( m_T(T)(v_i) = 0 \) for each \( 1 \leq i \leq n \). From this it is easy to show that...
Corollary 4.1.9. Let \( m_T(v) = 0 \) for all \( v \in V \). Since each \( m_{T,v_i}(x) \) has degree \( n \) and there are \( n \) polynomials, we must have \( \deg m_T(x) \leq \sum_{i=1}^{n} n = n^2 \).

It remains to show \( m_T(x) \) is unique. Suppose there exists \( r(x) \in F[x] \) with \( r(T)(v) = 0 \) for all \( v \in V \) but \( m_{T,v_j}(x) \nmid r(x) \) for some \( j \). This is a contradiction because if \( m_{T,v_j}(x) \nmid r(x) \), then \( r(T)(v_j) \neq 0 \). Thus, by the definition of least common multiple we must have \( m_T(x) \mid r(x) \) and so \( m_T(x) \) is the unique monic polynomial of minimal degree satisfying \( m_T(T)(v) = 0 \) for all \( v \in V \).

We note the following corollary that was shown in the process of proving the last result.

**Corollary 4.1.9.** Let \( T \in \text{Hom}_F(V,V) \) and suppose \( f(x) \in F[x] \) with \( f(T)(v) = 0 \) for all \( v \in V \). Then \( m_T(x) \mid f(x) \).

The bound on the degree of \( m_T(x) \) given above is far from optimal. In fact, we will see shortly that \( \deg m_T(x) \leq n \).

**Definition 4.1.10.** The polynomial \( m_T(x) \) is called the minimal polynomial of \( T \).

**Example 4.1.11.** We once again return to Example 4.1.1. Let \( F = \mathbb{Q} \). We saw \( m_{T,v_1}(x) = x^2 - 5x - 2 \) and one saw in the exercise that \( m_{T,v_2}(x) = x^2 - 5x - 2 \). Thus,

\[
m_T(x) = \text{lcm}(x^2 - 5x - 2, x^2 - 5x - 2) = x^2 - 5x - 2.
\]

**Example 4.1.12.** Let \( V = \mathbb{Q}^3 \) and let \( \mathcal{B} = \{e_1, e_2, e_3\} \) be the standard basis. Let \( T \) be given by the matrix \( \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & -1 \end{pmatrix} \). One can calculate from this that \( m_{T,e_1}(x) = x - 1, m_{T,e_2} = (x - 1)^2, m_{T,e_3} = (x - 1)^2(x + 1) \). Thus,

\[
m_T(x) = \text{lcm}((x - 1), (x - 1)^2, (x - 1)^2(x + 1)) = (x - 1)^2(x + 1).
\]

The following result is the first step in showing that one can realize the minimal polynomial of any linear map as the annihilator of an element of the vector space.

**Lemma 4.1.13.** Let \( T \in \text{Hom}_F(V,V) \). Let \( v_1, \ldots, v_k \in V \) and set \( p_i(x) = m_{T,v_i}(x) \). Suppose the \( p_i(x) \) are pairwise relatively prime. Set \( v = v_1 + \cdots + v_k \). Then \( m_{T,v}(x) = p_1(x) \cdots p_k(x) \).

**Proof.** We prove the case \( k = 2 \). The general case follows by induction and is left as an exercise. Let \( p_1(x) \) and \( p_2(x) \) be as in the statement of the lemma.
The fact that they are relatively prime gives polynomials \(q_1(x), q_2(x) \in F[x]\) such that \(p_1(x)q_1(x) + p_2(x)q_2(x) = 1\). Thus, we have

\[
v = 1 \cdot v = (p_1(T)q_1(T) + p_2(T)q_2(T))v = q_1(T)p_1(T)(v) + q_2(T)p_2(T)(v)
\]

\[
= q_1(T)p_1(T)(v_1) + q_1(T)p_1(T)(v_2) + q_2(T)p_2(T)(v_1) + q_2(T)p_2(T)(v_2)
\]

\[
= q_1(T)p_1(T)(v_2) + q_2(T)p_2(T)(v_1)
\]

where we have used \(p_1(T)(v_1) = 0\) and \(p_2(T)(v_2) = 0\). Set \(w_1 = q_2(T)p_2(T)(v_1)\) and \(w_2 = q_1(T)p_1(T)(v_2)\) so that \(v = w_1 + w_2\).

Observe that

\[
p_1(T)(w_1) = p_1(T)q_2(T)p_2(T)(v_1)
\]

\[
= p_2(T)q_2(T)p_1(T)(v_1)
\]

\[
= 0.
\]

Thus, \(w_1 \in \ker p_1(T)\). Similarly, \(w_2 \in \ker p_2(T)\).

Let \(r(x) \in F[x]\) such that \(r(T)(v) = 0\). Recall that \(v = w_1 + w_2\). Since \(w_2 \in \ker p_2(T)\) we have

\[
p_2(T)(v) = p_2(T)(w_1 + w_2)
\]

\[
= p_2(T)(w_1).
\]

Thus,

\[
0 = p_2(T)q_2(T)r(T)(v)
\]

\[
= r(T)q_2(T)p_2(T)(v)
\]

\[
= r(T)q_2(T)p_2(T)(w_1).
\]

Moreover, we have

\[
0 = r(T)p_1(T)q_1(T)(w_1)
\]

since \(w_1 \in \ker p_1(T)\). Thus,

\[
0 = r(T)(p_1(T)q_1(T) + p_2(T)q_2(T))(w_1).
\]

Using that \(p_1(T)q_1(T) + p_2(T)q_2(T) = 1\), we obtain \(0 = r(T)(w_1)\). Thus, we have \(0 = r(T)(w_1) = r(T)(p_2(T)q_2(T)(v_1))\). Since \(p_1(x)\) is the annihilating polynomial of \(v_1\), we obtain \(p_1(x)r(x)p_2(x)q_2(x)\). However, since \(p_1(x)q_1(x) + p_2(x)q_2(x) = 1\) we have that \(p_1\) is relatively prime to \(p_2(x)q_2(x)\), so \(p_1(x)\) is relatively prime to \(r(x)\). A similar argument shows \(p_2(x)\) is relatively prime to \(p_1(x)p_2(x)\). Observe that

\[
p_1(T)p_2(T)(v) = p_1(T)p_2(T)(v_1) + p_1(T)p_2(T)(v_2)
\]

\[
= 0.
\]

Thus \(p_1(x)p_2(x)\) is a monic polynomial and for any \(r(x) \in F[x]\) so that \(r(T)(v) = 0\), we have \(p_1(x)p_2(x) | r(x)\). This is exactly what it means for \(m_{T,v}(x) = p_1(x)p_2(x)\). \(\square\)
Theorem 4.1.14. Let $T \in \text{Hom}_F(V, V)$. There exists $v \in V$ such that $m_{T,v}(x) = m_T(x)$.

Proof. Let $v_1, \ldots, v_n$ be a basis of $V$ so that $m_T(x) = \text{lcm}_i m_{T,v_i}(x)$. Factor $m_T(x)$ into irreducible factors $p_1(x)^{e_1} \cdots p_k(x)^{e_k}$ with $e_i \geq 1$ and the $p_i(x)$ pairwise relatively prime. For each $1 \leq j \leq n$ there exists $i_j \in \{1, \ldots, k\}$ and $q_{i_j}(x) \in F[x]$ such that $m_{T,v_j}(x) = p_{i_j}(x)^{e_{i_j}} q_{i_j}(x)$. Set $u_{i_j} = q_{i_j}(T)(v_j)$. Then $m_{T,u_{i_j}}(x) = p_{i_j}(x)^{e_{i_j}}$. Thus, if we set $v = \sum u_{i_j}$, then Lemma 4.1.13 gives

$$m_{T,v}(x) = p_1(x)^{e_1} \cdots p_k(x)^{e_k} = m_T(x).$$

This result gives us the desired bound on the degree of $m_T(x)$.

Corollary 4.1.15. For $T \in \text{Hom}_F(V, V)$ we have $\deg m_T(x) \leq n$.

Proof. Since there exists $v \in V$ so that $m_T(x) = m_{T,v}(x)$ and we know $\deg m_{T,v}(x) \leq n$, this gives the result.

Definition 4.1.16. Let $A \in \text{Mat}_n(F)$. We define the characteristic polynomial of $A$ as $c_A(x) = \det(x1_n - A) \in F[x]$.

One must be very careful with what is meant here. As we have seen above, we will often be interested in evaluating a polynomial at a linear map or a matrix. Without the more rigorous treatment, we have to be careful what we mean by this. For example, the Cayley-Hamilton Theorem (see Theorem 4.1.30) says that $c_A(A) = 0$. At first glance this seems trivial, but that is only the case if you misinterpret what is meant by $c_A(A)$. What this actually means is form the polynomial $c_A(x)$ and then replace the $x$’s with $A$’s. One easy way to see the difference is that $c_A(B)$ is a matrix for any matrix $B$, but if you evaluated $B1_n - A$ and then took the determinant this would give a scalar. Note that a more rigorous treatment of the characteristic polynomial in the appendix. For a first reading of the material the more rigorous treatment can certainly be skipped.

Recall that we say matrices $A, B \in \text{Mat}_n(F)$ are similar if there exists $P \in \text{GL}_n(F)$ so that $A = PBP^{-1}$.

Lemma 4.1.17. Let $A, B \in \text{Mat}_n(F)$ be similar matrices. Then $c_A(x) = c_B(x)$. 

51
4.1. INVARIANT SUBSPACES  

CHAPTER 4.

**Proof.** Let \( P \in \text{GL}_n(F) \) such that \( A = PBP^{-1} \). Then we have

\[
c_A(x) = \det(xI_n - A) = \det(xI_n - PBP^{-1}) = \det(PxI_nP^{-1} - PBP^{-1}) = \det(P(xI_n - B)P^{-1}) = \det P \det(xI_n - B) \det P^{-1} = \det(xI_n - B) = c_B(x).
\]

We can use this result to define the characteristic polynomial of a linear map as well. Given \( T \in \text{Hom}_F(V,V) \), we define \( c_T(x) = c_{[T]_B}(x) \) for any basis \( B \). Note this is well-defined because choosing a different basis gives a similar matrix, which does not affect the characteristic polynomial. If \( \dim_F V = n \), then \( \deg c_T(x) = n \) and \( c_T(x) \) is monic.

**Definition 4.1.18.** Let \( f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x] \). The **companion matrix** of \( f \) is given by:

\[
C(f(x)) = \begin{bmatrix}
-a_{n-1} & 1 & 0 & 0 & \cdots & 0 \\
-a_{n-2} & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
-a_1 & 0 & 0 & 0 & \cdots & 1 \\
-a_0 & 0 & 0 & 0 & \cdots & 0
\end{bmatrix}
\]

The companion matrix will be extremely important when we study rational canonical form.

**Lemma 4.1.19.** Let \( f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x] \). Set \( A = C(f(x)) \). Then \( c_A(x) = f(x) \).

**Proof.** Observe we have

\[
xI_n - A = \begin{bmatrix}
x + a_{n-1} & -1 & 0 & \cdots & 0 \\
a_{n-2} & x & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
a_1 & 0 & 0 & x & -1 \\
a_0 & 0 & 0 & 0 & x
\end{bmatrix}
\]

We prove the result by induction on \( n \). First, suppose \( n = 1 \). Then we have \( f(x) = x + a_0 \) and \( A = (a_0) \). So \( xI_1 - A = x + a_0 \) and thus \( c_A(x) = \det(x + a_0) = x + a_0 = f(x) \) as claimed. Now suppose the result is true for all \( n \leq k - 1 \). We
show the result is true for $n = k$. We have
\[
c_A(x) = \det(x1_k - A)
= \det \begin{pmatrix}
x + a_{k-1} & -1 & 0 & \cdots & 0 \\
a_{k-2} & x & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
a_1 & 0 & 0 & x & -1 \\
a_0 & 0 & 0 & 0 & x
\end{pmatrix}
= (-1)^{k+1}a_0 \det \begin{pmatrix}
-1 & 0 & 0 & \cdots & 0 \\
x & -1 & 0 & \cdots & 0 \\
0 & x & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & x & -1
\end{pmatrix} + (-1)^{2k}x \det \begin{pmatrix}
x + a_{k-1} & -1 & 0 & \cdots & 0 \\
a_{k-2} & x & -1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
a_1 & 0 & 0 & x
\end{pmatrix}
= a_0 + x(a_{k-1} + a_{k-2} + \cdots + a_1)
= f(x).
\]
Thus, we have the result by induction. \hfill \Box

The next theorem gives us our first serious result towards our structure theorems.

**Theorem 4.1.20.** Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \in F[x]$. Let $A = C(f(x))$. Set $V = F^n$ and $T = T_A$. Let $\mathcal{E}_n = \{e_1, \ldots, e_n\}$ be the standard basis of $F^n$. Then the subspace $W \subseteq V$ given by
\[
W = \{g(T)(e_n) : g(x) \in F[x]\}
\]
is all of $V$. Moreover, $m_T(x) = m_{T,e_n}(x) = f(x)$.

**Proof.** Observe we have $T(e_n) = e_{n-1}, T^2(e_n) = e_{n-2}$, and in general $T^k(e_n) = e_{n-k}$ for $k \leq n-1$. We have that $W$ contains $\text{span}_F(T^{n-1}(e_n), \ldots, T(e_n), e_n) = \text{span}_F(e_1, \ldots, e_n) = V$. Since $W \subseteq V$, this implies that $W = V$.

Note, we have $n$ elements in $\{T^{n-1}(e_n), \ldots, T(e_n), e_n\}$ and they span, so they are linearly independent. So any polynomial $p(x)$ such that $p(T)(e_n) = 0$ must have degree at least $n$. We have
\[
T^n(e_n) = T(e_1)
= -a_{n-1}e_1 - a_{n-2}e_2 - \cdots - a_0e_n
= -a_{n-1}T^{n-1}(e_n) - \cdots - a_1T(e_n) - a_0e_n.
\]
Thus, $T^n(e_n) + \cdots + a_1T(e_n) + a_0e_n = 0$. This gives $f(T)(e_n) = 0$. Since $f$ is degree $n$ and is monic, we must have that $f(x) = m_{T,e_n}(x)$. We know $m_{T,e_n}(x)|m_T(x)$ and $\deg m_T(x) \leq n$, so $m_T(x) = m_{T,e_n}(x) = f(x)$. \hfill \Box

We now switch our attention to subspaces of $V$ that are well-behaved with respect to $T$, namely, ones that are preserved by $T$.
4.1. INVARIANT SUBSPACES

Definition 4.1.21. A subspace $W \subseteq V$ that satisfies $T(W) \subseteq W$ is referred to as a $T$-invariant subspace or a $T$-stable subspace.

As is the custom, we rephrase the definition in terms of matrices.

Theorem 4.1.22. Let $V$ be an $n$-dimensional $F$-vector space and $W \subseteq V$ a $k$-dimensional subspace. Let $B_W = \{v_1, \ldots, v_k\}$ be a basis of $W$ and extend to a basis $B = \{v_1, \ldots, v_n\}$ of $V$. Let $T \in \text{Hom}_F(V, V)$. Then $W$ is $T$-invariant if and only if $[T]_B$ is block upper triangular

$$[T]_B = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

where $A \in \text{Mat}_k(F)$ and is given by $A = [T|_W]_{B_W}$.

Note that if $W$ is a $T$-invariant subspace, then $T|_W \in \text{Hom}_F(W, W)$. Moreover, we have a natural map $\bar{T} : V/W \to V/W$ given by $T(v + W) = T(v) + W$. This is well-defined precisely because $W$ is $T$-invariant, which one should check as an exercise. We also have the following simple relations between the annihilating polynomials of $\bar{T}$ and $T$ as well as the minimal polynomials. These will be useful in induction proofs that follow.

Lemma 4.1.23. Let $W \subseteq V$ be $T$-invariant and let $\bar{T}$ be the induced linear map on $V/W$. Let $v \in V$. Then $m_{\bar{T},[v]}(x)|m_{T,v}(x)$ where $[v] = v + W$.

Proof. Observe we have

$$m_{\bar{T},[v]}(T([v])) = m_{T,v}(T(v)) + W$$

$$= 0 + W.$$

Thus, $m_{\bar{T},[v]}(x)|m_{T,v}(x)$. \hfill \Box$

The following result can either be proved by the same methods as used in the previous lemma, or one can use the previous result and the definition of the minimal polynomial.

Corollary 4.1.24. Let $W \subseteq V$ be $T$-invariant and let $\bar{T}$ be the induced linear map on $V/W$. Then $m_{\bar{T}}(x)|m_T(x)$.

Definition 4.1.25. Let $T \in \text{Hom}_F(V, V)$ and let $\mathcal{A} = \{v_1, \ldots, v_k\}$ be a set of vectors in $V$. The $T$-span of $\mathcal{A}$ is the subspace

$$W = \left\{ \sum_{i=1}^k p_i(T)(v_i) : p_i(x) \in F[x] \right\}.$$

We say $\mathcal{A}$ $T$-generates $W$.

Exercise 4.1.26. Check that $W$ given in the previous definition is a $T$-invariant subspace of $V$. Moreover, show it is the smallest (with respect to inclusion) $T$-invariant subspace of $V$ that contains $B$.  

54
Theorem 4.1.28. Let $T \in \text{Hom}_F(V,V)$. Let $w \in V$ and let $W$ be the subspace of $V$ that is $T$-generated by $w$. Then

$$\dim_F W = \deg m_{T,w}(x).$$

**Proof.** Let $\deg m_{T,w}(x) = k$. Then $\{w, T(w), \ldots, T^{k-1}(w)\}$ spans $W$ and if any subset spanned we would have $\deg m_{T,w}(x) < k$. This gives the result. \hfill \Box

Up to this point we have dealt with two separate polynomials associated to a linear map: the characteristic and minimal polynomials. It is natural to ask if there is a relation between the two. In fact, there is a very nice relation given by the following theorem.

**Theorem 4.1.28.** Let $T \in \text{Hom}_F(V,V)$. Then

1. $m_T(x)|c_T(x)$;
2. Every irreducible factor of $c_T(x)$ is a factor of $m_T(x)$.

**Proof.** We proceed by induction on $\dim_F V = n$. If $n = 1$ the result is true trivially, so our base case is done. Let $\deg m_T(x) = k \leq n$. Let $v \in V$ such that $m_T(x) = m_{T,v}(x)$. Let $W_1$ be the $T$-span of $v$, so by Lemma 4.1.27 we have $\dim_F W_1 = k$. Set $v_k = v$ and $v_{k-i} = T^i(v)$ for $i = 0, \ldots, k - 1$. Then we claim $B_1 = \{v_1, \ldots, v_k\}$ is a basis for $W_1$. Note that this set has the correct number of elements, so we only need to show it is linearly independent or spans. Suppose there exists $a_1, \ldots, a_k \in F$ so that $a_1 v_1 + \cdots + a_k v_k = 0$. Using the definition of the $v_i$ we have $a_1 T^{k-1}(v) + \cdots + a_k v = 0$. This implies that $m_{T,v}(x) | (a_1 x^{k-1} + \cdots + a_k - 0)$, which is a contradiction unless $a_1 = \cdots = a_{k-1} = 0$ since $\deg m_{T,v}(x) = 0$. This shows the set is a basis. Moreover, we have $\left[ T|_{W_1} \right]_{B_1} = C(m_T(x))$, the companion matrix.

If $k = n$, then $W_1 = V$ so $\left[ T \right]_{B_1} = C(m_T(x))$ has characteristic polynomial $m_T(x)$, i.e., $c_T(x) = m_T(x)$ and we are done.

Now suppose $k < n$. Let $V_2$ be the orthogonal complement of $W_1$ in $V$, and $B_2$ a basis of $V_2$. Set $B = B_1 \cup B_2$. We have $\left[ T \right]_{B} = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$ with $A$ the companion matrix of $m_T(x)$. This allows us to observe that

$$c_T(x) = \det(xI_n - [T]_B)$$
$$= \det \begin{pmatrix} xI_k - A & -B \\ 0 & xI_{n-k} - D \end{pmatrix}$$
$$= \det(xI_k - A) \det(xI_{n-k} - D)$$
$$= m_T(x) \det(xI_{n-k} - D).$$

Thus, we see $m_T(x)|c_T(x)$. This gives the first claim of the theorem.
4.2. \textit{T}-INVARIANT COMPLEMENTS

It remains to show \(c_T(x)\) and \(m_T(x)\) have the same irreducible factors. Note that since \(m_T(x) \mid c_T(x)\) we certainly have every irreducible factor of \(m_T(x)\) is an irreducible factor of \(c_T(x)\). If \(m_T(x)\) has degree \(n\), then \(m_T(x) = c_T(x)\) and the result is trivial. Assume \(\deg m_T(x) < n\). Consider \(T : V/W_1 \to V/W_1\). Set \(B = \pi_{W_1}(B)\). We have \(|T|_B = D\). Since \(\dim_F V/W_1 < \dim_F V\), we can use induction to conclude \(c_T(x)\) and \(m_T(x)\) have the same irreducible factors. Let \(p(x)\) be an irreducible factor of \(c_T(x)\). As above write \(c_T(x) = m_T(x) \det(xI_{n-k} - D)\), which we have is equal to \(m_T(x)c_T(x)\). Since \(p(x)\) is irreducible, it divides \(m_T(x)\) or \(c_T(x)\). If \(p(x) \mid m_T(x)\), we are done. If not, \(p(x) \mid c_T(x)\), and so \(p(x)\) is an irreducible factor of \(c_T(x)\). However, \(m_T(x)\) and \(c_T(x)\) have the same irreducible factors so \(p(x)\) is an irreducible factor of \(m_T(x)\). We now use that \(m_T(x)\mid m_T(x)\) to conclude \(p(x) \mid m_T(x)\).

\[\Box\]

Corollary 4.1.29. Let \(\dim_F V = n\), \(T \in \Hom_F(V, V)\). Then \(V\) is \(T\)-generated by a single element if and only if \(m_T(x) = c_T(x)\).

\textit{Proof.} Let \(w \in V\) and let \(W\) be the subspace of \(V\) that is \(T\)-generated by \(w\). We know that \(\dim_F W = \deg m_{T,w}(x)\). Thus, if there is an element that \(T\)-generates \(V\) we have \(\deg m_{T,w}(x) = n\), and since \(m_{T,w}(x) \mid m_T(x)\) for every \(w\), we have \(\deg m_T(x) = n\) and so it must be equal to \(c_T(x)\). Conversely, if \(m_T(x) = c_T(x)\) then we use the fact that there is an element \(w \in V\) so that \(m_{T,w}(x) = m_T(x)\). Thus, \(\deg m_{T,w}(x) = n\) and so \(\dim_F W = n\), i.e., \(W = V\).

\[\Box\]

These results trivially give the Cayley-Hamilton theorem.

Theorem 4.1.30. [Cayley-Hamilton Theorem]

1. Let \(T \in \Hom_F(V, V)\) with \(\dim_F V < \infty\). Then \(c_T(T) = 0\).
2. Let \(A \in \Mat_n(F)\) and \(c_A(x)\) the characteristic polynomial. Then \(c_A(A) = 0\).

As mentioned above, the term \(c_A(A)\) is a matrix and the content of the theorem above is that this is the zero matrix.

4.2 \textit{T}-invariant complements

We now turn our focus to invariant subspaces and the question of when a \(T\)-invariant subspace has a \(T\)-invariant complement. Recall that given any subspace \(W \subset V\), we always have a subspace \(W' \subset V\) so that \(V = W \oplus W'\). In particular, if \(W\) is \(T\)-invariant we have a complement \(W'\). However, it is not necessarily the case that \(W'\) will also be \(T\)-invariant. As we saw in the last section, this essentially comes down to whether there is a basis \(B\) so that the matrix for \(T\) is block diagonal.

Definition 4.2.1. Let \(W_1, \ldots, W_k\) be subspaces of \(V\) such that \(V = W_1 \oplus \cdots \oplus W_k\). Let \(T \in \Hom_F(V, V)\). We say \(V = W_1 \oplus \cdots \oplus W_k\) is a \(T\)-invariant direct sum if each \(W_i\) is \(T\)-invariant. If \(V = W \oplus W'\) is a \(T\)-invariant direct sum, we say \(W'\) is the \(T\)-invariant complement of \(W\).
We now give two fundamental examples of this. Not only are they useful for understanding the definition, they will be useful in understanding the arguments to follow.

**Example 4.2.2.** Let $T \in \text{Hom}_F(V,V)$ and assume $m_T(x) = c_T(x)$. For convenience write $f(x) = m_T(x)$. Suppose we can factor $f(x) = g(x)h(x)$ with $\gcd(g,h) = 1$. Let $v_0 \in V$ be so that $m_{T,v_0}(x) = m_T(x)$, i.e., $V$ is $T$-generated by $v_0$. 

Set $W_1 = h(T)(V)$ and $W_2 = g(T)(V)$. We claim that $V = W_1 \oplus W_2$ is a $T$-invariant direct sum. Note it is clear that $W_1$ and $W_2$ are both $T$-invariant, so we only need to show that $V = W_1 \oplus W_2$ to see it is a $T$-invariant direct sum.

First, we show that $W_1 \cap W_2 = \{0\}$. Let $w_1 \in W_1$. Then $w_1 = h(T)(v_1)$ for some $v_1 \in V$. We have $g(T)(w_1) = g(T)h(T)(v_1) = f(T)(v_1) = m_T(T)(v_1) = 0$. Thus, $m_{T,w_1} | g(x)$. Similarly, if $w_2 \in W_2$, then $m_{T,w_2} | h(x)$. If $w \in W_1 \cap W_2$, then $m_{T,w}(x) | g(x)$ and $m_{T,w}(x) | h(x)$. Thus $m_{T,w}(x) = 1$. This implies we must have $w = 0$, as desired.

It remains to show that $V = W_1 + W_2$. Since $\gcd(g,h) = 1$, there exists $s,t \in F[x]$ such that $1 = g(x)s(x) + h(x)t(x)$. Thus,

$$v_0 = (g(T)s(T) + h(T)t(T))v_0$$

$$= g(T)s(T)v_0 + h(T)t(T)v_0$$

$$= w_1 + w_2$$

where

$$w_1 = h(T)t(T)v_0$$

$$= h(T)(t(T)v_0) \in h(T)(V) = W_1$$

and

$$w_2 = g(T)s(T)v_0$$

$$= g(T)(s(T)v_0) \in g(T)(V) = W_2.$$

Thus, $v_0 \in W_1 + W_2$. Let $v \in V$. There exists $b(x) \in F[x]$ such that $v = b(T)(v_0)$. This gives

$$v = b(T)v_0$$

$$= b(t)(w_1 + w_2)$$

$$= b(T)(w_1) + b(T)(w_2) \in W_1 + W_2.$$

Thus $V = W_1 \oplus W_2$ as $T$-invariant spaces.

In summary, if $m_T(x) = c_T(x)$ and we can write $m_T(x) = g(x)h(x)$ with $\gcd(g,h) = 1$, then there exists $T$-invariant subspaces $W_1$ and $W_2$ so that $V = W_1 \oplus W_2$. Let $B_1$ be a basis for $W_1$. Then we have

$$[T]_{B_1 \cup B_2} = \begin{pmatrix} [T]_{B_1} & 0 \\ 0 & [T]_{B_2} \end{pmatrix}.$$
Example 4.2.3. As in the previous example, assume \( m_T(x) = c_T(x) \) and again write \( f(x) = m_T(x) \). However, in this case we assume \( f(x) = g(x)h(x) \) with \( \gcd(g, h) > 1 \). For example, \( f(x) \) could be a power of an irreducible polynomial. Let \( v_0 \in V \) such that \( m_T(x) = m_{T,v_0}(x) \), i.e., \( V \) is \( T \)-generated by \( v_0 \). Set \( W_1 = h(T)(V) \). Clearly we have \( W_1 \) is \( T \)-invariant. We will now show \( W_1 \) does not have a \( T \)-invariant complement. Suppose \( V = W_1 \oplus W_2 \) with \( W_2 \) a \( T \)-invariant subspace. Write \( T_1 = T|_{W_1} \) and \( T_2 = T|_{W_2} \).

We claim that \( m_{T_1}(x) = g(x) \). Let \( w_1 \in W_1 \) and write \( w_1 = h(T)(v_1) \) for some \( v_1 \in V \). We have

\[
g(T)(w_1) = g(T)h(T)(v_1) = f(T)(v_1) = m_T(T)(v_1) = 0.
\]

Thus, we must have \( m_{T_1}(x) \mid g(x) \). Set \( w'_1 = h(T)(v_0) \) and \( k(x) = m_{T,w'_1}(x) \). Then we have

\[
0 = k(T)(w'_1) = k(T)h(T)(v_0).
\]

Thus, we have \( m_{T,v_0}(x) = g(x)h(x) \mid k(x)h(x) \). This gives \( g(x) \mid k(x) = m_{T,w_1}(x) \). However, \( m_{T,w'_1}(x) \mid m_{T_1}(x) \) and so \( g(x) \mid m_{T_1}(x) \). This gives \( g(x) = m_{T_1}(x) \) as claimed.

Our second claim is that \( m_{T_2}(x)h(x) \). Let \( w_2 \in W_2 \). Then \( h(T)(w_2) \in W_1 \) because \( h(T)(V) = W_1 \). We also have \( h(T)(w_2) \in W_2 \) because \( W_2 \) is \( T \)-invariant by assumption. However, \( W_1 \cap W_2 = \{0\} \) so it must be the case that \( h(T)(w_2) = 0 \). Hence \( m_{T_2}(x) \mid h(x) \) as claimed.

As we are assuming \( V = W_1 + W_2 \) we can write \( v_0 = w_1 + w_2 \) for some \( w_i \in W_i \). Set \( b(x) = \text{lcm}(g(x), h(x)) \). We have \( b(T)(v_0) = b(T)(w_1) + b(T)(w_2) = 0 \) since \( m_{T_1}(x) = g(x) \mid b(x) \) and \( m_{T_2}(x) \mid h(x) \mid b(x) \). Thus, \( b(x) \) is divisible by \( f(x) = m_{T,v_0}(x) \). However, since \( \gcd(g, h) \neq 1 \) we have \( b(x) \) is a proper factor of \( f(x) \) that vanishes on \( v_0 \), so on \( V \). This contradicts \( m_{T,v_0}(x) = f(x) = c_T(x) \).

Thus, \( W_1 \) does not have a \( T \)-invariant complement.

Now that we have these examples at our disposal, we return to the general situation.

Theorem 4.2.4. Let \( T \in \text{Hom}_F(V, V) \). Let \( m_T(x) = p_1(x) \cdots p_k(x) \) be a factorization into relatively prime polynomials. Set \( W_i = \text{ker}(p_i(T)) \) for \( i = 1, \ldots, k \). Then \( W_i \) is \( T \)-invariant for each \( i \) and \( V = W_1 \oplus \cdots \oplus W_k \) is a \( T \)-invariant direct sum decomposition of \( V \).

Proof. Let \( w_i \in W_i \). We have

\[
p_i(T)(T(w_i)) = T(p_i(T)(w_i)) = T(0) = 0.
\]

58
Thus, $T(w_i) \in \ker(p_i(T)) = W_i$. This gives each $W_i$ is $T$-invariant so it only remains to show $V$ is a direct sum of the $W_i$.

For each $i$ set $q_i(x) = m_T(x)/p_i(x)$. The collection $\{q_1, \ldots, q_k\}$ is relatively prime (not in pairs, just overall) so there are polynomials $r_1, \ldots, r_k \in F[x]$ such that

$$1 = q_1 r_1 + \cdots + q_k r_k.$$ 

Let $v \in V$. We have

$$v = 1 \cdot v = r_1(T) q_1(T) v + \cdots + r_k(T) q_k(T) v = w_1 + \cdots + w_k$$

where we set $w_i = r_i(T) q_i(T)(v)$. We claim that $w_i \in W_i$. Observe

$$p_i(T)(w_i) = p_i(T) q_i(T) r_i(T)(v) = r_i(T) m_T(T)(v) = 0.$$

Thus, $w_i \in W_i$ as claimed. This shows we have $V = W_1 + \cdots + W_k$.

Suppose there exists $w_i \in W_i$ so that $0 = w_1 + \cdots + w_k$. We need to show this implies $w_i = 0$ for each $i$. We have

$$0 = q_1(T)(0) = q_1(T)(w_1 + \cdots + w_k) = q_1(T)(w_1) + \cdots + q_1(T)(w_k)$$

where we have used $p_i | q_1$ for all $i \neq 1$ so $q_1(T)(w_i) = 0$ for $i \neq 1$. Thus, we have $q_1(T)(w_1) = 0$ and by definition $p_i(T)(w_1) = 0$. Since $p_1$ and $q_1$ are relatively prime there exists $r, s \in F[x]$ so that

$$1 = r(x)p_1(x) + s(x)q_1(x).$$

Thus,

$$w_1 = (r(T)p_1(T) + s(T)q_1(T)) w_1 = 0.$$ 

The same argument gives $w_i = 0$ for all $i$. Thus, $V = W_1 \oplus \cdots \oplus W_k$ with $W_i T$-invariant.

One thing to note in the previous result is that given $T$ as in the statement of the theorem, the map $\pi_i = r_i(T) q_i(T)$ is a projection map from $V$ onto $W_i$. This fact is important in the following result. We can now determine exactly how all $T$-invariant subspaces of $V$ arise from the minimal polynomial of $T$.
4.2. **T-INvariant Complements**

**Corollary 4.2.5.** Let $T \in \text{Hom}_F(V, V)$, write $m_T(x) = p_1(x)^{e_1} \cdots p_k(x)^{e_k}$ be the irreducible factorization of $m_T(x)$. Let $W_i = \ker(p_i(T)^{e_i})$. For $i = 1, \ldots, k$ let $U_i$ be a $T$-invariant subspace of $W_i$. (Note we allow $U_i$ to be zero.) Then $U = U_1 \oplus \cdots \oplus U_k$ is a $T$-invariant subspace of $V$. Moreover, if $W$ is any $T$-invariant subspace of $V$, we have

$$W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k),$$

i.e., all $T$-invariant subspaces arise by looking at $T$-invariant subspaces of the $W_i$.

**Proof.** The previous result shows $V = W_1 \oplus \cdots \oplus W_k$ with the $W_i$ being $T$-invariant, so it is clear such a $U$ is a $T$-invariant subspace.

Let $W$ be a $T$-invariant subspace of $V$. Let $w \in W$ and write $w = w_1 + \cdots + w_k$ with $w_i \in W_i$. We know that $w_i = \pi_i(w) = h_i(T)(w)$ for some $h_i(x) \in F[x]$ (see the remark preceding this result.) Now since $W$ is $T$-invariant, each $w_i$ is in $W$ as well. Thus, each $w_i \in W \cap W_i$. This gives

$$W = (W \cap W_1) \oplus \cdots \oplus (W \cap W_k),$$

as desired.

This result reinforces the importance of the minimal polynomial in understanding the structure of $T$. We obtain $T$-invariant subspaces by factoring the minimal polynomial and using the irreducible factors.

The following lemma is essential to prove the theorem that will allow us to give our first significant result on the structure of a linear map: the rational canonical form. The proof of this lemma is a bit of work though and can be skipped without losing any of the main ideas.

**Lemma 4.2.6.** Let $T \in \text{Hom}_F(V, V)$. Let $w_1 \in V$ such that $m_{T,w_1}(x) = m_T(x)$ and let $W_1$ be the $T$-invariant subspace of $V$ that is $T$-generated by $w_1$. Suppose $W_1 \subset V$ is a proper subspace and there is a vector $v_2 \in V$ so that $V$ is $T$-generated by $\{w_1, v_2\}$. Then there is a vector $v_2 \in V$ such that if $W_2$ is the subspace $T$-generated by $v_2$, then $V = W_1 \oplus W_2$.

**Proof.** Let $v_2$ be as in the statement of the theorem. Let $V_2$ be the subspace of $V$ that is $T$-generated by $v_2$. We certainly have $V = W_1 + V_2$, but in general we will not have $W_1 \cap V_2 = \{0\}$. We will obtain the direct sum by changing $v_2$ into a different element so that there is no overlap.

Let $\dim_F V = n$ and $\dim_F W_1 = k$. Then $B_1 = \{T^{k-1}(w_1), \ldots, T(w_1), w_1\}$ is a basis for $W_1$. Set $u_i = T^{k-i}(w_1)$ to ease the notation. We have $V$ is spanned by $B_1 \cup \{v_2', T(v_2'), \ldots\}$ where $v_2' = v_2 + w$ for any $w \in W_1$. Note that for any such $v_2'$ that $\{w_1, v_2\}$ also $T$-generates $V$. The point now is to choose $w$ carefully so that if we set $w_2 = v_2'$ we will have the result. Set $B_2 = \{v_2', T(v_2'), \ldots, T^{n-k-1}(v_2')\}$. The first claim is that $B_1 \cup B_2$ is a basis for $V$. We know that $B_1$ is linearly independent, so we can add elements to it to
form a basis. We add $v'_2, T(v'_2)$, etc. until we obtain a set that is no longer linearly independent. Certainly we cannot go past $T^{n-k-1}(v'_2)$ because then we will have more than $n$ vectors. To see we can go all the way to $T^{n-k-1}(v'_2)$, observe that if $T^j(v'_2)$ is a linear combination of $B_1$ and $\{v'_2, \ldots, T^j-1(v'_2)\}$, then the latter set consisting of $k+j$ vectors spans $V$ so $j \geq n-k$. (Note one must use $W_1$ is $T$-invariant to conclude this. Make sure you understand this point.) However, we know $j \leq n-k$ and so $j = n-k$ as desired. Thus, $B' = B_1 \cup B_2$ is a basis for $V$. Set $u'_{k+1} = T^{n-k-1}(v'_2), \ldots, u'_n = v'_2$.

We now consider $T^{n-k}(u'_n)$. We can uniquely write

$$T^{n-k}(u'_n) = \sum_{i=1}^k b_i u_i + \sum_{i=k+1}^n b_i u'_i$$

for some $b_i \in F$. Set

$$p(x) = x^{n-k} - b_{k+1} x^{n-k-1} - \cdots - b_n x - b_n.$$ 

Set $u = p(T)(v'_2)$ and observe

\[
\begin{align*}
  u &= p(T)(v'_2) \\
  &= p(T)(u'_n) \\
  &= \sum_{i=1}^k b_i u_i \in W_1.
\end{align*}
\]

We now break into two cases:

Case 1: Suppose $u = 0$.

In this case we have $\sum_{i=1}^k b_i u_i = 0$, and so $b_i = 0$ for $i = 1, \ldots, k$ since the $u_i$ are linearly independent. Set $V'_2 = \text{span}_F B_2$. We have

$$T^{n-k}(v'_2) = T^{n-k}(u'_n) = \sum_{i=k+1}^n b_i u'_i \in \text{span}_F B_2.$$ 

Thus, we have $T^j(v'_2) \in \text{span}_F B_2$ for all $j$, so $V_2$ is a $T$-invariant subspace of $V$. By construction we have $W_1 \cap V'_2 = \{0\}$, so we have that $V = W_1 \oplus V'_2$ is a $T$-invariant direct sum decomposition of $V$ and we are done in this case.

Case 2: Suppose $u \neq 0$.

The goal here is to reduce this case to the previous case. We must adjust $v'_2$ for this to work. We now set $V'_2$ to be the space $T$-generated by $v'_2$. We claim that $b_{2k-n+1}, \ldots, b_k$ are all zero so $u = \sum_{i=1}^{2k-n} b_i u_i$. Suppose there exists $b_m$ with $2k-n+1 \leq m \leq k$ so that $b_m \neq 0$ and let $m$ be the largest such index. Since $T$ acts by shifting the $u_i$, we have

$$T^{m-1}(u) = b_m u_1$$

$$T^{m-2}(u) = b_m u_2 + b_{m-1} u_1$$

\[ \vdots \]
4.2. T-IN Variant COMPLEMENTS

Thus, we have
\[ \{ T^{m-1}p(T)(v'_2), T^{m-2}p(T)(v'_2), \ldots, p(T)v'_2, T^{n-k-1}(v'_2), T^{n-k-2}(v'_2), \ldots, v'_2 \} \]
is a linearly independent subset of \( V_2' \). Thus, \( \dim_F V_2' \geq m + n - k \geq k + 1 \). This gives that the degree of \( m_{T,v'_2}(x) \) is at least \( k + 1 \). However, since \( m_{T,v'_2}(x) \) must divide \( m_T(x) \), and \( m_T(x) = m_{T,w_1}(x) \) which has degree \( k \), this is a contradiction. Thus, \( b_{2k-n+1} = \cdots = b_k = 0 \) as claimed.

Set
\[ w = -\sum_{i=1}^{2k-n} b_i u_{i+n-k} \]
Note that the \( u_i \)'s range from \( u_{n-k+1} \) to \( u_k \), so \( w \in W_1 \). Set \( w_2 = v'_2 + w \). Define \( B_1 = \{ u_1, \ldots, u_k \} \) as above, but now set \( B_2 = \{ u_{k+1}, \ldots, u_n \} = \{ T^{n-k-1}(w_2), \ldots, w_2 \} \). Set \( B = B_1 \cup B_2 \). We have
\[
T^{n-k}(u_n) = T^{n-k}(v'_2 + w)
\]
\[
= T^{n-k}(v'_2) + T^{n-k}(w)
\]
\[
= \sum_{i=1}^{2k-n} b_i u_i + T^{n-k} \left( -\sum_{i=1}^{2k-n} b_i u_{i+n-k} \right)
\]
\[
= \sum_{i=1}^{2k-n} b_i u_i + \sum_{i=1}^{2k-n} (-b_i u_i)
\]
\[
= 0.
\]
Thus, we are back in the previous case so we have the result. \( \square \)

This lemma now allows us to prove the following theorem that will be essential to developing the rational and Jordan canonical forms.

Theorem 4.2.7. Let \( T \in \text{Hom}_F(V,V) \) and let \( w_1 \in V \) such that \( m_{T,w_1}(x) = m_T(x) \). Let \( W_1 \) be the subspace \( T \)-generated by \( w_1 \). Then \( W_1 \) has a \( T \)-invariant complement \( W_2 \).

Proof. If \( W_1 = V \), we can set \( W_2 = 0 \) and we are done. Suppose that \( W_1 \) is properly contained in \( V \). Consider the collection of \( T \)-invariant subspaces of \( V \) that intersect \( W_1 \) trivially. We have this is a nonempty set because \( \{0\} \) is a subspace that intersects \( W_1 \) trivially. Partially ordering this by inclusion, we apply Zorn’s lemma to choose a maximal subspace \( W_2 \) that is \( T \)-invariant and intersects \( W_1 \) trivially. We now show that \( V = W_1 \oplus W_2 \).

Suppose that \( W_1 \oplus W_2 \) is properly contained in \( V \). Let \( v \in V \) so that \( v \notin W_1 \oplus W_2 \). Let \( V_2 \) be the subspace of \( V \) that is \( T \)-generated by \( v \). Set \( U_2 = W_2 + V_2 \). If \( W_1 \cap U_2 = \{0\} \) we have a contradiction to \( W_2 \) being maximal. Thus, we must have \( W_1 \cap U_2 \neq \{0\} \). Set \( V' = W_1 + U_2 \). We have \( V' \) is a \( T \)-invariant subspace of \( V \). Set \( T' = T|_{V'} \). Since \( W_2 \) is a \( T \)-invariant space, it is a \( T' \)-invariant subspace of \( V' \). Consider
\[
T' : V'/W_2 \rightarrow V'/W_2.
\]
Set $X = V'/W_2$ and $S = T'$. Let $\pi_{W_2} : V' \rightarrow X$ be the natural projection map and set $\overline{w_1} = \pi_{W_2}(w_1)$ and $\overline{w_2} = \pi_{W_2}(w_2)$. Set $Y_1 = \pi_{W_2}(W_1) \subset X$ and $Z_2 = \pi_{W_2}(U_2) \subset X$. Clearly we have $Y_1$ and $Z_2$ are $S$-invariant subspaces of $X$. The space $Y_1$ is $S$-spanned by $\overline{w_1}$ and $Z_2$ is $S$-spanned by $\overline{w_2}$, so $X$ is $S$-spanned by $\{\overline{w_1}, \overline{w_2}\}$. Finally, since $W_1 \cap W_2 = \{0\}$ we have $\pi_{W_2}|_{W_1}$ is injective.

We have that $m_{T'}(x) | m_T(x)$. We also have $m_S(x) | m_{T'}(x)$. We assumed $m_{T', w_1}(x) = m_T(x)$ and since $\pi_{W_2} : W_1 \rightarrow Y_1$ is injective, we must have $m_{S, \overline{w_1}}(x) = m_{T', w_1}(x)$. Since $w_1 \in V'$, $m_{T', w_1}(x) | m_T(x)$. Finally, $m_{S, \overline{w_1}}(x) | m_S(x)$. Combining all of these gives

$$m_{S, \overline{w_1}}(x) | m_S(x) | m_{T'}(x) | m_T(x) = m_{T', w_1}(x) = m_{S, \overline{w_1}}(x),$$

which gives equality throughout.

We are now in a position to apply the previous lemma to $S, X, \overline{w_1},$ and $\overline{w_2}$. This gives a vector $\overline{w_2}$ so that $X = Y_1 + Y_2$ where $Y_2$ is the subspace of $X$ that is $S$-generated by $\overline{w_2}$. Let $w_2'$ be any element of $V'$ with $\pi_{W_2}(w_2') = \overline{w_2}$, and set $V_2'$ to be the subspace of $V'$ that is $T'$-generated by $w_2'$ (equivalently, the subspace of $V$ that is $T$-generated by $w_2'$). This gives $\pi_{W_2}(V_2') = Y_2$.

We are finally in a position to finish the proof. Observe we have

$$V'/W_2 = X = Y_1 + Z_2 = Y_1 + Y_2.$$

Set $U_2' = W_2 + V_2'$. Then

$$V = W_1 + V_2' + W_2 = W_1 + (W_2 + V_2') = W_1 + U_2'.$$

We have $W_1 \cap U_2' = \{0\}$. To see this, observe that if $x \in W_1 \cap U_2'$, then $\pi_{W_2}(x) \in \pi_{W_2}(W_1) \cap \pi_{W_2}(U_2') = Y_1 \cap Y_2 = \{0\}$. However, if $x \in W_1 \cap V_2'$, then $x \in W_1$ and $\pi_{W_2}|_{W_1}$ is injective so $x = 0$. Thus, we have $V' = W_1 + U_2'$ and $U_2'$ properly contains $W_2$. This contradicts the maximality of $W_2$. \qed

4.3 Rational canonical form

In this section we will give the rational canonical form for a linear transformation. The idea is that we show a “nice” basis exists so that the matrix with respect to the linear transformation is particularly simple. One important feature of the rational canonical form is that this result works over any field. This is in contrast to the Jordan canonical form, which will be presented in the next section.

**Definition 4.3.1.** Let $T \in \text{Hom}_F(V, V)$. An ordered set $C = \{w_1, \ldots, w_k\}$ is a rational canonical $T$-generating set of $V$ if it satisfies

1. $V = W_1 \oplus \cdots \oplus W_k$ where $W_i$ is the subspace $T$-generated by $w_i$;
2. For all \( i = 1, \ldots, k-1 \) we have \( m_{T,w_i}(x) \mid m_{T,w_i}(x) \).

One should note that some textbooks will reverse the order of divisibility of the annihilating polynomials.

The first goal of this section is to show such a set exists. Showing such a set exists is straightforward given Theorem 4.2.7; the work now lies in showing uniqueness.

**Theorem 4.3.2.** Let \( T \in \text{Hom}_F(V, V) \). Then \( V \) has a rational canonical \( T \)-generating set \( C = \{ w_1, \ldots, w_k \} \). Moreover, if \( C' = \{ w'_1, \ldots, w'_l \} \) is any other rational canonical \( T \)-generating set, then \( k = l \) and \( m_{T,w_i}(x) = m_{T,w_i}(x) \) for \( i = 1, \ldots, k \).

**Proof.** Let \( \dim_F V = n \). We induct on \( n \) to prove existence. Let \( w_1 \in V \) such that \( m_T(x) = m_{T,w_1}(x) \) and let \( W_1 \) be the subspace \( T \)-generated by \( w_1 \). If \( W_1 = V \) we are done, so assume not. Let \( W' \) be the \( T \)-invariant complement of \( W_1 \), which we know exists by Theorem 4.2.7. Consider \( T' = T \mid_{W'} \). Clearly we have \( m_{T'}(x) \mid m_T(x) \). Moreover, \( \dim_F W' < n \). By induction \( W' \) has a rational canonical \( T \)-generating set \( \{ w_2, \ldots, w_k \} \). The rational canonical \( T \)-generating set of \( V \) is just \( \{ w_1, \ldots, w_k \} \). This gives existence. It remains to prove uniqueness.

Let \( C = \{ w_1, \ldots, w_k \} \) and \( C' = \{ w'_1, \ldots, w'_l \} \) be rational canonical \( T \)-generating sets with corresponding decompositions

\[
V = W_1 \oplus \cdots \oplus W_k
\]

and

\[
V = W'_1 \oplus \cdots \oplus W'_l.
\]

Let \( p_i(x) = m_{T,w_i}(x) \), \( p'_i(x) = m_{T,w'_i}(x) \), \( d_i = \deg p_i(x) \), and \( d'_i = \deg p'_i(x) \). We proceed by induction on \( k \). If \( k = 1 \), then \( V = W_1 = W'_1 \oplus \cdots \oplus W'_l \). We have \( p_1(x) = m_{T,w_1}(x) = m_T(x) = m_{T,w'_1}(x) = p'_1(x) \). This gives \( d_1 = n \). However, we also have \( \dim_F W_1 = n = \deg m_{T,w_1}(x) = \deg m_T(x) = \deg p'_1(x) \). Thus, \( \dim_F W'_1 = n \) and so \( l = 1 \) and we are done in this case.

Now suppose for some \( m \geq 1 \) we have \( p'_i(x) = p_i(x) \) for all \( 1 \leq i \leq m \). If \( V = W_1 \oplus \cdots \oplus W_m \), then \( n = d_1 + \cdots + d_m = d'_1 + \cdots + d'_m \). Thus \( k = m = m' = l \) and we are done. The same argument works if \( V = W'_1 \oplus \cdots \oplus W'_m \).

Suppose now that \( V \neq W_1 \oplus \cdots \oplus W_m \), i.e., \( k > m \). Consider \( p_{m+1}(T)(V) \). This is \( T \)-invariant. By assumption we have

\[
V = W_1 \oplus \cdots \oplus W_m \oplus W_{m+1} \oplus \cdots.
\]

This gives

\[
p_{m+1}(T)(V) = p_{m+1}(W_1) \oplus \cdots \oplus p_{m+1}(T)(W_m) \oplus p_{m+1}(T)(W_{m+1}) \cdots.
\]

Since \( p_{m+1}(T)(x) = m_{T,w_{m+1}}(x) \), we have \( p_{m+1}(T)(w_{m+1}) = 0 \). We now use that \( W_{m+1} \) is generated by \( w_{m+1} \) to conclude \( p_{m+1}(T)(W_{m+1}) = 0 \) as well. We
also have that $p_{m+j}(x) \mid p_{m+1}(x)$ for all $j \geq 1$. This gives $p_{m+1}(T)(W_{m+j}) = 0$ for all $j \geq 1$. Thus,

$$p_{m+1}(T)(V) = p_{m+1}(T)(W_1) \oplus \cdots \oplus p_{m+1}(T)(W_m).$$

Since $p_{m+1}(x) \mid p_i(x)$ for $1 \leq i \leq m$, we get $\dim_F p_{m+1}(T)(W_i) = d_i - d_{m+1}$.

(See the homework problems.) Thus,

$$\dim_F p_{m+1}(T)(V) = d = (d_1 - d_{m+1}) + \cdots + (d_m - d_{m+1}).$$

We do the same thing to $V = W'_1 \oplus \cdots \oplus W'_j$. We now apply the same argument to $V = W'_1 \oplus \cdots \oplus W'_i \oplus \cdots$ to obtain

$$p_{m+1}(T)(V) = \bigoplus_{j \geq 1} p_{m+1}(T)(W'_i).$$

This has a subspace of dimension $d$ given by $\bigoplus_{j=1}^m p_{m+1}(T)(W'_j)$ by our induction hypothesis. Thus, this must be the entire space since it has dimension equal to the dimension of $p_{m+1}(T)(V)$. Thus, $p_{m+1}(T)(W'_j) = 0$ for $j \geq m+1$.

The annihilator of $W'_{m+1}$ is $p'_{m+1}(x)$, so we must have $p_{m+1}(x) \mid p'_{m+1}(x)$. We now run the same argument with $p'_{m+1}(x)$ instead of $p_{m+1}(x)$ to obtain $p'_{m+1}(x) \mid p_{m+1}(x)$. Thus, $p_{m+1}(x) = p'_{m+1}(x)$ and we are done by induction.

We now rephrase this in terms of matrices.

**Definition 4.3.3.** Let $M \in \text{Mat}_n(F)$. We say $M$ is in *rational canonical form* if $M$ is a block diagonal matrix

$$M = \begin{pmatrix}
C(p_1(x)) & & \\
& C(p_2(x)) & \\
& & \ddots \\
& & & C(p_k(x))
\end{pmatrix}$$

where $C(p_i(x))$ denotes the companion matrix of $p_i(x)$ where $p_1(x), \ldots, p_k(x)$ is a sequence of polynomials satisfying $p_{i+1}(x) \mid p_i(x)$ for $i = 1, \ldots, k - 1$.

**Corollary 4.3.4.**

1. Let $T \in \text{Hom}_F(V, V)$. Then $V$ has a basis $B$ such that $[T]_B = M$ is in rational canonical form. Moreover, $M$ is unique.

2. Let $A \in \text{Mat}_n(F)$. Then $A$ is similar to a unique matrix in rational canonical form.

**Proof.** Let $C = \{w_1, \ldots, w_k\}$ be a rational canonical $T$-generating set for $V$ with $p_i(x) = m_{T,w_i}(x)$. Set $d_i = \deg p_i(x)$. Then

$$B = \{T^{d_1-1}(w_1), \ldots, T(w_1), w_1, \ldots, T^{d_k-1}(w_k), \ldots, w_k\}$$

is the desired basis. To prove the second statement just apply the first part to $T = T_A$. \qed
4.3. RATIONAL CANONICAL FORM

Definition 4.3.5. Let $T$ have rational canonical form with diagonal blocks $C(p_1(x)), \ldots, C(p_k(x))$ with $p_i(x)$ divisible by $p_{i+1}(x)$. We call the polynomials $p_i(x)$ the invariant factors of $T$.

One should note that in some places, such as [4], this are referred to as the elementary divisors. Please see Section 4.6 for clarification on this.

Remark 4.3.6. 1. The rational canonical form is a special case of the fundamental structure theorem for modules over a principal ideal domain. This is where the terminology comes from.

2. Some sources will use invariant factors so that $p_1(x) \mid p_2(x) \mid \cdots \mid p_k(x)$. This is clearly equivalent to our presentation upon a change of basis.

Corollary 4.3.7. Let $A \in \text{Mat}_n(F)$.

1. The matrix $A$ is determined up to similarity by its sequence of invariant factors $p_1(x), \ldots, p_k(x)$.

2. The sequence of invariant factors is determined recursively as follows.

   (a) Set $p_1(x) = m_T(x)$.

   (b) Let $w_1 \in V$ so that $m_T(x) = m_{T,w_1}(x)$ and let $W_1$ be the subspace $T$-generated by $w_1$.

   (c) Let $\overline{T} : V/W_1 \to V/W_1$ and set $p_2(x) = m_{\overline{T}}(x)$.

   (d) Now repeat this process where the starting space for the next step is $V/W_1$.

Proof. The proof is left as an exercise.

Corollary 4.3.8. Let $T \in \text{Hom}_F(V, V)$ have invariant factors $p_1, \ldots, p_k$. Then we have

1. $p_1(x) = m_T(x)$;

2. $c_T(x) = p_1(x) \cdots p_k(x)$.

Proof. We have already seen the first statement. The second statement follows immediately from the definition of $c_T(x)$ and the fact that $\det(C(p(x))) = p(x)$.

The important thing about rational canonical form, as opposed to Jordan canonical form, is that it is defined over any field. It does not depend on the field which one considers the vector space defined over, as the following result shows.

Corollary 4.3.9. Let $F \subseteq K$ fields. Let $A, B \in \text{Mat}_n(F) \subseteq \text{Mat}_n(K)$.

1. The rational canonical form of $A$ is the same whether computed over $F$ or $K$. The minimal polynomial, characteristic polynomial, and invariant factors are the same whether considered over $F$ or $K$.
2. The matrices $A$ and $B$ are similar over $K$ if and only if they are similar over $F$. In particular, there exists some $P \in \text{GL}_n(K)$ such that $B = PAP^{-1}$ if and only if there exists $Q \in \text{GL}_n(F)$ such that $B = QAQ^{-1}$.

**Proof.** Write $M_F$ for the rational canonical form of $A$ computed over $F$. Note this also satisfies the condition for being a rational canonical form over $K$ as well so uniqueness of the rational canonical form gives that $M_F$ is the rational canonical form for $A$ over $K$ as well. Thus, the invariant factors must agree. However, one now just uses Corollary 4.3.8 to obtain the statement about the minimal and characteristic polynomials.

It is important to note for the above result to be applied both matrices must actually be defined over $F$! This does not say if one has two matrices defined over $K$ that they are similar over any subfield of $K$ since they may not even be defined over that subfield.

It is particularly easy to compute the rational canonical form for a matrix in $\text{Mat}_2(F)$ or $\text{Mat}_3(F)$ since the invariant factors are completely determined by the characteristic and minimal polynomials in this case. We illustrate this in the following example. However, we then work the same example with a general method that works for any size matrix. We do not prove this method works as it relies on working with modules over principal ideal domains.

**Example 4.3.10.** Set

$$A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix} \in \text{Mat}_3(\mathbb{Q}).$$

We have

$$c_A(x) = \det \begin{pmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix} = (x-2)^2(x-3).$$

Since we know $m_A(x)$ must have the same irreducible factors as $c_A(x)$, the only possibilities for $m_A(x)$ are $(x-2)(x-3)$ and $(x-2)^2(x-3)$. One easily checks that $(A-2\cdot 1_3)(A-3\cdot 1_3) = 0$ so $m_A(x) = (x-2)(x-3)$. Thus we have $p_1(x) = (x-2)(x-3) = x^2 - 5x + 6$. We now use that $c_A(x)$ is the product of the invariant factors to conclude that $p_2(x) = (x-2)$. Thus, we have
4.3. RATIONAL CANONICAL FORM

CHAPTER 4.

\[ C(p_1(x)) = \begin{pmatrix} 5 & 1 \\ -6 & 0 \end{pmatrix} \] and \[ C(p_2(x)) = 2. \] Thus, the rational canonical form for \( A \) is

\[
\begin{pmatrix}
5 & 1 & 0 \\
-6 & 0 & 0 \\
0 & 0 & 2
\end{pmatrix}
\]

**Example 4.3.11.** Note the minimal and characteristic polynomials do not in general provide enough information to determine the invariant factors for matrices in \( \text{Mat}_n(F) \) for \( n \geq 4 \). For example, if \( A \in \text{Mat}_4(\mathbb{Q}) \) with \( c_A(x) = (x - 1)^4 \), \( m_A(x) = (x - 1)^2 \), then the invariant factors could be \( p_1(x) = (x - 1)^2, p_2(x) = (x - 1)^2 \) or \( p_1(x) = (x - 1)^2, p_2(x) = (x - 1), p_3(x) = x - 1 \). Without more information about \( A \) we cannot determine which is the correct list of invariant factors.

**Example 4.3.12.** We now compute the rational canonical form of \( A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix} \in \text{Mat}_3(\mathbb{Q}) \) in a way that generalizes to matrices in \( \text{Mat}_n(F) \). Consider the matrix

\[
x \cdot 1_3 - A = \begin{pmatrix} x - 2 & 2 & -14 \\ 0 & x - 3 & 7 \\ 0 & 0 & x - 2 \end{pmatrix}
\]

We now apply elementary row and column operations on this matrix to diagonalize it. The fact that this matrix is always diagonalizable is a fact from abstract algebra having to do with the fact that \( F[x] \) is a principal ideal domain. We use standard notation for elementary row and column operations. For example, \( R_1 + R_2 \rightarrow R_1 \) means we replace row 1 by row 1 + row 2.

\[
x \cdot 1_n - A \rightarrow \begin{pmatrix} x - 2 & x - 1 & -7 \\ 0 & x - 3 & 7 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } R_1 + R_2 \rightarrow R_1)
\]

\[
\rightarrow \begin{pmatrix} 1 & x - 1 & -7 \\ x - 3 & x - 3 & 7 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } -C_1 + C_2 \rightarrow C_1)
\]

\[
\rightarrow \begin{pmatrix} 1 & x - 1 & -7 \\ 0 & -x^2 + 5x - 6 & 7x - 14 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } -(x - 3)R_1 + R_2 \rightarrow R_2)
\]

\[
\rightarrow \begin{pmatrix} 1 & x - 1 & -7 \\ 0 & -x^2 + 5x - 6 & 7x - 14 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } -(x - 1)C_1 + C_2 \rightarrow C_2)
\]

\[
\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -x^2 + 5x - 6 & 7x - 14 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } 7C_1 + C_3 \rightarrow C_3)
\]

\[
\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -x^2 + 5x - 6 & 0 \\ 0 & 0 & x - 2 \end{pmatrix} \quad (\text{via } R_2 - 7R_3 \rightarrow R_2).
\]

68
Once the matrix has been diagonalized, the polynomials of degree 1 or greater are the invariant factors. In this case we have \( p_1(x) = x^2 - 5x + 6 \) and \( p_2(x) = x - 2 \), as was found in the previous example. If one keeps track of the elementary row and column operations one can compute \( P \) that converts \( A \) to its rational canonical form as well.

**Example 4.3.13.** In this example we compute a representative of each similarity class of matrices \( A \in \text{Mat}_3(\mathbb{Q}) \) that satisfy \( A^4 = 1_3 \). Note that if \( A^4 = 1_3 \), then \( m_A(x) \mid x^4 - 1 \). We can factor \( x^4 - 1 \) into irreducibles over \( \mathbb{Q} \) to obtain
\[
x^4 - 1 = (x - 1)(x + 1)(x^2 + 1).
\]
Thus, \( m_A(x) \) is a polynomial of degree at most 3 that divides \( (x - 1)(x + 1)(x^2 + 1) \). Conversely, if \( B \in \text{Mat}_3(\mathbb{Q}) \) satisfies that \( m_B(x) \mid x^4 - 1 \), then \( B^4 = 1_3 \). We now list all possible minimal polynomials:

1. \( x - 1 \);
2. \( x + 1 \);
3. \( x^2 + 1 \);
4. \( (x - 1)(x + 1) \);
5. \( (x - 1)(x^2 + 1) \);
6. \( (x + 1)(x^2 + 1) \).

Note we cannot have \( m_A(x) = x^4 - 1 \) because \( \deg m_A(x) \leq 3 \). We can now list the possible invariant factors, keeping in mind the product must have degree 3 and they must divide each other. The possible invariant factors are

1. \( x - 1, x - 1, x - 1 \);
2. \( x + 1, x + 1, x + 1 \);
3. \( (x - 1)(x + 1), x - 1 \);
4. \( (x - 1)(x + 1), x + 1 \);
5. \( (x - 1)(x^2 + 1) \);
6. \( (x + 1)(x^2 + 1) \).

Note the first factor cannot be \( x^2 + 1 \) because then we cannot obtain \( p_2(x) \) so that \( p_2(x) \mid x^2 + 1 \) and \( p_1(x)p_2(x) \) has degree 3. Thus, the elements of \( \text{GL}_3(\mathbb{Q}) \) of order dividing 4, up to similarity, are given by

1. \( 1_3 \);
2. \( -1_3 \);
3. \[
\begin{pmatrix}
0 & 1 \\
1 & 0 \\
1 & 0
\end{pmatrix}
\]
4.3. RATIONAL CANONICAL FORM

4.
\[
\begin{pmatrix}
0 & 1 \\
1 & 0 \\
-1
\end{pmatrix};
\]

5.
\[
\begin{pmatrix}
-1 & 1 & 0 \\
-1 & 0 & 1 \\
-1 & 0 & 0
\end{pmatrix};
\]

6.
\[
\begin{pmatrix}
1 & 1 & 0 \\
-1 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}.
\]

Note in the above matrices we omit the 0’s in some spots to help emphasize the blocks giving the rational canonical form.

In the previous example the possibilities were limited by how \( m_A(x) \) factored over our field. For the next example we consider the same set-up, but working over a different field.

**Example 4.3.14.** Consider \( A \in \text{Mat}_3(\mathbb{Q}(i)) \) satisfying \( A^4 = 1_3 \). In this example we classify all such matrices up to similarity. As above, we have for such an \( A \) that \( m_A(x) \mid x^4 - 1 \), but unlike in the previous example we have \( x^4 - 1 = (x - 1)(x + 1)(x - i)(x + i) \) when we factor it into irreducibles over \( \mathbb{Q}(i) \). This vastly increases the possibilities for the minimal polynomials and hence invariant factors. The possible minimal polynomials are given by

1. \( x - 1; \)
2. \( x + 1; \)
3. \( x - i; \)
4. \( x + i; \)
5. \( x^2 - 1; \)
6. \( x^2 + 1; \)
7. \( (x - 1)(x - i); \)
8. \( (x - 1)(x + i); \)
9. \( (x + 1)(x - i); \)
10. \( (x + 1)(x + i); \)
11. \( (x - 1)(x^2 + 1); \)
12. \( (x + 1)(x^2 + 1); \)
13. \( (x - i)(x^2 - 1); \)
14. \( (x + i)(x^2 - 1). \)
From this, we see the possible invariant factors are given by

1. \( x - 1, x - 1, x - 1; \)
2. \( x + 1, x + 1, x + 1; \)
3. \( x - i, x - i, x - i; \)
4. \( x + i, x + i, x + i; \)
5. \( x^2 - 1, x - 1; \)
6. \( x^2 - 1, x + 1; \)
7. \( x^2 + 1, x - i; \)
8. \( x^2 + 1, x + i; \)
9. \( (x - 1)(x - i), x - 1; \)
10. \( (x - 1)(x - i), x - i; \)
11. \( (x - 1)(x + i), x - 1; \)
12. \( (x - 1)(x + i), x + i; \)
13. \( (x + 1)(x - i), x + 1; \)
14. \( (x + 1)(x - i), x - i; \)
15. \( (x + 1)(x + i), x + 1; \)
16. \( (x + 1)(x + i), x + i; \)
17. \( x^3 - x^2 + x - 1; \)
18. \( x^3 + x^2 + x + 1; \)
19. \( x^3 - ix^2 - x + i; \)
20. \( x^3 + ix^2 - x - i. \)

This gives the following possible rational canonical forms:

1. \( 1_3; \)
2. \( -1_3; \)
3. \( i \cdot 1_3; \)
4. \( -i \cdot 1_3; \)
5. \( \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \)
6. \[
\begin{pmatrix}
0 & 1 \\
1 & 0 \\
-1
\end{pmatrix};
\]
7. \[
\begin{pmatrix}
0 & 1 \\
-1 & 0 \\
-i
\end{pmatrix};
\]
8. \[
\begin{pmatrix}
0 & 1 \\
-1 & 0 \\
i
\end{pmatrix};
\]
9. \[
\begin{pmatrix}
1+i & 1 \\
-i & 0 \\
1
\end{pmatrix};
\]
10. \[
\begin{pmatrix}
1+i & 1 \\
-i & 0 \\
i
\end{pmatrix};
\]
11. \[
\begin{pmatrix}
1-i & 1 \\
i & 0 \\
1
\end{pmatrix};
\]
12. \[
\begin{pmatrix}
1-i & 1 \\
i & 0 \\
-i
\end{pmatrix};
\]
13. \[
\begin{pmatrix}
i-1 & 1 \\
i & 0 \\
-1
\end{pmatrix};
\]
14. \[
\begin{pmatrix}
i-1 & 1 \\
i & 0 \\
i
\end{pmatrix};
\]
15. \[
\begin{pmatrix}
-i-1 & 1 \\
-i & 0 \\
-1
\end{pmatrix};
\]
16. \[
\begin{pmatrix}
-i-1 & 1 \\
-i & 0 \\
-i
\end{pmatrix};
\]
17. \[
\begin{pmatrix}
1 & 1 & 0 \\
-1 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix};
\]
4.4 Jordan canonical form

In this section we use the results on the rational canonical form to deduce the Jordan canonical form. One nice aspect of the Jordan canonical form is that it is given in terms of the eigenvalues of the matrix. One serious drawback is it is only defined over fields containing the splitting field of the minimal polynomial, i.e., the field must contain all the roots of the minimal polynomial. Most treatments of the Jordan canonical form work over an algebraically closed field, and one can certainly do this to be safe. However, for a particular matrix it is not necessary to move all the way to an algebraically closed field so we do not do so.

Definition 4.4.1. Let $T \in \text{Hom}_F(V,V)$ and $\lambda \in F$. If $\ker(T - \lambda \text{id}) \neq 0$, then we say $\lambda$ is an eigenvalue of $T$. Any nonzero vector in this kernel is called an eigenvector of $T$ or a $\lambda$-eigenvector of $T$ if we need to emphasize the eigenvalue. The space $E_1^\lambda = \ker(T - \lambda \text{id})$ is called the eigenspace associated to $\lambda$. More generally, for $k \geq 1$ the $k$th eigenspace of $T$ is given by

$$E_k^\lambda = \{v \in V : (T - \lambda \text{id})^k v = 0\} = \ker((T - \lambda \text{id})^k).$$

We refer to a nonzero element in $E_k^\lambda$ as a generalized $\lambda$-eigenvector of $T$. Write $E_\infty^\lambda = \bigcup_{k=1}^\infty E_k^\lambda$.

One should note that $E_1^\lambda \subset E_2^\lambda \subset \cdots \subset E_k^\lambda \subset \cdots$.

Example 4.4.2. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of $V$ and consider $T \in \text{Hom}_F(V,V)$ with matrix given by

$$A = [T]_\mathcal{B} = \begin{pmatrix}
\lambda & 1 & 0 & 0 & \cdots & 0 \\
0 & \lambda & 1 & 0 & \cdots & 0 \\
& & & & & \\
& & & & & \\
0 & \cdots & 0 & \lambda & 1 & 0 \\
0 & \cdots & 0 & 0 & \lambda & 1 \\
0 & \cdots & 0 & 0 & 0 & \lambda
\end{pmatrix}.$$
We have that each \( v_k \) is a generalized eigenvector. Note that

\[
A - \lambda \cdot 1_n = \begin{pmatrix}
0 & 1 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
& \ddots & \ddots & \ddots & & \\
0 & \cdots & 0 & 1 & 0 \\
0 & \cdots & 0 & 0 & 1 \\
0 & \cdots & 0 & 0 & 0 & 0
\end{pmatrix}.
\]

Thus,

\[
(T - \lambda \cdot \text{id})(v_1) = 0 \\
(T - \lambda \cdot \text{id})(v_2) = v_1 \\
\vdots \\
(T - \lambda \cdot \text{id})(v_n) = v_{n-1}.
\]

This clearly gives \( v_2, \ldots, v_n \notin E^1_\lambda \). Moreover, observe that we have \( \{v_1, \ldots, v_{n-1}\} \subset \text{Im}(T - \lambda \text{id}) \), so \( \dim_F \text{Im}(T - \lambda \text{id}) \geq n - 1 \). However, since \( \dim_F V = n \) and \( \dim_F \ker(T - \lambda \text{id}) \geq 1 \), this gives \( \dim_F \ker(T - \lambda \text{id}) = 1 \) and \( \dim_F \text{Im}(T - \lambda \text{id}) = n - 1 \). This allows us to conclude that \( E^1_\lambda = \text{span}_F \{v_1\} \). Next we consider \( E^2_\lambda \).

It is immediate that \( \{v_1, v_2\} \subset E^2_\lambda \). Since \( (A - \lambda \cdot 1_n)[v_k]_B = [v_{k-1}]_B \), we have \( v_k \notin E^2_\lambda \) for \( k > 2 \). Thus, as above we have \( E^2_\lambda = \text{span}_F \{v_1, v_2\} \). More generally, the same argument gives \( E^k_\lambda = \text{span}_F \{v_1, \ldots, v_k\} \) for \( k = 1, \ldots, n \) and \( E^\infty_\lambda = E^1_\lambda = V \).

**Exercise 4.4.3.** Describe the generalized eigenspaces of the map given by

\[
A = \begin{pmatrix}
\lambda_1 & 1 & 0 & 0 \\
0 & \lambda_1 & 0 & 0 \\
0 & 0 & \lambda_2 & 0 \\
0 & 0 & 0 & \lambda_3
\end{pmatrix}
\]

where \( \lambda_1, \lambda_2, \) and \( \lambda_3 \) are distinct elements of \( F \).

**Exercise 4.4.4.** Let \( T \in \text{Hom}_F(V, V) \). Then \( c_T(x) \) has a root in \( F \) if and only if \( T \) has an eigenvalue in \( F \).

Note the previous exercise shows that if \( (x - \lambda) \mid c_T(x) \), then there is necessarily a nonzero vector in \( V \) of eigenvalue \( \lambda \). Since \( c_T(x) \) and \( m_T(x) \) have the same irreducible factors, this is equivalent to saying \( \lambda \) is an eigenvalue of \( T \) if and only if \( (x - \lambda) \mid m_T(x) \). In fact, we can do better in terms of describing the dimensions of the eigenspaces in terms of the minimal polynomial and characteristic polynomial.

**Lemma 4.4.5.** Let \( T \in \text{Hom}_F(V, V) \) and suppose that \( m_T(x) = (x - \lambda)^e p(x) \) with \( p(\lambda) \neq 0 \). Then \( E^\infty_\lambda = E^e_\lambda \).
Proof. Let \( v \in V \) and let \( m \) be the least positive integer so that \((T - \lambda \text{id})^m(v) = 0\). Suppose that \( m > e \). Then we have \( m_{T,v}(x) \mid (x - \lambda)^m \), but \( m_{T,v}(x) \mid (x - \lambda)^{m-1} \) and so \( m_{T,v}(x) = (x - \lambda)^m \). However, we know \( m_{T,v}(x) \mid m_T(x) \), which is a contradiction if \( m > e \). Thus, \( E_\lambda^\infty = E_\lambda^e \).

**Lemma 4.4.6.** Let \( T \in \text{Hom}_F(V,V) \) and suppose \( m_T(x) = c_T(x) = (x - \lambda)^n \) for some \( \lambda \in F \). Then \( V \) is \( T \)-generated by a single element \( w_1 \) and \( V \) has a basis \( \{v_1, \ldots, v_n\} \) where \( v_n = w_1 \), and \( v_i = (T - \lambda \text{id})(v_{i+1}) \) for \( i = 1, \ldots, n - 1 \).

**Proof.** Let \( w_1 \in V \) be such that \( m_{T,w_1}(x) = m_T(x) = c_T(x) \). Let \( W_1 \) be the subspace \( T \)-generated by \( w_1 \). Then \( \dim_F W_1 = \deg m_{T,w_1}(x) = \deg c_T(x) = \dim_F V \) and so \( W_1 = V \). Set \( v_n = w_1 \) and define \( v_i = (T - \lambda \text{id})^{n-i}(v_n) \). Observe we have
\[
v_i = (T - \lambda \text{id})^{n-i}(v_n) \\
= (T - \lambda \text{id})(T - \lambda I)^{n-i-1}(v_n) \\
= (T - \lambda \text{id})(v_{i+1}).
\]
Thus, we only need to show that \( B = \{v_1, \ldots, v_n\} \) is a basis for \( V \). This has \( \dim_F V \) elements, so it is enough to check \( B \) is linearly independent. Suppose there are scalars \( c_1, \ldots, c_n \in F \) such that
\[
c_1v_1 + \cdots + c_nv_n = 0.
\]
This gives
\[
c_1(T - \lambda \text{id})^{n-1}(v_n) + \cdots + c_{n-1}(T - \lambda \text{id})v_n + c_nv_n = 0.
\]
If we set \( p(x) = c_1(x - \lambda)^{n-1} + \cdots + c_{n-1}(x - \lambda) + c_n \), then \( p(T)(v_n) = 0 \), i.e., \( p(T)(w_1) = 0 \). This gives \( m_{T,w_1}(x) \mid p(x) \). However, \( m_{T,w_1}(x) = c_T(x) \) has degree \( n \) and \( \deg p(x) = n - 1 \), so it must be that \( p(x) = 0 \), i.e., \( c_j = 0 \) for all \( j \).

**Corollary 4.4.7.** Let \( T \) and \( B \) be as in the previous lemma. Then
\[
[T]_B = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \lambda & 1 & 0 \\ 0 & \cdots & 0 & 0 & \lambda & 1 \\ 0 & \cdots & 0 & 0 & 0 & \lambda \end{pmatrix}.
\]

**Proof.** We have \((T - \lambda \text{id})v_1 = 0\), so \( Tv_1 = \lambda v_1 \). For \( i > 1 \) we have \((T - \lambda \text{id})v_{i+1} = v_i \), so \( T(v_{i+1}) = v_i + \lambda v_{i+1} \). This gives the correct form for the matrix.

**Definition 4.4.8.** A basis \( B \) as in Lemma 4.4.6 is called a *Jordan basis for* \( V \). Moreover, generally if \( V = V_1 \oplus \cdots \oplus V_k \) is a \( T \)-invariant decomposition and each \( V_i \) has a Jordan basis \( B_i \), then we call \( B = \cup B_i \) a *Jordan basis for* \( V \).
4.4. JORDAN CANONICAL FORM

Definition 4.4.9. 1. A matrix $A \in \text{Mat}_k(F)$ of the form
\[
\begin{pmatrix}
\lambda & 1 & 0 & 0 & \cdots & 0 \\
0 & \lambda & 1 & 0 & \cdots & 0 \\
& & \ddots & \ddots & & \vdots \\
0 & \cdots & 0 & \lambda & 1 & 0 \\
0 & \cdots & 0 & 0 & \lambda & 1 \\
0 & \cdots & 0 & 0 & 0 & \lambda
\end{pmatrix}
\]

is called a $k \times k$ Jordan block associated to the eigenvalue $\lambda$.

2. A matrix $J$ is said to be in Jordan canonical form if it is a block diagonal matrix where each $J_i$ is a Jordan Block.

Theorem 4.4.10. 1. Let $T \in \text{Hom}_F(V,V)$. Suppose that
\[
c_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}
\]
over $F$. Then $V$ has a basis $B$ such that $J = [T]_B$ is in Jordan canonical form. Moreover, $J$ is unique up to the order of the blocks.

2. Let $A \in \text{Mat}_n(F)$. Suppose
\[
c_A(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}
\]
over $F$. Then $A$ is similar to a matrix $J$ in Jordan canonical form. Moreover, $J$ is unique up to the order of the blocks.

Proof. Set $p_i(x) = (x - \lambda_i)^{e_i}$. Set $V^i = E_{\lambda_i}^i = \ker(p_i(T))$. Note this $V^i$ is the $\lambda_i$-eigenspace of $V$ by Lemma 4.4.5. We can apply Theorem 4.2.4 to conclude that
\[
V = E_{\lambda_1}^{e_1} \oplus \cdots \oplus E_{\lambda_k}^{e_k}.
\]
Set $T_i = T|_{V^i}$. Then one can show that $c_{T_i}(x) = (x - \lambda_i)^{e_i}$. Choose a rational canonical $T_i$-generating set $C = \{w_{i1}^1, \ldots, w_{im_i}^i\}$ with corresponding direct sum decomposition
\[
V^i = W_{i1}^1 \oplus \cdots \oplus W_{im_i}^i
\]
where $W_j^i$ is the $T_i$-generated subspace by $w_j^i$. Note that each $W_j^i$ satisfies the hypotheses of Lemma 4.4.6 so we have a Jordan basis $B_j^i$ for $W_j^i$ with respect to $T_i$. The desired basis is
\[
B = \bigcup_{j=1}^k \bigcup_{i=1}^{m_i} B_j^i.
\]
The uniqueness follows from uniqueness of the rational canonical form and our construction.

The second claim follows easily from the first. \qed
One should note that if $F$ is algebraically closed the characteristic polynomial always splits completely into linear factors. In particular, if $F$ contains all the roots of $c_T(x)$ then $T$ can be put in Jordan canonical form over $F$. If $c_T(x)$ does not split over $F$ we cannot put $T$ in Jordan canonical form over $F$. This makes the rational canonical form more useful in such situations.

We now give an algorithm for computing the Jordan canonical form. Our first worked example is particularly simple as the matrix is already in Jordan canonical form. However, it is nice to see how the algorithm works on a very simple example before giving a less trivial one.

**Example 4.4.11.** Consider the matrix $A = \begin{pmatrix} 6 & 1 & 0 \\ 0 & 6 & 1 \\ 0 & 0 & 6 \end{pmatrix}$. One easily calculate that $c_A(x) = (x - 6)^5(x - 7)^3$. This immediately gives the only nontrivial generalized eigenspaces are associated to 6 and 7. Moreover, from the powers on the linear terms we know $E_6^\infty = E_j^6$ for some $j = 1, \ldots, 5$ and $E_7^\infty = E_i^7$ for some $i = 1, 2, 3$.

Consider $\lambda = 6$. We compute the dimension of $\ker(A - 6 \cdot 1_8)$. Note that

$A - 6 \cdot 1_8 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

It is now immediate that $\dim_F E_6^3 = \dim_F (A - 6 \cdot 1_8) = 3$ and has basis $\{e_1, e_4, e_5\}$. The next step is to compute $E_6^2 = \ker(A - 6 \cdot 1_8)^2$. This has dimension 4 with basis $\{e_1, e_2, e_4, e_5\}$. Finally, $E_6^1 = \ker(A - 6 \cdot 1_8)^3$ has dimension 5 and is spanned by $\{e_1, e_2, e_3, e_4, e_5\}$. We represent this graphically as follows. We begin with a horizontal line with nodes the basis elements of $E_6^1$. 

77
We then add a second row of nodes by adding basis elements that are in $E_6^2 - E_6^1$. In this case it means we only add $e_2$. Note we can put it over any element of the first row since we are only looking for the size of the blocks here, so for convenience we put it over $e_1$.

Finally, we build on that by adding a third row with basis elements that are in $E_6^3 - E_6^2$. Since there is only one element, it must go over the $e_2$ as there can be no gaps in the vertical lines through the nodes we add.

This gives us all the information we need for the Jordan blocks associated to the eigenvalue 6. The number of nodes on the first horizontal line tells us the
number of Jordan blocks with 6 on the diagonal, and the height over each of these nodes tells the size of the block. So we have three blocks with 6’s on the diagonal: one of size 3, and two of size 1.

We now move on to the eigenvalue 7. We compute $E^1_7 = \ker(A - 7 \cdot 1_8)$ has dimension 2 with basis $\{e_6, e_8\}$. We have $E^2_7$ has dimension 7 with basis $\{e_6, e_7, e_8\}$. We now position these on a graph as above, giving only the final result here:

\[
\begin{align*}
&\begin{align*}
&e_7 \\
&e_6 \\
&e_8
\end{align*}
\end{align*}
\]

Thus, we have two blocks with 7’s on the diagonal, one of size 2 and one of size 1. Putting this into matrix form gives us what we already knew, namely, that $A$ is already in Jordan canonical form.

Our next example is a bit more involved.

**Example 4.4.12.** Consider the matrix

\[
A = \begin{pmatrix}
3 & 3 & 0 & 0 & 0 & -1 & 0 & 2 \\
-3 & 4 & 1 & -1 & -1 & 0 & 1 & -1 \\
0 & 6 & 3 & 0 & 0 & -2 & 0 & -4 \\
-2 & 4 & 0 & 1 & -1 & 0 & 2 & -5 \\
-3 & 2 & 1 & -1 & 2 & 0 & 1 & -2 \\
-1 & 1 & 0 & -1 & -1 & 3 & 1 & -1 \\
-5 & 10 & 1 & -3 & -2 & -1 & 6 & -10 \\
-3 & 2 & 1 & -1 & -1 & 0 & 1 & 1
\end{pmatrix}
\]

One calculates

\[
c_A(x) = (x - 2)(x - 3)^5(x^2 - 6x + 21) = (x - 2)(x - 3)^5(x - (3 + 2\sqrt{-3}))(x - (3 - 2\sqrt{-3})).
\]

Thus, $A$ does not have Jordan canonical form over $\mathbb{Q}$, but does over $\mathbb{Q}(\sqrt{-3})$. We now compute the Jordan canonical form over $\mathbb{Q}(\sqrt{-3})$. Note that the eigenvalues $2, 3 \pm 2\sqrt{-3}$ are very easily to deal with. Since they only occur to multiplicity
one, there can only be one Jordan block for each of these eigenvalues, each of size 1. Thus, the only work comes in determining the block structure for the eigenvalue 3.

We compute the Jordan block structure for the eigenvalue 3 just as in the previous example. The difference here is it requires more work to compute the size of the eigenspaces, and we don’t keep track of the bases of the eigenspaces since we are only looking for the canonical form and not the basis that realizes the form. One could keep track of the bases using elementary linear algebra at each step if one so desired. One computes, using elementary linear algebra, that \( \dim_{\mathbb{Q}(\sqrt{-3})} E_3^1 = 2 \). Let \( \{v_1, v_2\} \) be a basis for this space. One then calculates \( \dim_{\mathbb{Q}(\sqrt{-3})} E_3^2 = 4 \), so we add vectors \( v_3 \) and \( v_4 \) to obtain a basis \( \{v_1, v_2, v_3, v_4\} \) for \( E_3^2 \). We know \( E_3^\infty \) has dimension 5, so it must be that \( E_3^\infty = E_3^3 \) and so we have a basis \( \{v_1, v_2, v_3, v_4, v_5\} \) for \( E_3^3 \). Graphically, we have for the first step:

Since we add two vectors going from \( E_3^1 \) to \( E_3^2 \), we add \( v_3 \) over \( v_1 \) and \( v_4 \) over \( v_2 \) obtaining:

Finally, we add the final vector over \( v_3 \) to obtain the final graph:
Thus, we have two Jordan blocks associated to the eigenvalue 3, one of size 3 and one of size 2. Thus, the Jordan canonical form of $A$ is given by

$$
\begin{pmatrix}
3 & 1 & 0 \\
0 & 3 & 1 \\
0 & 0 & 3 \\
\end{pmatrix}
\begin{pmatrix}
3 & 1 \\
0 & 3 \\
\end{pmatrix}
\begin{pmatrix}
2 \\
3 + 2\sqrt{-3} \\
3 - 2\sqrt{-3} \\
\end{pmatrix}
$$

4.5 Semi-simple and diagonalizable operators

In the past few sections we have seen examples of how to pick a nice basis so that a linear map has a nice form with respect to this basis. In this section we use these results to recover some results from elementary linear algebra as well as add a few results that can be obtained when one does not have Jordan canonical form or when one has multiple linear transformations.

Definition 4.5.1. Let $T \in \text{Hom}_F(V,V)$. We say $T$ is diagonalizable if there is a basis $\mathcal{B}$ of $V$ so that $[T]_{\mathcal{B}}$ is a diagonal matrix.

We have the following result from elementary linear algebra on diagonalizability.

Corollary 4.5.2. Let $T \in \text{Hom}_F(V,V)$. If $c_T(x)$ does not split into a product of linear factors, then $T$ is not diagonalizable. If $c_T(x)$ does split into a product of linear factors then the following are equivalent:
1. $T$ is diagonalizable;

2. $m_T(x)$ splits into a product of distinct linear factors;

3. For every eigenvalue $\lambda$, $E_\lambda^\infty = E_\lambda^1$;

4. For every eigenvalue $\lambda$ of $T$ if $c_T(x) = (x-\lambda)^{e_\lambda}p(x)$ with $p(\lambda) \neq 0$, then $e_\lambda = \dim_F E_\lambda^1$;

5. If we set $d_\lambda = \dim_F E_\lambda^1$, then $\sum d_\lambda = \dim_F V$;

6. If $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of $T$, then $V = E_{\lambda_1}^1 \oplus \cdots \oplus E_{\lambda_m}^1$.

Proof. Suppose $T$ is diagonalizable. There is a basis $B$ such that

$$[T]_B = D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix}$$

for some $\lambda_1, \ldots, \lambda_n \in F$ where the $\lambda_i$ are not assumed to be distinct. Then $c_T(x) = \det(x1_m - D) = (x-\lambda_1) \cdots (x-\lambda_m)$. Thus, $c_T(x)$ splits into linear factors. Equivalently, if $c_T(x)$ does not split into linear factors then $T$ is not diagonalizable.

Now suppose $c_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_m)^{e_m}$ for some $\lambda_i \in F$ and $e_i \in \mathbb{Z}_{\geq 1}$ where we now assume the $\lambda_i$ are distinct. Note that since $c_T(x)$ splits into linear factors, the Jordan canonical form for $T$ exists over $F$.

Suppose $T$ is diagonalizable. We have from the proof of the existence of the Jordan canonical form of $T$ that $V = V^1 \oplus \cdots \oplus V^m$ where $V^i = \ker(T-\lambda_i)^{e_i} = E_{\lambda_i}^{e_i}$. Each $V^i$ splits into a direct sum of subspaces coming from the rational canonical form of $T_i = T|_{V^i}$. In particular, we can write

$$V^i = W^i_1 \oplus \cdots \oplus W^i_{m_i}$$

where each $W^i_j$ is generated by an element $w^i_j$. The dimension of each $W^i_j$ corresponds to the size of the corresponding block in the rational canonical form. Thus, if $\dim_F W^i_j > 1$, then the rational canonical form for $T$ restricted to $W^i_j$ is larger than a 1 by 1 matrix, i.e., $T$ is not diagonalizable. This shows we have each $W^i_j$ is 1-dimensional. Now we have $W^i_j$ is generated by a single element $w^i_j$, $m_{T_i,w^i_j}(x)| m_{T_i,W^i_j}(x)$ for each $i$, and $m_{T_i}(x) = m_{T_i,w^i_j}(x)$. Thus, $m_{T_i}(x) = (x-\lambda_i)$. We have $m_{T_i}(x)$ is the least common multiple of the $m_{T_i}(x)$, and so $m_T(x)$ splits into a product of distinct linear factors, i.e., we have 1) implies 2).

Now suppose that $m_T(x)$ splits into a product of distinct linear terms. We saw above that if $m_T(x) = (x-\lambda_1) \cdots (x-\lambda_m)$ with the $\lambda_i$ distinct, then Lemma 4.4.5 gives $E_\lambda^\infty = E_\lambda^1$ and so 2) implies 3).

Observe that if $c_T(x) = (x-\lambda)^{e}p(x)$, then we have $\dim_F E_\lambda^\infty = e$. (This is a homework exercise.) Assume that for every eigenvalue $\lambda$ we have $E_\lambda^\infty = E_\lambda^1$. This gives the result and so 3) implies 4).
Suppose that we are given for each \( \lambda \) that \( \dim_F E^1_\lambda = e_\lambda \). Then since \( \deg c_T(x) = n \), we must have \( d_{\lambda_1} + \cdots + d_{\lambda_m} = e_{\lambda_1} + \cdots + e_{\lambda_m} = n \), i.e., 4) implies 5).

We now show 5) implies 6). Suppose \( d_{\lambda_1} + \cdots + d_{\lambda_m} = n \). We know \( E^1_{\lambda_1} \oplus \cdots \oplus E^1_{\lambda_m} \subset V \). However, since they have the same dimension they must be equal.

Finally, it only remains to show 6) implies 1). Suppose we have \( V = E^1_{\lambda_1} \oplus \cdots \oplus E^1_{\lambda_m} \). This gives that each Jordan block can have size at most 1, so in particular the Jordan canonical form is a diagonal matrix.

One does not need the existence of Jordan canonical form to prove the above result as there are elementary proofs. However, this proof has the added benefit of reinforcing the important points of the proof of the existence of Jordan canonical form. One should note that \( c_T(x) \) factoring into linear factors does not imply \( T \) is diagonalizable (this is the entire point of the Jordan canonical form), but if it splits into distinct linear factors then one does have \( T \) is diagonalizable since \( m_T(x) \mid c_T(x) \).

We saw in the previous section that a linear map could fail to be diagonalizable for a couple of reasons. The first is fairly easy to deal with. For instance, consider the matrix \( A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \text{Mat}_2(\mathbb{R}) \). The characteristic polynomial of this is \( c_A(x) = x^2 + 1 \), so \( A \) is not diagonalizable over \( \mathbb{R} \) because its eigenvalues, namely \( \pm i \), are not in \( \mathbb{R} \). However, if we consider this as a matrix in \( \text{Mat}_2(\mathbb{C}) \), then since the eigenvalues are distinct we see the Jordan canonical form is \( \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \), and so it is diagonalizable over \( \mathbb{C} \). This motivates the following definition.

**Definition 4.5.3.** Let \( T \in \text{Hom}_F(V, V) \). We say \( T \) is **potentially diagonalizable** if \( T \) is diagonalizable over a field extension \( K/F \). (This is also referred to as **absolutely semi-simple**.)

We have a criterion for a linear map to be diagonalizable in terms of its minimal polynomial, namely, the minimal polynomial should split into distinct linear factors over the field. We also have a criterion in terms of the minimal polynomial for a linear map to be potentially diagonalizable. First, we recall a definition from abstract algebra.

**Definition 4.5.4.** A polynomial \( f \in F[x] \) is said to be **separable** if it has no repeated roots.

**Exercise 4.5.5.** Show \( f \) is separable if and only if \( \gcd(f, f') = 1 \) where \( f' \) denotes the formal derivative of \( f \).

The power of the previous exercise is it allows one to check for separability without having to find the roots of \( f \). Finding the greatest common divisor of two polynomials depends only on the Euclidean algorithm in \( F[x] \), which is very fast and easy to use; finding roots of polynomials is very hard.
Proposition 4.5.6. A linear map $T \in \text{Hom}_F(V, V)$ is potentially diagonalizable if and only if the minimal polynomial $m_T(x) \in F[x]$ is separable.

Proof. Suppose $T$ is potentially diagonalizable. Let $K/F$ be the field so that $T$ is diagonalizable over $K$. Then we know that the minimal polynomial of $T$ is the same when considered over $F$ or $K$. Since $T$ is diagonalizable over $K$, the minimal polynomial can have no repeated roots over $K$, so certainly no repeated roots over $F$.

Now suppose $m_T(x)$ is separable. Let $K$ be the splitting field of $m_T(x)$ over $F$, i.e., the field obtained upon adjoining all the roots of $m_T(x)$ to $F$. Since $m_T(x)$ is separable and $K$ contains all the roots of $m_T(x)$, it splits into distinct linear factors over $K$. Thus, $T$ is diagonalizable over $K$. Of course, there are linear maps that are not diagonalizable or potentially diagonalizable. For example, we know the matrix
$$
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}
$$
in $\text{Mat}_2(Q)$ is not even potentially diagonalizable because it is already in Jordan canonical form, so it cannot be diagonalized over a larger field because the Jordan canonical form is unique when it exists. We consider two more types of linear maps here.

Definition 4.5.7. Let $T \in \text{Hom}_F(V, V)$. We say $T$ is simple if the only $T$-invariant subspaces of $V$ are $\{0\}$ and $V$.

Example 4.5.8. Consider $V = \mathbb{R}^2$ and let $T$ be the linear map given by rotation by any fixed $\alpha$ radians with $\alpha \neq 0$. This is clearly a linear map, but the only fixed subspaces are $V$ and $\{0\}$, so it is a simple map.

As in the above cases, we can characterize simple linear maps in terms of the minimal polynomial.

Theorem 4.5.9. Let $T \in \text{Hom}_F(V, V)$. The following are equivalent:

1. $T$ is simple,
2. $m_T(x)$ is irreducible in $F[x]$ and has degree $\text{dim}_F V$,
3. $c_T(x)$ is irreducible in $F[x]$,
4. $c_T(x) = m_T(x)$ is irreducible in $F[x]$.

Proof. If $m_T(x)$ is reducible or has degree less than $n = \text{dim}_F V$, we use the rational canonical form of $T$ to obtain a contradiction. Namely, write $c_T(x) = h(x)m_T(x) = h(x)g_1(x)g_2(x)$. Recall that if $p_1(x), \ldots, p_k(x)$ are the invariant factors, then $c_T(x) = p_1(x) \cdots p_k(x)$ and $p_1(x) = m_T(x)$. Thus, we must have $\deg p_2(x) \geq 1$, and so $\ker p_2(T)$ gives a non-trivial proper $T$-invariant subspace. So it must be the case that $\deg m_T(x) = \deg c_T(x)$. If $m_T(x)$ is reducible, then Theorem 4.2.4 gives non-trivial proper $T$-invariant subspaces of $V$. Thus, we see that if $m_T(x)$ is reducible or has degree less than the degree of $c_T(x)$, then $T$ is not simple. Thus, we have (1) implies (2).
4.5. DIAGONALIZABLE OPERATORS

If \( m_T(x) \) is irreducible and has degree \( \dim_F V \), then since \( m_T(x) \mid c_T(x) \) and \( \deg c_T(x) = n \), we have \( m_T(x) = c_T(x) \) and so \( c_T(x) \) is irreducible. This shows (2) implies (3) and (4). We also immediately have (3) implies (4) because \( m_T(x) \mid c_T(x) \).

Finally, to see (4) implies (1) just apply Corollary 4.2.5. \( \square \)

The last class of linear maps we will deal with are semi-simple linear maps.

**Definition 4.5.10.** Let \( T \in \text{Hom}_F(V, V) \). We say \( T \) is semi-simple if every \( T \)-invariant subspace has a \( T \)-invariant complement.

Just as above, we can classify these in terms of the minimal polynomial of \( T \).

**Theorem 4.5.11.** Let \( T \in \text{Hom}_F(V, V) \). Then \( T \) is semi-simple if and only if \( m_T(x) \) is square-free.

**Proof.** Assume \( m_T(x) \) is square-free and let \( W \) be a \( T \)-invariant subspace of \( V \). Let \( m_T(x) = p_1(x) \cdots p_k(x) \) be the factorization of \( m_T(x) \) into irreducible factors. We have that \( V = W_1 \oplus \cdots \oplus W_k \) where \( W_i = \ker p_i(T) \). Set \( W_i = W \cap V_i \), so \( W = \oplus_{j=1}^k W_j \). Finding a \( T \)-invariant complement of \( W \) in \( V \) is equivalent to finding a \( T \)-invariant complement of \( W_i \) in \( V_i \). The easiest way to do this is to use a little bit of module theory, but it can be done with vectors as well. Observe that since \( p_i(x) \) is an irreducible polynomial, we have \( K = F[x]/(p_i(x)) \) is a field. We claim \( V_i \) is a \( K \)-vector space. We define scalar multiplication by

\[
[f(x)] \cdot v = f(T)(v).
\]

The fact that \( p_i(T) \) kills \( V_i \) gives that this scalar multiplication is well-defined. It is now elementary to check that \( V_i \) is a \( K \)-vector space. Moreover, we have that \( W_i \) is a subspace of the \( K \)-vector space \( V_i \), and so has a complement \( W'_i \) as a \( K \)-vector space. Since \( W'_i \) is an \( K \)-subspace, it is necessarily an \( F \)-subspace that is also \( T \)-invariant. Thus, \( W' = \sum_{i=1}^k W'_i \) is the \( T \)-invariant complement of \( W \).

Now assume that \( m_T(x) \) is divisible by \( p_1(x)^2 \) for \( p_1(x) \) irreducible. We now just apply the argument given in Example 4.2.3 to construct a \( T \)-invariant subspace with not \( T \)-invariant complement. Thus, if \( m_T(x) \) is not square-free then \( T \) is not semi-simple. \( \square \)

As an easy corollary we have the following.

**Corollary 4.5.12.** Let \( T \in \text{Hom}_F(V, V) \) with \( c_T(x) \) square-free. Then \( T \) is semi-simple and \( m_T(x) = c_T(x) \).

**Proof.** We have \( m_T(x) \mid c_T(x) \), so clearly \( m_T(x) \) must be square-free so \( T \) is semi-simple. Moreover, \( m_T(x) \) and \( c_T(x) \) have the same irreducible factors, so if \( c_T(x) \) is square-free necessarily \( m_T(x) = c_T(x) \). \( \square \)
Note that a polynomial that splits with distinct roots is clearly separable. Thus, diagonalizable implies potentially diagonalizable, which also follows immediately from the definition. One also has a separable polynomial has no repeated factors, so potentially diagonalizable implies semi-simple. Clearly a simple linear map is semi-simple as well. The reverse implications are not true. For example, we have seen above already a map that is potentially diagonalizable but not diagonalizable. We also have the following exercise.

**Exercise 4.5.13.** Give an example of a semi-simple linear map that is not simple.

It isn’t as easy to give a linear map that is semi-simple but not potentially diagonalizable. This has to do with the subtle difference of a polynomial being separable as opposed to having distinct factors. For example, if the field one is working over has characteristic 0, it is not possible to give a polynomial that has distinct factors but is not separable. (This is true over any perfect field; it does not really require characteristic 0.) Over a perfect field semisimplicity and potential diagonalizability are the same thing; over an algebraically closed field semisimplicity is the same as diagonalizability. To see these are not the same in general, we consider a non-perfect field.

**Example 4.5.14.** Let $F = \mathbb{F}_2(t)$, i.e., the field consisting of ratios of polynomials with coefficients in $\mathbb{F}_2$. This is not a perfect field. For instance, $t$ is not a square in $F$. Consider the matrix $A = \begin{pmatrix} 0 & t \\ 1 & 0 \end{pmatrix}$. Then $c_A(x) = x^2 - t$, so $c_A(x)$ is irreducible and so it is square-free. Thus, $A$ is a semi-simple linear map on $F^2$. (It is actually a simple linear map.) Now if we consider the splitting field of $c_A(x)$, this is given by $K = \mathbb{F}(\sqrt{t})$. Note that over $K$ we have $c_A(x) = (x - \sqrt{t})^2$ since $K$ has characteristic 2. Thus, $A$ is not potentially diagonalizable. Thus, we have a linear map that is semi-simple but not potentially diagonalizable.

We have seen in the last two sections that in many instances one can choose a basis so that the matrix representing a linear transformation is in a very nice form. However, it may be the case that one is interested in more than one linear transformation at a time. In that case, the results just given aren’t as useful because while a basis might put one linear transformation into a nice form, it does not necessarily put the other transformation in a nice form. We do have the following result in this direction.

**Theorem 4.5.15.** Let $S, T \in \text{Hom}_F(V, V)$ with each diagonalizable. The following are equivalent:

1. There is a basis $\mathcal{B}$ of $V$ such that $[S]_{\mathcal{B}}$ and $[T]_{\mathcal{B}}$ are diagonal, i.e., $\mathcal{B}$ consists of common eigenvalues of $S$ and $T$;

2. $S$ and $T$ commute.

**Proof.** First suppose there is a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ so that $[S]_{\mathcal{B}}$ and $[T]_{\mathcal{B}}$ are both diagonal. Then there exists $\lambda_i, \mu_i \in F$ such that $S(v_i) = \lambda_i v_i$ and $T(v_i) = \mu_i v_i$. Since $S$ and $T$ commute, $[S]_{\mathcal{B}} [T]_{\mathcal{B}} = [T]_{\mathcal{B}} [S]_{\mathcal{B}}$. This implies that $\lambda_i \mu_j = \mu_i \lambda_j$ for all $i, j$. Hence, $\lambda_i = \mu_i$ for all $i$, so $S$ and $T$ are simultaneously diagonalizable.

Conversely, suppose $S$ and $T$ commute. Then $[S]_{\mathcal{B}} [T]_{\mathcal{B}} = [T]_{\mathcal{B}} [S]_{\mathcal{B}}$. This implies that $\lambda_i \mu_j = \mu_i \lambda_j$ for all $i, j$. Hence, $\lambda_i = \mu_i$ for all $i$, so $S$ and $T$ are simultaneously diagonalizable.
4.5. DIAGONALIZABLE OPERATORS

CHAPTER 4.

\[ T(v_i) = \mu_i v_i. \] To show \( S \) and \( T \) commute, it is enough to check they commute on a basis. We have

\[
ST(v_i) = S(\mu_i v_i) \\
= \mu_i S(v_i) \\
= \mu_i \lambda_i v_i \\
= \lambda_i \mu_i v_i \\
= \lambda_i T(v_i) \\
= TS(v_i).
\]

Thus, \( S \) and \( T \) commute.

Now suppose that \( S \) and \( T \) commute. Since \( T \) is diagonalizable, we can write \( V = V_1 \oplus \cdots \oplus V_k \) with \( V_i \) the \( \mu_i \)-eigenspace of \( T \). For \( v_i \in V_i \) we have

\[
TS(v_i) = ST(v_i) \\
= \mu_i S(v_i).
\]

Thus, \( S(v_i) \in V_i \) and so the \( V_i \) are \( S \)-invariant. We claim each \( V_i \) has a basis consisting of eigenvectors of \( S \), i.e., \( S_i = S|_{V_i} \) is diagonalizable. We know from the previous section that a linear transformation is diagonalizable if and only if the minimal polynomial splits into distinct linear factors. Thus, \( m_S(x) \) splits into distinct linear factors by assumption. However, \( m_{S_i}(x) \mid m_S(x) \), so \( S_i \) is diagonalizable as well. Let \( B_i \) be a basis of \( V_i \) consisting of eigenvectors of \( S \). The elements of \( B_i \) are in \( V_i \), so are eigenvectors of \( T \) automatically. Set \( B = B_1 \cup \cdots \cup B_k \) and we are done.

**Definition 4.5.16.** Let \( T \in \text{Hom}_F(V,V) \). We say \( T \) is nilpotent if there is a positive integer \( r \) so that \( T^r \) is 0.

Let \( A \in \text{Mat}_n(F) \) be a matrix and assume that \( c_A(x) \) splits over \( F \) so that \( A \) has a Jordan canonical form. Note that if \( J \) is the Jordan canonical form of \( A \), we can write \( J = S + N \) where \( S \) is a semi-simple matrix and \( N \) is a nilpotent matrix. In particular, the matrix \( S \) is the diagonal matrix and \( N \) is the super-diagonal matrix. Moreover, \( S \) and \( N \) are unique and they satisfy \( SN = NS \). This can be show in much greater generality. First, we need the following result on

**Theorem 4.5.17.** Let \( F \) be a subfield of \( \mathbb{C} \) and let \( T \in \text{Hom}_F(V,V) \). There is a unique semi-simple map \( S \in \text{Hom}_F(V,V) \) and a unique nilpotent map \( N \in \text{Hom}_F(V,V) \) so that

1. \( T = S + N; \)
2. \( SN = NS. \)

Furthermore, \( S \) and \( N \) are given by polynomials in \( T \).

87
Proof. Write $m_T(x) = p_1(x)^{e_1} \cdots p_k(x)^{e_k}$ be the factorization of $m_T(x)$ into irreducible components. Let $r = \max\{e_1, \ldots, e_k\}$. Set $f(x) = p_1(x) \cdots p_k(x)$. We have $f(x)$ is a product of distinct irreducibles and $f(x)^r$ is divisible by $m_T(x)$.

We construct a sequence of polynomials $g_0(x), g_1(x), \ldots$ so that $f \left( x - \sum_{j=0}^{m} g_j(x) f(x)^j \right)$ is divisible by $f(x)^{m+1}$ for $m = 0, 1, 2, \ldots$. To begin, set $g_0(x) = 0$. Then clearly we have $f(x)$ divides $f(x)$, so the result is true in this case. Suppose we have constructed $g_0(x), \ldots, g_{m-1}(x)$ with the required properties. Set

$$h(x) = x - \sum_{j=0}^{m-1} g_j(x) f(x)^j.$$ 

Then we have $f(x)^{m-1} | f(h(x))$. We apply Taylor’s formula to $f(h(x) - g_m(x) f(x)^m)$ to obtain

$$f(h(x) - g_m(x) f(x)^m) = f(h(x)) - g_m(x) f(x)^m f'(h(x)) + f(x)^{m+1} b(x)$$

where $b(x) \in F[x]$ is some polynomial. Our assumption gives that $f(h(x)) = q(x) f(x)^m$ for some $q(x) \in F[x]$. Thus, if we can choose $g_m(x)$ so that $q(x) - g_m(x) f'(h(x))$ is divisible by $f(x)$. This can be done because $\gcd(f(x), f'(x)) = 1$.

Now take $m = r - 1$. Since $f(T)^r = 0$ we have

$$f \left( T - \sum_{j=0}^{r-1} g_j(T) f(T)^j \right) = 0.$$ 

Set

$$N = \sum_{j=1}^{r-1} g_j(T) f(T)^j = \sum_{j=0}^{r-1} g_j(T) f(T)^j$$

where the second equality is because we chose $g_0(x) = 0$. We use the fact that $\sum_{j=1}^{m} g_j(x) f(x)^j$ is divisible by $f$ to see that $N^r = 0$ and so $N$ is nilpotent. Set $S = T - N$. Clearly we have $T = S + N$. To see that $S$ is semi-simple, just note that $f(S) = f(T - N) = 0$ by the construction of $N$ and $f$ has distinct irreducible factors by definition. It is also clear $S$ and $N$ are given by polynomials in $T$ since that is how they were constructed.

It only remains to prove that $S$ and $N$ are unique. To prove uniqueness it is enough to work over an algebraically closed field. Suppose there exists $N'$ and $S'$ that satisfy $S'$ being semi-simple, $N'$ nilpotent, $T = S' + N'$ and $S'N' = N'S'$. We now show that $S' = S$ and $N' = N$. We immediately see since $S'$ and $N'$ commute with each other and $T = S' + N'$, $S'$ and $N'$ commute with $T$. Since $S$ and $N$ are given by polynomials in $T$, we have $S'$ and $N'$ commute with $S$ and $N$ as well. Observe we have $S + N = S' + N'$, i.e., $S - S' = N' - N$ and all four of these operators commute with each other. Since we are working over an algebraically closed field, semi-simple is the same as diagonalizable so $S$ and
4.5. DIAGONALIZABLE OPERATORS

$S'$ are diagonalizable. Moreover, since they commute they are simultaneously diagonalizable. This gives $S - S'$ is diagonalizable, which in turns gives $N' - N$ is diagonalizable. Moreover, since $N$ and $N'$ are both nilpotent and commute with each other, $N' - N$ is nilpotent. In particular, we have

$$(N' - N)^m = \sum_{j=0}^{m} \binom{m}{j} (N')^{m-j} (-N)^j.$$  

If $\dim_F V = n$, if we take $m = 2n$ that is large enough so that $(N' - N)^m = 0$. (In fact $n$ is large enough, but that isn’t important for this result.) So we have $S - S'$ is diagonalizable and nilpotent. The minimal polynomial for $S - S'$ must be $x^r$ for some $r \leq m$ because the operator is nilpotent, but since $S - S'$ is semi-simple its minimal polynomial must split into distinct irreducible factors. Thus, $r = 1$ and so $S - S' = 0$, i.e., $S = S'$. Since $S = S'$, we immediately get $N = N'$ and we have the result.

We close this section with a few further results on the Jordan and rational canonical structure of products of matrices.

Lemma 4.5.18. Let $S, T \in \text{Hom}_F(V, V)$ with $V$ a finite dimensional $F$-vector space. Let $p(x) = a_0 x^r + \cdots + a_1 x + a_0 \in F[x]$ with $a_0 \neq 0$. Then $\dim_F(\ker p(ST)) = \dim_F(\ker p(TS))$.

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis for $\ker(p(ST))$. We claim $\{T(v_1), \ldots, T(v_k)\}$ is linearly independent. Suppose

$$c_1 T(v_1) + \cdots + c_k T(v_k) = 0.$$  

This gives $T(c_1 v_1 + \cdots + c_k v_k) = 0$, so $ST(c_1 v_1 + \cdots + c_k v_k) = 0$. Let $v = c_1 v_1 + \cdots + c_k v_k$. Then $ST(v) = 0$. Moreover, we also have $v \in \ker(p(ST))$ because $\{v_1, \ldots, v_k\}$ is a basis for $\ker(p(ST))$. Thus

$$0 = p(ST)v = a_r (ST)^r(v) + \cdots + a_1 ST(v) + a_0 v = a_0 v$$

where we have used $v \in \ker(ST)$. However, since $a_0 \neq 0$, this gives $v = 0$. Thus $0 = c_1 v_1 + \cdots + c_k v_k$. But $\{v_1, \ldots, v_k\}$ is linearly independent so we must have $c_i = 0$ for all $i$ as desired. Thus $\{T(v_1), \ldots, T(v_k)\}$ is linearly independent.

We now claim $T(v_i) \in \ker(p(TS))$ for each $i$. We have

$$(TS)^j T = \underbrace{(TS) \cdots (TS)}_{j \text{ times}} T = T \underbrace{(ST) \cdots (ST)}_{j \text{ times}} = T(ST)^j.$$  

89
Thus,
\[
p(TS)T(v_i) = (a_r(TS)^r + \cdots + a_0)T(v_i)
= T((a_r(ST)^r + \cdots + a_0)(v_i))
= T(p(ST)(v_i))
= T(0)
= 0.
\]

Thus, \(\{T(v_1), \ldots, T(v_k)\}\) is a linearly independent set in \(\ker p(TS)\), so \(\dim_F \ker(p(TS)) \geq k = \dim_F \ker(p(ST))\). One now uses the same argument with \(S\) replacing \(T\) to get the other direction of the inequality.

**Theorem 4.5.19.** Let \(T \in \text{Hom}_F(V,V)\) and \(S \in \text{Hom}_F(V,V)\) with \(F\) algebraically closed. Then \(ST\) and \(TS\) have the same nonzero eigenvalues, and for a common eigenvalue \(\lambda\), they have the same Jordan block structure at \(\lambda\).

**Proof.** Let \(\lambda \neq 0\) be a nonzero eigenvalue of \(ST\) and \(v \in V\) a nonzero eigenvector. Set \(w = T(v)\). Then \(ST(v) = \lambda v = S(w)\). We also have
\[
TS(w) = T(\lambda v)
= \lambda T(v)
= \lambda w.
\]
We have \(T(v)\) is nonzero since if it were 0 we would have \(S(0) = \lambda v\) implies \(\lambda = 0\) or \(v = 0\), a contradiction. Thus \(w \neq 0\) and so \(\lambda\) is an eigenvalue of \(TS\) as well. The same argument in the other direction shows every nonzero eigenvalue of \(TS\) is a nonzero eigenvalue of \(ST\). It remains to deal with the Jordan block structure.

Consider the polynomials \(p_{j,\lambda}(x) = (x - \lambda)^j\) for \(j \geq 1\). The Jordan block structure of \(ST\) at \(\lambda\) is given by the dimensions of \(\ker(p_{j,\lambda}(ST))\). However, Lemma 4.5.18 gives \(\dim \ker(p_{j,\lambda}(ST)) = \dim \ker(p_{j,\lambda}(TS))\) for each \(j\) as long as the constant term of \(p_{j,\lambda}(x) \neq 0\). However, as long as \(\lambda \neq 0\) this is satisfied so the Jordan block structures must be equal.

Over a general field we have the following result.

**Corollary 4.5.20.** Let \(S, T \in \text{Hom}_F(V,V)\). Then \(c_{TS}(x) = c_{ST}(x)\).

**Proof.** We claim it is enough to prove the result for \(F\) algebraically closed. This follows because the characteristic polynomial is the same regardless of which field one consider \(S\) and \(T\) to be defined over.

Let \(\dim_F V = n\) and let \(\lambda_1, \ldots, \lambda_k\) be the distinct nonzero eigenvalues of \(ST\). Write
\[
c_{ST}(x) = x^{e_0}(x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}
\]
with \(e_0 = n - (e_1 + \cdots + e_k)\). The previous theorem gives that \(ST\) and \(TS\) have the same Jordan block structure at \(\lambda_1, \ldots, \lambda_k\). Thus, \(c_{TS}(x) = x^{f_0}(x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}\) where \(f_0 = n - (e_1 + \cdots + e_k) = e_0\). Thus \(c_{TS}(x) = c_{ST}(x)\). \(\square\)
The previous results do not give that the matrices associated to $ST$ and $TS$ are similar. Consider the following example.

**Example 4.5.21.** Let $A = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Then $AB = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ has characteristic polynomial $c_{AB}(x) = x^2$ and $BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ has characteristic polynomial $c_{BA}(x) = x^2$. Note the rational canonical form of $BA$ is $BA$, but the rational canonical form of $AB$ is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Thus, they cannot be similar.

**Theorem 4.5.22.** Let $T \in \text{Hom}_F(V, V)$. Then the following are equivalent:

1. $V$ is $T$-generated by a single element, i.e., the rational canonical form of $T$ is a single block;
2. every linear transformation $S \in \text{Hom}_F(V, V)$ that commutes with $T$ is a polynomial in $T$.

**Proof.** Suppose that $V$ is $T$-generated by a single element $v_0$. Let $S \in \text{Hom}_F(V, V)$ commute with $T$. There exists $p_0(x) \in F[x]$ so that $S(v_0) = p_0(T)(v_0)$ since $v_0$ is a $T$-generator of $V$. Let $v \in V$ and write $v = g(T)(v_0)$ for some $g(x) \in F[x]$. We have

$$S(v) = S(g(T)(v_0)) = g(T)S(v_0) = g(T)p_0(T)(v_0) = p_0(T)g(T)(v_0) = p_0(T)(v).$$

Thus, $S = p_0(T)$ since they agree on every element of $V$.

Now suppose that $V$ is not $T$-generated by a single element. Let $\{v_1, \ldots, v_k\}$ be a rational canonical generating set for $T$. This gives $V = V_1 \oplus \cdots \oplus V_k$ with $V_i$ $T$-generated by $v_i$. In particular the $V_i$ are $T$-invariant. Define $S : V \to V$ by

$$S(v) = \begin{cases} 0 & \text{if } v \in V_1 \\ v & \text{if } v \in V_i \text{ for } i > 0. \end{cases}$$

Since the direct sum is $T$-invariant we get that $S$ and $T$ commute. Suppose $S = p(T)$ for some $p(x) \in F[x]$. We have $0 = S(v_1) = p(T)(v_1)$. Thus, $p_1(x) \mid p(x)$ where $p_1(x) = m_{T,v_1}(x)$. However, $p_j(x) \mid p_1(x)$ for all $j \geq 1$ and so $p_j(x) \mid p(x)$ for all $j$. Thus, $S(v_j) = p(T)(v_j) = 0$ because $p_j(x) \mid p(x)$. This is a contradiction if $k > 1$.

### 4.6 Canonical forms via modules

This section is not part of the main course and is simply included to provide some clarification to those that have had some exposure to modules and would
like to fit the contents of this chapter into what they already know. We will assume basic familiarity with modules; for those that lack this but would like to read this section, please see [1].

Let $R$ be a PID and $M$ an $R$-module. Let $X$ be a finite subset of $M$. We define

$$XR = \{x_1r_1 + \cdots + x_nr_n : r_i \in R, x_i \in X, m \in \mathbb{N}\}.$$ 

We say that $M$ is a \textit{finitely generated} $R$-module if $M = XR$ for some finite subset $X \subset M$. Recall that we say $M$ is a \textit{cyclic $R$-module} if there exists $m \in M$ so that $M = mR$. If we define a map $\phi : R \to M$ by $r \mapsto rm$, then we obtain

$$M \cong R/\ker \phi.$$ 

Since we are assuming $R$ is a PID, there exists $a \in R$ so that $\ker \phi = aR = \{ar : r \in R\}$. Thus, $M \cong R/aR$.

The following example is the key example for relating modules to what we’ve done in this chapter.

\textbf{Example 4.6.1.} Let $F$ be a field and $V$ a finite dimensional $F$-vector space. Let $T \in \text{Hom}_F(V,V)$. We have that $V$ is an $F[x]$-module via $f(x) \cdot v = f(T)(v)$. Note that here we have $V$ is a cyclic $F[x]$-module if and only if $V$ is $T$-generated by a single element. Now consider what it means for a subspace $W \subset V$. Note that here we have $W$ to be $T$-invariant. This means that $T(W) \subset W$ and $W$ is a subspace of $V$. However, one can easily see this is equivalent to requiring that $W$ be an $F[x]$-submodule of $V$. Thus, the entire theory of $T$-invariant subspaces amounts to studying submodules of $V$ when considered as an $F[x]$-module with $x$ acting via $T$.

We can now state the first form of the Fundamental Theorem of Finitely Generated Modules over a PID.

\textbf{Theorem 4.6.2.} Let $R$ be a PID and $M$ a finitely generated $R$-module. Then

$$M \cong R^r \oplus R/a_1R \oplus R/a_2R \oplus \cdots \oplus R/a_mR$$

for some integer $r \geq 0$ and some nonzero elements $a_1, \ldots, a_m \in R$ which are not units in $R$ and satisfy the divisibility relations $a_m \mid a_{m-1} \mid \cdots \mid a_2 \mid a_1$.

The elements $a_1, \ldots, a_m$ in the above theorem are referred to as the \textit{invariant factors} of $M$ and $r$ is referred to as the \textit{rank} of $M$. Equation 4.1 is the \textit{invariant factor decomposition} of $M$.

Proving this theorem takes us too far afield for this course, so we assume its validity and show how it implies every matrix has a rational canonical form. Let $T \in \text{Hom}_F(V,V)$ and view $V$ as an $F[x]$-module as in the example about. The theorem gives $p_1(x), \ldots, p_m(x)$ in $F[x]$ with $p_m(x) \mid p_{m-1}(x) \mid \cdots \mid p_1(x)$ so that

$$V \cong F[x]/(p_1(x)) \oplus \cdots \oplus F[x]/(p_m(x)).$$

Now to see that this gives rational canonical form we just need to choose an appropriate basis for each space $F[x]/(p_j(x))$. Write $p_j(x) = x^{n_j} + a_{n_j-1}x^{n_j-1} + \cdots + a_0$.
4.7. PROBLEMS

\( \cdots + a_1 x + a_0 \). Consider the basis \( B = \{ [x]^{n_j-1}, [x]^{n_j-1}, \cdots, [x], [1] \} \). We have that \( T \) acts via multiplication by \( x \), thus \( T(x^k) = x^{k+1} \). Thus, in \( F[x]/(p_j(x)) \) we have \( T \) acts as

\[
\begin{align*}
[1] & \mapsto [x] \\
[x] & \mapsto [x]^2 \\
& \vdots \\
[x]^{n_j-1} & \mapsto [x]^{n_j} = -a_{n_j-1} [x]^{n_j-1} - \cdots - a_1 [x] - a_0.
\end{align*}
\]

This shows that the matrix of \( T \) with respect to \( B \) on \( F[x]/(p_j(x)) \) is precisely the companion matrix. Thus, we have recovered rational canonical form.

One should keep in mind that even though it was very easy to deduce rational canonical form from the structure theorem, proving the structure theorem itself requires a considerable amount of effort. The reason this method is generally preferable to the approach given earlier in this chapter is that the structure theorem for finitely generated modules has many other applications, such as classifying all finitely generated abelian groups.

We can give a different version of the structure theorem that is useful for obtaining Jordan canonical form. We assumed above that \( R \) is a PID. This implies that \( R \) is necessarily a unique factorization domain as well, i.e., a UFD. Given any \( a \in R \) there are primes \( p_1, \ldots, p_s \), a unit \( u \), and positive integers \( e_1, \ldots, e_s \) so that

\[ a = u p_1^{e_1} \cdots p_s^{e_s}. \]

Recall here that “prime” means that if \( p \mid ab \), then \( p \mid a \) or \( p \mid b \). The factorization of \( a \) is unique up to units. We can apply this to the structure theorem by decomposing each \( a_i \) into its prime factorization, i.e., we can write

\[ R/aR \cong R/p_1^{e_1} R \oplus \cdots \oplus R/p_s^{e_s} R. \]

We now restate the theorem in this form.

**Theorem 4.6.3.** Let \( R \) be a PID and \( M \) a finitely generated \( R \)-module. Then we have

\[ M \cong R^r \oplus R/p_1^{e_1} R \oplus \cdots \oplus R/p_s^{e_s} R \]

where \( r \geq 0 \) is an integer and the \( p_i^{e_i} \) are positive powers of primes in \( R \). (Note we do not assume the \( p_i \) are distinct here!)

The \( p_i^{e_i} \) are referred to as the *elementary divisors* of \( M \).

To recover Jordan Canonical Form, one applies this version of the structure theorem with the terms \( F[x]/(x - \lambda)^k \).

4.7 Problems

For all of these problems \( V \) is a finite dimensional \( F \)-vector space.
1. Let $T \in \text{Hom}_F(V,V)$. Prove that the intersection of any collection of $T$-invariant subspaces of $V$ is $T$-invariant.

2. Let $T \in \text{Hom}_F(V,V)$ and $w \in V$. Let $m_{T,w}(x) \in F[x]$ be the annihilating polynomial of $w$.
   (a) Show that if $m_{T,w}(x) = p(x)q(x)$, then $p(x) = m_{T,q(T)(w)}(x)$.
   (b) Let $W$ be the subspace of $V$ that is $T$-generated by $w$. If $\deg m_{T,w}(x) = d$ and $\deg q(x) = e$, show that $\dim_W q(T)(W) = d - e$.

3. Let $A \in \text{Mat}_4(\mathbb{Q})$ be defined by
   \[
   \begin{pmatrix}
   1 & 2 & -4 & 4 \\
   2 & -1 & 4 & -8 \\
   1 & 0 & 1 & -2 \\
   0 & 1 & -2 & 3
   \end{pmatrix}.
   \]
   First find the rational canonical form of $A$ by hand. Check your answer using SAGE.

4. Prove that two $3 \times 3$ matrices are similar if and only if they have the same characteristic and minimal polynomials. Give an explicit counterexample to this assertion for $4 \times 4$ matrices.

5. We say $A \in \text{Mat}_2(F)$ has multiplicative order $n$ if $A^n = I$ and $A^m \neq I$ for any $0 < m < n$. Show that $x^5 - 1 = (x - 1)(x^2 - 4)(x^2 + 5)$ in $\mathbb{F}_{19}[x]$. Use this to determine all similarity all elements of $\text{Mat}_2(\mathbb{F}_{19})$ of multiplicative order 5.

6. In a group $G$ we say two elements $a$ and $b$ are conjugate if there exists $g \in G$ so that $a = gbg^{-1}$. The conjugacy class of an element is the collection of all elements conjugate to it. Given a conjugacy class $C$, any element in $C$ is referred to as a representative for $C$. Determine representatives for all the conjugacy classes for $\text{GL}_3(\mathbb{F}_2)$.

7. Prove that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of a matrix $A \in \text{Mat}_n(F)$, then $\lambda_1^k, \ldots, \lambda_n^k$ are the eigenvalues of $A^k$ for any $k \geq 0$.

8. Let $c_T(x) = (x - \lambda)^e p(x)$ with $p(\lambda) \neq 0$. Show that $\dim_F E^\infty_{\lambda} = e$.

9. Prove that the matrices
   \[
   \begin{pmatrix}
   2 & 0 & 0 & 0 \\
   -4 & -1 & -4 & 0 \\
   2 & 1 & 3 & 0 \\
   -2 & 4 & 9 & 1
   \end{pmatrix}
   \quad \text{and} \quad
   \begin{pmatrix}
   5 & 0 & -4 & -7 \\
   3 & -8 & 15 & -13 \\
   2 & -4 & 7 & -7 \\
   1 & 2 & -5 & 1
   \end{pmatrix}
   \]
   are similar.
10. Prove that the matrices
\[
\begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 \\
1 & 1 & 1 & 0
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
5 & 2 & -8 & -8 \\
-6 & -3 & 8 & 8 \\
-3 & -1 & 3 & 4 \\
3 & 1 & -4 & -5
\end{pmatrix}
\]
both have characteristic polynomial \((x - 3)(x + 1)^3\). Determine the Jordan canonical form for each matrix and determine if they are similar.

11. Determine all possible Jordan canonical forms for a linear transformation with characteristic polynomial \((x - 2)^3(x - 3)^2\).

12. Prove that any matrix \(A \in \text{Mat}_n(\mathbb{C})\) satisfying \(A^3 = A\) can be diagonalized. Is the same statement true over any field \(F\)? If so, prove it. If not, give a counterexample.

13. Determine the Jordan canonical form for a matrix \(A \in \text{Mat}_n(\mathbb{Q})\) with entries all equal to 1.
Chapter 5

Bilinear and sesquilinear forms

In this chapter we study bilinear and sesquilinear forms. In the special cases of symmetric, skew-symmetric, Hermitian, and skew-Hermitian forms we provide the standard classification theorems.

5.1 Basic definitions and facts

In this chapter we will be interested in maps $V \times V \to F$ that are linear in each variable separately. In particular, we will be interested in bilinear forms.

**Definition 5.1.1.** Let $V$ be an $F$-vector space. A function $\phi: V \times V \to F$ is said to be a bilinear form if $\phi$ is an $F$-linear map in each variable separately, i.e., for all $v_1, v_2, v \in V$ and $c \in F$ we have

1. $\phi(cv_1 + v_2, v) = c\phi(v_1, v) + \phi(v_2, v)$;
2. $\phi(v, cv_1 + v_2) = c\phi(v, v_1) + \phi(v, v_2)$.

We denote the collection of bilinear forms by $\text{Hom}_F(V, V; F)$.

**Exercise 5.1.2.** Show that $\text{Hom}_F(V, V; F)$ is an $F$-vector space.

We now give some elementary examples of bilinear forms. Checking each of these satisfy the criterion to be bilinear form is left as an exercise.

**Example 5.1.3.** The first example, and the one that should be kept in mind throughout this and the next chapter, is that familiar example of a dot product from multivariable calculus. Let $V = \mathbb{R}^n$ for some $n \in \mathbb{Z}_{\geq 1}$. Define $\varphi: V \times V \to$
5.1. BASIC DEFINITIONS AND FACTS

CHAPTER 5.

\[ \varphi(v, w) = v \cdot w \]
\[ = \sum_{i=1}^{n} a_i b_i \]

where \( v = (a_1, \ldots, a_n) \) and \( w = (b_1, \ldots, b_n) \).

**Example 5.1.4.** Let \( A \in \text{Mat}_n(F) \). We have a bilinear form \( \varphi_A \) on \( V = F^n \) defined by
\[ \varphi_A(v, w) = v^t A w. \]

Much as one saw that upon choosing a basis any linear map between finite dimensional vector spaces could be realized as a matrix, we will soon see that any bilinear form on a finite dimensional vector space can be represented as a matrix as in this example upon choosing a basis for \( V \).

**Example 5.1.5.** Let \( V = C^0([0,1], \mathbb{R}) \) be the \( \mathbb{R} \)-vector space of continuous functions from \([0,1]\) to \( \mathbb{R} \). Define \( \varphi : V \times V \to \mathbb{R} \) by setting
\[ \varphi(f, g) = \int_0^1 f(x)g(x)dx. \]

This gives a bilinear form. More generally, given any positive integer \( n \), one can consider the vector space of paths \( V = C^0([0,1], \mathbb{R}^n) \). Given \( f, g \in V \), we have \( f(x), g(x) \in \mathbb{R}^n \) for each \( x \in [0,1] \). Thus, for each \( x \) we have \( f(x) \cdot g(x) \) is well-defined from Example 5.1.3 above. We can define
\[ \varphi(f, g) = \int_0^1 f(x) \cdot g(x)dx. \]

This defines a bilinear form on \( V \).

In the next section we will be interested in classifying certain “nice” bilinear forms. In particular, we will often restrict to the case our bilinear form is non-degenerate.

**Definition 5.1.6.** Let \( \varphi \in \text{Hom}_F(V, V; F) \) be a bilinear form. We say that \( \varphi \) is right non-degenerate if given \( w_0 \in V \) so that \( \varphi(v, w_0) = 0 \) for every \( v \in V \) one has \( w_0 = 0 \). We say it is left non-degenerate if given any \( v_0 \in V \) so that \( \varphi(v_0, w) = 0 \) for every \( w \in V \) one has \( v_0 = 0 \). We say \( \varphi \) is non-degenerate if it is left and right non-degenerate.

We will see below that if \( V \) is a finite dimensional vector space that \( \varphi \) is left non-degenerate if and only if it is right non-degenerate. However, if \( V \) is infinite dimensional these may not be the same as the following example shows.
Example 5.1.7. Consider the $\mathbb{R}$-vector space
\[ \ell^2 = \left\{ (a_n)_{n \geq 1} : \sum_{n=1}^{\infty} |a_n|^2 < \infty \right\}. \]

Define a bilinear form $\varphi$ on $\ell^2$ by setting
\[ \varphi(a, b) = \sum_{n=1}^{\infty} a_{n+1} b_n \]
for $a = (a_n)$ and $b = (b_n)$. Set $x = (1, 0, 0, \ldots)$. Then we have $\varphi(x, b) = 0$ for all $b$, so $\varphi$ is not left non-degenerate, i.e., it is left degenerate. However, if $\varphi(a, b_0) = 0$ for all $a = (a_n)$, it is easy to check this implies $b_0$ is the sequence consisting of all 0’s. Thus, $\varphi$ is right non-degenerate.

Non-degenerate forms arise in a very natural way in a context we have already studied, namely, in terms of isomorphisms between a finite dimensional vector space and its dual. In particular, recall that when studying dual spaces we saw that for a finite dimensional $F$-vector space $V$ that one has $V \cong V^\vee$, but that this isomorphism depends upon picking a basis and so is non-canonical. It turns out that non-degenerate bilinear forms are in bijection with the collection of such isomorphisms for finite dimensional vector spaces. For infinite dimensional spaces one only obtains $V$ injects into $V^\vee$.

Let $\varphi \in \text{Hom}_F(V, V; F)$. If we fix $v_0 \in V$ then we have a linear map $\varphi(\cdot, v_0) \in V^\vee$ given by $w \mapsto \varphi(w, v_0)$. In particular, we define $R_\varphi : V \to V^\vee$ by setting
\[ R_\varphi(v) = \varphi(\cdot, v). \]
Note that for any $v \in V$, the map $R_\varphi(v) \in V^\vee$ is given by $R_\varphi(v)(w) = \varphi(w, v)$.

We claim that $R_\varphi$ is a linear map. We want to show that for any $a \in F$ and $v_1, v_2 \in V$ that $R_\varphi(av_1 + v_2) = aR_\varphi(v_1) + R_\varphi(v_2)$. As this is an equality of maps, we must show these maps agree on elements. We have
\[ R_\varphi(av_1 + v_2)(w) = \varphi(w, av_1 + v_2) \]
\[ = a\varphi(w, v_1) + \varphi(w, v_2) \]
\[ = aR_\varphi(v_1)(w) + R_\varphi(v_2)(w), \]
as desired. Thus, $R_\varphi \in \text{Hom}_F(V, V^\vee)$.

Of course, one could just as easily have fixed $w_0 \in V$ and considered the linear map $\varphi(w_0, \cdot) \in V^\vee$ given by $v \mapsto \varphi(w_0, v)$. In this case we define $L_\varphi : V \to V^\vee$ by setting
\[ L_\varphi(w) = \varphi(w, \cdot). \]
Just as above one obtains $L_\varphi \in \text{Hom}_F(V, V^\vee)$.

Lemma 5.1.8. A bilinear form $\varphi$ is non-degenerate if and only if $L_\varphi$ and $R_\varphi$ are injections.
5.1. BASIC DEFINITIONS AND FACTS

Proof. First, suppose that $L_\varphi$ and $R_\varphi$ are injective. Let $w \in V$. If $\varphi(v, w) = 0$ for all $v \in V$, this implies that $R_\varphi(w)(v) = 0$ for every $v \in V$. However, this says that $R_\varphi(w)$ is the zero map. Since we are assuming $R_\varphi$ is injective, this gives $w = 0$. Thus, $\varphi$ is right non-degenerate. Similarly, $L_\varphi$ injective implies that $\varphi$ is left non-degenerate.

Now suppose that $L_\varphi$ or $R_\varphi$ is not injective. As the argument is essentially the same, we suppose there exists $w \in V$ with $w \neq 0$ so that $R_\varphi(w) = 0$, i.e., $R_\varphi$ is not an injection. This translates to the statement that $0 = R_\varphi(w)(v) = \varphi(v, w)$ for all $v \in V$. This gives that $\varphi$ is not right non-degenerate, i.e., $\varphi$ is degenerate.

One can note in the previous result that if $V$ is finite dimensional then $R_\varphi$ being an injection is the same as $R_\varphi$ being an isomorphism since $V$ and $V^\vee$ have the same dimension. So in the finite dimensional case one can replace the condition $L_\varphi$ and $R_\varphi$ are injective with them being isomorphisms.

One can define the left and right kernel of $\varphi$ by setting

$$\ker R_\varphi = \ker R_\varphi = \{ w \in V : \varphi(v, w) = 0, \text{ for every } v \in V \}$$

and

$$\ker L_\varphi = \ker L_\varphi = \{ v \in V : \varphi(v, w) = 0, \text{ for every } w \in V \}.$$ 

This allows us to rephrase the condition that $\varphi$ is non-degenerate to be that the left and right kernels of $\varphi$ are trivial.

In the case that $V$ is a finite dimensional vector space one only needs to consider left or right non-degenerate as the other comes for free. This is due to the following result.

Theorem 5.1.9. Let $V$ be a finite dimensional vector space over a field $F$. Let $\varphi \in \text{Hom}_F(V, V; F)$. Then $L_\varphi$ and $R_\varphi$ are dual to each other, namely, given $L_\varphi : V \to V^\vee$, if we dualize this to obtain $L_\varphi^\vee : (V^\vee)^\vee \to V^\vee$, then upon identifying $(V^\vee)^\vee$ with $V$ via the canonical isomorphism we have $L_\varphi^\vee = R_\varphi$. Similarly one has $R_\varphi^\vee = L_\varphi$.

Proof. Recall that given $F$-vector spaces $V, W$ and $T \in \text{Hom}_F(V, W)$, the dual map $T^\vee$ is defined by

$$T^\vee(\psi)(v) = \psi(T(v))$$

for $v \in V$ and $\psi \in V^\vee$. We also recall the canonical isomorphism between $V$ and $(V^\vee)^\vee$ is given by sending $v$ to the evaluation map $\text{eval}_v$. For $v, w \in V$ we have

$$L_\varphi^\vee(\text{eval}_v)(w) = \text{eval}_v(L_\varphi(w)) = \text{eval}_v(\varphi(w, \cdot)) = \varphi(w, v) = R_\varphi(v)(w).$$

Thus, we have $L_\varphi^\vee = R_\varphi$ upon identifying $V$ and $(V^\vee)^\vee$. 

99
This result shows that if \( V \) is finite dimensional then \( R_\varphi \) is injective if and only if \( L_\varphi \) is injective. Equivalently, \( \varphi \) is non-degenerate if and only if \( R_\varphi \) is an isomorphism.

**Lemma 5.1.10.** Let \( V \) be a finite dimensional vector space. There is a bijection between isomorphisms \( T : V \to V^\vee \) and non-degenerate bilinear forms \( \varphi \in \text{Hom}_F(V,V;F) \).

**Proof.** Let \( \varphi \) be a non-degenerate bilinear form. Then we associate the required isomorphism by sending \( \varphi \) to \( R_\varphi \).

Now suppose we have an isomorphism \( T : V \to V^\vee \). Define \( \varphi \in \text{Hom}_F(V,V;F) \) by
\[
\varphi(v,w) = T(w)(v).
\]
It is elementary to check this is a bilinear form and \( R_\varphi = T \), so is non-degenerate. Moreover, since \( R_\varphi = T \) it is immediate that these maps are inverse to each other and so provide the required bijection.

It is well known from multivariable calculus that Example 5.1.3 is used to define the length of a vector in \( \mathbb{R}^n \). However, if we make the same definition on \( \mathbb{C} \) we do not recover the length of a complex number. In order to find the length of \( z \in \mathbb{C} \), we want to consider \( z \overline{z} \). This leads to the definition of a different type of form, a sesquilinear form, that keeps track of conjugation as well. Before defining these forms, we need to define conjugation for more general fields than \( \mathbb{C} \).

**Definition 5.1.11.** Let \( F \) be a field and \( \text{conj} : F \to F \) a map that satisfies:

1. \( \text{conj}(\text{conj}(x)) = x \) for every \( x \in F \);
2. \( \text{conj}(x + y) = \text{conj}(x) + \text{conj}(y) \) for every \( x, y \in F \);
3. \( \text{conj}(xy) = \text{conj}(x) \text{conj}(y) \) for every \( x, y \in F \).

We call such a map a *conjugation map* on \( F \). We say \( \text{conj} \) is nontrivial if \( \text{conj} \) is not the identity map on \( F \).

We give a couple of familiar examples of conjugation maps.

**Example 5.1.12.** Let \( F = \mathbb{C} \). Then \( \text{conj} \) is the familiar conjugation map sending \( x + iy \) to \( x - iy \) for \( x, y \in \mathbb{R} \).

**Example 5.1.13.** Let \( D \in \mathbb{Z} \) and consider \( F = \mathbb{Q}(\sqrt{D}) \) where we recall
\[
\mathbb{Q}(\sqrt{D}) = \{ a + b\sqrt{D} : a, b \in \mathbb{Q} \}.
\]
It is easy to check the map sending \( a+b\sqrt{D} \) to \( a-b\sqrt{D} \) is a nontrivial conjugation map on \( F \) if \( D \) is not a perfect square.

To emphasize that one should think of conjugation as a generalization of complex conjugation as well as to save writing we will denote conjugation maps by \( x \mapsto \overline{x} \).
Lemma 5.1.14. Let $F$ be a field with nontrivial conjugation and assume $\text{char}(F) \neq 2$. Then:

1. Let $F_0 = \{ z \in F : \bar{z} = z \}$. Then $F_0$ is a subfield of $F$.

2. There is a nonzero element $j \in F$ so that $\bar{j} = -j$.

3. Every element of $F$ can be written uniquely as $z = x + jy$ for some $x, y \in F_0$.

Proof. The fact that $F_0$ is a subfield of $F$ is left as an exercise. Let $a \in F$ so that $a \neq \bar{a}$. Since the conjugation is nontrivial there is always such an $a$. Set $j = \frac{a + \bar{a}}{2}$. Then one has $\bar{j} = -j$. Given $z \in F$, we have $\frac{z + \bar{z}}{2}$ and $\frac{z - \bar{z}}{2j}$ are both elements of $F_0$ and $z = \left( \frac{z + \bar{z}}{2} \right) + j \left( \frac{z - \bar{z}}{2j} \right)$.

Example 5.1.15. Returning to the above examples, in the case $F = \mathbb{C}$ we have $F_0 = \mathbb{R}$ and $j = \sqrt{-1}$. In the case $F = \mathbb{Q}(\sqrt{D})$ we have $F_0 = \mathbb{Q}$ and $j = \sqrt{D}$.

Given a field $F$ that admits a conjugation map and a vector space $V$ over $F$, we can also consider a conjugation map on $V$.

Definition 5.1.16. Let $V$ be an $F$-vector space where $F$ is a field with conjugation. A conjugation map on $V$ is a map $\text{conj} : V \to V$ that satisfies

1. $\text{conj}(\text{conj}(v)) = v$ for every $v \in V$;

2. $\text{conj}(v + w) = \text{conj}(v) + \text{conj}(w)$ for every $v, w \in V$;

3. $\text{conj}(av) = \pi \text{conj}(v)$ for every $a \in F$, $v \in V$.

As in the case of conjugation on a field, we will often write $v \mapsto \bar{v}$ to denote a conjugation map.

Given a finite dimensional $F$-vector space, one always has a conjugation map on $V$ if $F$ has a conjugation map. Namely, upon choosing a basis one has $V \cong F^n$ for $n = \dim_F V$. Given $v \in V$, we have an element $\left( \begin{array}{c} a_1 \\ \vdots \\ a_n \end{array} \right) \in F^n$.

Since $F$ has a nontrivial conjugation, we have $\left( \begin{array}{c} \bar{a}_1 \\ \vdots \\ \bar{a}_n \end{array} \right) \in F^n$. We set $\bar{v}$ to be the element of $V$ corresponding to this element. This clearly gives a non-trivial conjugation on $V$.

The third property in the definition of conjugation for a vector space gives immediately that linear maps are not the correct maps to work with if one wants to take nontrivial conjugation into account. To remedy this, we define conjugate linear maps.
**Definition 5.1.17.** Let $F$ be a field with conjugation and let $V$ and $W$ be $F$-vector spaces. A map $T : V \to W$ is said to be **conjugate linear** if it satisfies

1. $T(v_1 + v_2) = T(v_1) + T(v_2)$ for every $v_1, v_2 \in V$;
2. $T(av) = \pi T(v)$ for every $a \in F$, $v \in V$.

We say $T$ is a **conjugate isomorphism** if it is conjugate linear and bijective.

Of course, one might not wish to introduce a completely new definition to deal with conjugation. We can get around this by defining a new vector space associated to $V$ that takes into account the conjugation. Let $\overline{V}$ be equal to $V$ as a set, and let $\overline{V}$ have the same addition as $V$. However, we define scalar multiplication on $\overline{V}$ by setting

$$F \times V \to V$$

$$(a, v) \mapsto a \cdot v := \pi v.$$

**Exercise 5.1.18.** Check that $\overline{V}$ is an $F$-vector space.

We now consider maps in $\text{Hom}_F(\overline{V}, W)$. Let $T \in \text{Hom}_F(\overline{V}, W)$. Then we have

1. $T(v_1 + v_2) = T(v_1) + T(v_2)$ for every $v_1, v_2 \in V$;
2. $T(a \cdot v) = aT(v)$ for every $a \in F$, $v \in V$, i.e., $T(\pi v) = aT(v)$ for every $a \in F$, $v \in V$.

The second equation translates to $T(sv) = \pi T(v)$, i.e., $T$ is a conjugate linear map. Thus, the set of conjugate linear maps forms an $F$-vector space, in particular, it is equal to $\text{Hom}_F(\overline{V}, W)$. In the case that $W = F$ we write $\overline{V}' = \text{Hom}_F(\overline{V}, F)$.

There is one subtlety to keep in mind here. We have a natural $F$-vector space structure on $\text{Hom}_F(\overline{V}, W)$ just as we would for any set of linear maps between two $F$-vector spaces, namely, given $a \in F$ and $T \in \text{Hom}_F(\overline{V}, W)$, we set $a \cdot T$ to be the map defined by $a \cdot T(v) = aT(v)$. The scalar multiplication on $\text{Hom}_F(\overline{V}, W)$ is really derived from the scalar multiplication in $W$. The scalar multiplication here is not given by the conjugate multiplication; if we wanted that vector space we would have to write $\text{Hom}_F(\overline{V}, \overline{W})$, or, equivalently, we would have $\text{Hom}_F(\overline{V}, W)$.

**Definition 5.1.19.** A **sesquilinear form** $\phi : V \times V \to F$ is a map that is linear in the first variable and conjugate linear in the second variable.

While bilinear forms are linear in each variable, a sesquilinear form is linear in the first variable and conjugate linear in the second variable. Note that “sesqui” means one and a half, and this is what it refers to.

**Exercise 5.1.20.** Show that a sesquilinear form $\phi : V \times V \to F$ is the same as a bilinear form from $V \times \overline{V}$ to $F$. As such, we can denote the collection of sesquilinear forms by $\text{Hom}_F(V, \overline{V}; F)$. 
The above examples of bilinear forms are easily adjusted to give sesquilinear forms.

**Example 5.1.21.** Let $V = \mathbb{C}^n$. Define $\varphi : V \times V \to \mathbb{C}$ by

$$\varphi(v, w) = iv \overline{w} = \sum_{i=1}^{n} v_i \overline{w}_i$$

where $v = (v_1, \ldots, v_n)$ and $w = (w_1, \ldots, w_n)$. Observe that $\varphi(v, v) = ||v||^2$, and so for any $v \in \mathbb{C}^n$ we have $\varphi(v, v) \in \mathbb{R}_{\geq 0}$.

**Example 5.1.22.** Let $V = F^n$ where $F$ is a field with conjugation. Let $A \in \text{Mat}_n(F)$ and define $\varphi : V \times V \to F$ by

$$\varphi(v, w) = i v \overline{A} w.$$ 

This is a sesquilinear form.

**Example 5.1.23.** Let $V = \mathcal{C}^0([0, 1], \mathbb{C})$, i.e., $V$ is the set of paths in $\mathbb{C}$. Define $\varphi : V \times V \to \mathbb{C}$ to be the function

$$\varphi(f, g) = \int_0^1 f(z) \overline{g(z)} \, dz.$$ 

One can easily check that $\varphi$ is a sesquilinear form.

We define left and right non-degenerate as well as non-degenerate for sesquilinear forms just as for bilinear forms. We would like to give a classification for this in terms of maps analogous to $L_\varphi$ and $R_\varphi$ as for bilinear forms. There is one major difference here though in that above $L_\varphi$ and $R_\varphi$ could both be viewed as maps from $V$ to $V^\vee$, where here we need to take into account the conjugate linearity of the maps. As in the case of bilinear forms for each $v \in V$ we define the map $L_\varphi(v)$ by setting $L_\varphi(v)(u) = \varphi(v, u)$ for each $u \in V$. In this case for $v, v_1, v_2, u, u_1, u_2 \in V$ and $c \in F$ we have

$$L_\varphi(v_1 + cv_2)(u) = \varphi(v_1 + cv_2, u) = \varphi(v_1, u) + c \varphi(v_2, u) = L_\varphi(v_1)(u) + cL_\varphi(v_2)(u)$$

and

$$L_\varphi(v)(u_1 + cu_2) = \varphi(v, u_1 + cu_2) = \varphi(v, u_1) + c \varphi(v, u_2) = L_\varphi(v)(u_1) + cL_\varphi(v)(u_2).$$
Thus, we see that $L_\varphi : V \to \overline{V}^\vee$. As above, for each $v \in V$ define $R_\varphi(v)$ to be the map defined by $R_\varphi(v)(u) = \varphi(u, v)$ for each $u \in V$. Now for $v_1, v_2, u \in V$ and $c \in F$ we have

$$R_\varphi(v_1 + cv_2)(u) = \varphi(u, v_1 + cv_2) = \varphi(u, v_1) + \bar{c}\varphi(u, v_2) = R_\varphi(v_1)(u) + \bar{c}R_\varphi(v_2)(u),$$

e.g., $R_\varphi(v_1 + cv_2) = R_\varphi(v_1) + \bar{c}R_\varphi(v_2)$. Moreover, for each fixed $v$ and all $u_1, u_2 \in V, c \in F$, it is easy to check that

$$R_\varphi(v)(u_1 + cu_2) = R_\varphi(v)(u_1) + cR_\varphi(v)(u_2).$$

This shows that $R_\varphi$ is a conjugate-linear map and for each $v \in V$ we have $R_\varphi(v)$ is a linear map, i.e., $R_\varphi : V \to \overline{V}^\vee$.

Our next goal is to relate $L_\varphi$ and $R_\varphi$. This proceeds much as in the bilinear case, but the scalar actions can be a bit confusing so we provide the details here. The first thing to recall is the canonical isomorphism $V \to \overline{V}^\vee$; in our case we will need

$$\Phi : V \to (\overline{V}^\vee)^\vee
\quad v \mapsto \text{eval}_v.$$

We just briefly recall the proof that $\Phi$ is a linear map due to the potential confusion for the scalar multiplications. In particular, recall for $v \in V$ and $a \in F$ we have $a \cdot v = \overline{a}v$ but the scalar multiplication on $\overline{V}^\vee$ and $(\overline{V}^\vee)^\vee$ is given by the scalar multiplication induced by the multiplication in $F$. So, for example if $f \in \overline{V}^\vee, a \in F$, and $v_1, v_2 \in V$, then we have

$$f(a \cdot v_1 + v_2) = af(v_1) + f(v_2)$$

e.g., $f(\overline{a}v_1 + v_2) = af(v_1) + f(v_2)$. Thus, given $v_1, v_2 \in V, a \in F$, and $f \in \overline{V}^\vee$ we have

$$\Phi(a \cdot v_1 + v_2)(f) = \text{eval}_{a\overline{v_1} + v_2}(f) = f(\overline{a}v_1 + v_2) = af(v_1) + f(v_2) = \Phi(v_1)(f) + \Phi(v_2)(f) = (a\Phi(v_1) + \Phi(v_2))(f).$$

The proof that this is an injection in general and an isomorphism for finite dimensional cases is omitted as there is nothing confusing added to that. We now follow the argument given for bilinear forms to relate $L_\varphi$ and $R_\varphi$. We have $L_\varphi : (\overline{V}^\vee)^\vee \to \overline{V}^\vee$ and $R_\varphi : \overline{V} \to V^\vee$. We identify $\overline{V}$ and $(\overline{V}^\vee)^\vee$ as above. The
5.1. BASIC DEFINITIONS AND FACTS

5.1. BASIC DEFINITIONS AND FACTS

CHAPTER 5.

We can define the left and right kernels of $\varphi$ just as was done for bilinear forms. Once again we obtain $\varphi$ is non-degenerate if and only if $L_\varphi$ and $R_\varphi$ are injections. If $V$ is finite-dimensional, this is equivalent to $L_\varphi$ and $R_\varphi$ being isomorphisms.

Exercise 5.1.24. Show there is a bijection between isomorphisms $T : V \rightarrow V^\vee$ and non-degenerate sesquilinear forms $\varphi : V \times V \rightarrow V$.

We now define the matrix associated to a bilinear (resp. sesquilinear) form.

Definition 5.1.25. Let $\varphi : V \times V \rightarrow F$ be a bilinear or sesquilinear form. Let $B = \{v_1, \ldots, v_n\}$ be a basis of $V$. Set $a_{ij} = \varphi(v_i, v_j)$. The matrix associated to $\varphi$ is given by

\[ [\varphi]_B = A = (a_{ij}) \in \text{Mat}_n(F). \]

Theorem 5.1.26. Let $\varphi$ be a bilinear or sesquilinear form on a finite dimensional vector space $V$. Let $B$ be a basis of $V$. Then for $w, v \in V$ we have

\[ \varphi(v, w) = \sum_i [v]_B A[w]_B \]

if $\varphi$ is bilinear and

\[ \varphi(v, w) = \sum_i [v]_B A[w]_B \]

if $\varphi$ is sesquilinear.

Proof. This follows immediately upon calculating on a basis.

The definition of the matrix associated to a bilinear (resp. sesquilinear) form seems rather arbitrary in how we defined it. At this point the only justification is that the previous theorem shows this definition works how we expect it to. However, there is a more natural reason this is the correct definition to use. Let $\varphi$ be a bilinear or sesquilinear form. Let $B = \{v_1, \ldots, v_n\}$ be a basis of $V$ and $B^\vee = \{v_1^\vee, \ldots, v_n^\vee\}$ the dual basis of $V^\vee$. (Note that $B$ is also a basis for $V$.)

Observe we have

\[ R_\varphi(v_j) = \varphi(\cdot, v_j) \in V^\vee. \]

If $\varphi$ is a bilinear map this is a linear map from $V$ to $V^\vee$, so we can ask for $[R_\varphi]_B^{B^\vee}$. Similarly, if $\varphi$ is sesquilinear this is a linear map from $V^\vee$ to $V^\vee$ and so again we can consider $[R_\varphi]_B^{B^\vee}$. To calculate this, we need to expand $R_\varphi(v_j)$ in terms of $B^\vee$ for each $j$. Write

\[ R_\varphi(v_j) = c_1 v_1^\vee + \cdots + c_n v_n^\vee \]

for $c_i \in F$. The goal is to find $c_i$. Observe we have

\[ a_{ij} = \varphi(v_i, v_j) = R_\varphi(v_j)(v_i) = c_1 v_1^\vee(v_i) + \cdots + c_n v_n^\vee(v_i) = c_i. \]
Thus, $a_{ij} = c_i$ and so
\[ R_\varphi(v_j) = a_{1j} v_1^\vee + \cdots + a_{nj} v_n^\vee, \]
which gives
\[ [R_\varphi]_B^\vee = [\varphi]_B. \]
This shows that the matrix of a bilinear or sesquilinear form $\varphi$ is really just the matrix associated to the linear map $R_\varphi$.

**Corollary 5.1.27.** A bilinear or sesquilinear form $\varphi$ on a finite dimensional vector space $V$ is non-degenerate if and only if $[\varphi]_B$ is nonsingular for any basis $B$.

**Proof.** This follows immediately from the definition of non-degenerate and relation between $[\varphi]_B$ and the linear map $R_\varphi$.

Just as when studying linear maps, we would like to know how a change of basis effects the matrix of $\varphi$.

**Theorem 5.1.28.** Let $V$ be a finite dimensional $F$-vector space and let $B$ and $C$ be bases of $V$. If $P$ is the change of basis matrix from $C$ to $B$ then if $\varphi$ is bilinear we have
\[ [\varphi]_C = {}^t P [\varphi]_B P \]
and if $\varphi$ is sesquilinear we have
\[ [\varphi]_C = {}^t P [\varphi]_B \overline{P}. \]

**Proof.** We consider the case that $\varphi$ is sesquilinear. The proof when $\varphi$ is bilinear can be obtained by taking the conjugation to be trivial in this proof.

We have from the definition that for every $v, w \in V$
\[ \varphi(v, w) = {}^t [v]_B [\varphi]_B [w]_B, \]
and
\[ \varphi(v, w) = {}^t [v]_C [\varphi]_C [w]_C. \]

Moreover, since $P$ is the change of basis matrix from $C$ to $B$ we have $[v]_B = P[v]_C$ and $[w]_B = \overline{P}[w]_C$. Combining all of this we obtain
\[ {}^t [v]_C [\varphi]_C [w]_C = \varphi(v, w) = {}^t [v]_B [\varphi]_B [w]_B = {}^t (P[v]_C)[\varphi]_B P[w]_C = {}^t [v]_C {}^t (P\varphi_B \overline{P}[w]_C. \]

Since this is true for all $v, w \in V$, we must have
\[ [\varphi]_C = {}^t P [\varphi]_B \overline{P} \]
as claimed.
### 5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

It is very important to note the difference between this and changing bases for linear maps. We do not conjugate the matrix here, i.e., in Chapter 3 we had the equation 

\[ T_C = P^{-1}TP. \]

This means that you cannot in general use the results from Chapter 4 to find a nice basis for the matrix \( \phi_B \). Later we will discuss this further and determine some cases when one can use rational and Jordan canonical forms to obtain nice bases for forms as well.

**Definition 5.1.29.** Let \( A, B \in \text{Mat}_n(F) \). We say \( A \) and \( B \) are congruent (resp. conjugate congruent) if there exists \( P \in \text{GL}_n(F) \) so that \( B = tPAP \) (resp. \( B = tPA^TP \)).

**Exercise 5.1.30.** Show that congruence on matrices defines an equivalence relation.

The above results show that matrices \( A \) and \( B \) in \( \text{Mat}_n(F) \) represent the same bilinear (resp. sesquilinear) form if and only if \( A \) and \( B \) are congruent.

### 5.2 Symmetric, skew-symmetric, and Hermitian forms

In this section we specialize the forms we are looking at. This allows us to prove some very nice structure theorems. For the rest of this section, all of the bilinear and sesquilinear forms will be one of the forms given in the following definition. When we wish to emphasize a form \( \phi \) on \( V \), we write \((V, \phi)\) for the vector space.

**Definition 5.2.1.** Let \( V \) be a vector space over a field \( F \).

1. A bilinear form \( \phi \) on \( V \) is said to be symmetric if \( \phi(v, w) = \phi(w, v) \) for all \( v, w \in V \).
2. A bilinear form \( \phi \) on \( V \) is said to be skew-symmetric if \( \phi(v, w) = -\phi(w, v) \) for all \( v, w \in V \) and \( \phi(v, v) = 0 \) for all \( v \in V \).
3. A sesquilinear form \( \phi \) on \( V \) is said to be Hermitian if \( \phi(v, w) = \overline{\phi(w, v)} \) for all \( v, w \in V \).
4. A sesquilinear form \( \phi \) on \( V \) is said to be skew-Hermitian if \( \text{char}(F) \neq 2 \) and \( \phi(v, w) = -\overline{\phi(w, v)} \) for all \( v, w \in V \).

**Exercise 5.2.2.** Show that if \( \text{char}(F) \neq 2 \) then \( \phi \) is skew-symmetric if and only if \( \phi(v, w) = -\phi(w, v) \) for all \( v, w \in V \).

**Exercise 5.2.3.** Show that if \( \phi \) is any of the forms given in this definition \( \ker \phi \) is well-defined since \( \ker_R \phi = \ker_L \phi \).

One should note that it is not often one encounters skew-Hermitian forms. The reason for this is as follows. Let \( F \) be a field with \( \text{char}(F) \neq 2 \) with non-trivial conjugation. Then we saw before that there is an element \( j \in F \) so that \( \overline{j} = -j \) and every element \( z \in F \) can be written uniquely in the form \( z = x + jy \)
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

with $x, y \in F_0$. Let $\varphi$ be a skew-Hermitian form. Set $\psi(v, w) = j \varphi(v, w)$. We claim that $\psi$ is Hermitian. To see this, observe

\[
\psi(v, w) = j \varphi(v, w) \\
= j \varphi(v, w) \\
= (-j)(-\varphi(w, v)) \\
= \psi(v, w).
\]

Similarly, given a Hermitian form one can multiply by $j$ to obtain a skew-Hermitian form. Thus, we see that we can move back and forth between Hermitian and skew-Hermitian forms simply by multiplying by $j$. This means there is essentially nothing new in studying skew-Hermitian forms.

**Lemma 5.2.4.** Let $(V, \varphi)$ be a finite dimensional $F$-vector space. Let $B$ be a basis for $V$ and set $A = [\varphi]_B$. Then we have:

1. $\varphi$ is symmetric if and only if $^tA = A$.
2. $\varphi$ is skew-symmetric if and only if $^tA = -A$.
3. $\varphi$ is Hermitian if and only if $^tA = \overline{A}$.

The proof of this lemma is left as an exercise. It amounts to converting the definition to matrix language.

**Definition 5.2.5.** Let $A \in \text{Mat}_n(F)$.

1. We say $A$ is **symmetric** if $^tA = A$.
2. We say $A$ is **skew-symmetric** if $^tA = -A$.
3. We say $A$ is **Hermitian** if $^tA = \overline{A}$.

One should keep in mind that the notion we are generalizing is that of the dot product on $\mathbb{R}^n$. As such, the correct notion of equivalence between spaces $(V, \varphi)$ and $(W, \psi)$ should preserve “distance”, i.e., it should be a generalization of the notion of isometry from Euclidean geometry.

**Definition 5.2.6.** Let $(V, \varphi)$ and $(W, \psi)$ be $F$-vector spaces. We say $T \in \text{Hom}_F(V, W)$ is an **isometry** if $T$ is an isomorphism and satisfies

\[
\varphi(v_1, v_2) = \psi(T(v_1), T(v_2))
\]

for all $v_1, v_2 \in V$.

As is the custom, we rephrase this definition in terms of matrices.

**Lemma 5.2.7.** Let $T \in \text{Hom}_F(V, W)$ with $(V, \varphi)$ and $(W, \psi)$ finite dimensional $F$-vector spaces of dimension $n$. Let $B$ be a basis for $V$ and $C$ a basis for $W$. Set $P = [T]_B^C$. Then we have $T$ is an isometry if and only if $P \in \text{GL}_n(F)$ and

\[
^tP[\psi]_C P = [\varphi]_B
\]
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

CHAPTER 5.

for $\varphi$ and $\psi$ bilinear and

$^t\!P [\psi]_C \bar{P} = [\varphi]_B$

if $\varphi$ and $\psi$ are sesquilinear.

Proof. Exercise. \hfill $\Box$

Definition 5.2.8. Let $\varphi$ be a form on $V$. The isometry group of $\varphi$ is given by

$\text{Isom}(V, \varphi) = \{ T \in \text{Hom}_F (V, V) : T \text{ is an isometry} \}.$

When $V$ is clear from context we will write $\text{Isom}(\varphi)$.

Exercise 5.2.9. Let $(V, \varphi)$ and $(W, \psi)$ be $F$-vector spaces. Show that $(V, \varphi)$ is isometric to $(W, \psi)$ if and only if the matrix relative to $\varphi$ is congruent to the matrix of $\psi$ if $\varphi$ and $\psi$ are bilinear. If they are sesquilinear, show the same statement with conjugate congruent.

Exercise 5.2.10. Show that $\text{Isom}(V, \varphi)$ is a group under composition.

One should keep in mind that studying isometry groups of a space gives information back about the geometry of the space $(V, \varphi)$. We will see more detailed examples of this in the next chapter when we restrict ourselves to inner product spaces.

Depending on the type of form $\varphi$ we give the isometry group special names.

1. Let $\varphi$ be a nondegenerate symmetric bilinear form. We write $O(\varphi)$ for the isometry group and refer to this as the orthogonal group associated to $\varphi$. We call isometries in this group orthogonal maps.

2. Let $\varphi$ be a nondegenerate Hermitian form. We write $U(\varphi)$ for the isometry group and refer to this as the unitary group associated to $\varphi$. We call isometries in this group unitary maps.

3. Let $\varphi$ be a skew-symmetric bilinear form. We write $\text{Sp}(\varphi)$ for the isometry group and refer to this as the symplectic group associated to $\varphi$. We call isometries in this group symplectic maps.

Recall that when studying linear transformations $T \in \text{Hom}_F (V, V)$ in Chapter 4, the first step was to break $V$ into a direct sum of $T$-invariant subspaces. We want to follow the same general pattern here, but first need the appropriate notion of breaking $V$ up into pieces for a form $\varphi$.

Definition 5.2.11. Let $\varphi$ be a form on $V$. We say $v, w \in V$ are orthogonal with respect to $\varphi$ if

$\varphi(v, w) = \varphi(w, v) = 0.$

We say subspaces $V_1, V_2 \subset V$ are orthogonal if

$\varphi(v_1, v_2) = \varphi(v_2, v_1) = 0$

for all $v_1 \in V_1, v_2 \in V_2$. 

109
We once again emphasize that one should be keeping in mind the notion of a dot product on $\mathbb{R}^n$ when thinking of these definitions. One then sees that in this special case this notion of orthogonal is the same as the notion that was introduced in vector calculus.

**Definition 5.2.12.** Let $V_1, V_2 \subset V$ be orthogonal subspaces. We say $V$ is the **orthogonal direct sum of $V_1$ and $V_2$** and write $V = V_1 \perp V_2$ if $V = V_1 \oplus V_2$.

**Exercise 5.2.13.** 1. Show that $V = V_1 \perp V_2$ if $V = V_1 \oplus V_2$ and given any $v, v' \in V$, when we write $v = v_1 + v_2$ and $v' = v_1' + v_2'$ for $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ we have

$$\varphi(v, v') = \varphi(v_1, v_1') + \varphi(v_2, v_2').$$

2. Suppose $V = V_1 \perp V_2$ and $V$ is finite dimensional. Let $\varphi_1 = \varphi|_{V_1 \times V_1}$ and $\varphi_2 = \varphi|_{V_2 \times V_2}$. If we let $B_1$ be a basis for $V_1$, $B_2$ a basis for $V_2$, and $B = B_1 \cup B_2$, then show that

$$[\varphi]_B = \left( \begin{array}{cc} [\varphi_1]_{B_1} & 0 \\ 0 & [\varphi_2]_{B_2} \end{array} \right).$$

The next result shows that even if one is given a degenerate form, one can split off the kernel of this form and then be left with a nondegenerate form. This will allow us to focus our attention on nondegenerate forms when proving our classification theorems.

**Lemma 5.2.14.** Let $\varphi$ be a form on a finite dimensional $F$-vector space $V$. Then there is a subspace $V_1$ of $V$ so that $\varphi_1 = \varphi|_{V_1}$ is nondegenerate and $V = \ker(\varphi) \perp V_1$. Moreover, $(V_1, \varphi_1)$ is well-defined up to isometry.

**Proof.** The exercise above gives that $\ker(\varphi)$ is a subspace of $V$, so it certainly has a complement $V_1$, i.e., we can write $V = \ker(\varphi) \oplus V_1$ for some subspace $V_1 \subset V$. Set $\varphi_1 = \varphi|_{V_1}$. The definition of $\ker(\varphi)$ immediately gives that $V = \ker(\varphi) \perp V_1$.

The next step is to show that $\varphi_1$ is nondegenerate, i.e., $R_{\varphi_1}$ is injective. (Note we use finite dimensional here to conclude $R_{\varphi_1}$ being injective is equivalent to being an isomorphism.) Suppose there exists $v_1 \in V_1$ so that $R_{\varphi_1}(v_1) = 0$, i.e., $0 = R_{\varphi_1}(v_1)(w_1) = \varphi(v_1, w_1) = \varphi(w_1, v_1)$ for every $w_1 \in V_1$. Moreover, if $w \in V - V_1$, then $w \in \ker(\varphi)$ and so $\varphi(w, v_1) = 0$. Thus, $\varphi(w, v_1) = 0$ for all $w \in V$, i.e. $v_1 \in \ker(\varphi)$. However, this is a contradiction since $V_1 \cap \ker(\varphi) = \{0\}$. Thus, $\varphi_1$ must be nondegenerate.

The last thing to show is that $V_1$ is well-defined up to isometry. Set $V' = V/\ker(\varphi)$. We define a form $\varphi'$ on $V'$ as follows. Let $\pi : V \to V'$ be the standard projection map. For $v', w' \in V'$, let $v, w \in V$ such that $\pi(v) = v'$, $\pi(w) = w'$. Set

$$\varphi'(v', w') = \varphi(v, w).$$

This is a well-defined form on $V'$ as we are quotienting out by the kernel. Thus, $\pi|_{V_1}$ is an isometry between $(V_1, \varphi_1)$ and $(V', \varphi')$. This gives the result. \qed
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

CHAPTER 5.

Definition 5.2.15. Let \( \varphi \) be a form on a finite dimensional vector space \( V \). The rank of \( \varphi \) is the dimension of \( V / \ker(\varphi) \).

We saw in the proof of the previous lemma that given \( \varphi \), there is a subspace \( V_1 \subset V \) so that \( V = \ker(\varphi) \perp V_1 \). In general, given a subspace \( W \subset V \), we can ask if there is a subspace \( W^\perp \) so that \( V = W \perp W^\perp \).

Definition 5.2.16. Let \( W \subset V \) be a subspace. The orthogonal complement of \( W \) is given by

\[
W^\perp = \{ v \in V : \varphi(w, v) = 0 \text{ for all } w \in W \}.
\]

Exercise 5.2.17. Show that \( W^\perp \) is a subspace of \( V \).

Lemma 5.2.18. Let \( W \) be a finite dimensional subspace of \( (V, \varphi) \) and set \( \psi = \varphi|_{W \times W} \). If \( \psi \) is nondegenerate then \( V = W \perp W^\perp \). If \( V \) is finite dimensional and \( \varphi \) is nondegenerate as well then \( \psi^\perp = \varphi|_{W^\perp \times W^\perp} \) is also nondegenerate.

Proof. By definition we have \( W \) and \( W^\perp \) are orthogonal, so to show \( V = W \perp W^\perp \) we need only show that \( V = W \oplus W^\perp \).

Let \( v_0 \in W \cap W^\perp \). Observe that \( \varphi(v_0, v_0) = 0 \) because \( v_0 \in W \) and \( v_0 \in W^\perp \). Using that \( \psi \) is nondegenerate we obtain \( v_0 = 0 \) and so \( W \cap W^\perp = \{0\} \).

Let \( v_0 \in V \). Note that since \( \psi \) is assumed to be nondegenerate and \( W \) is finite dimensional, \( R_\psi : W \to W^\vee \) is an isomorphism. Thus, for any \( T \in W^\vee \) there is an element \( w_T \in W \) so that \( R_\psi(w_T) = T \). We now apply this fact to our situation. We have \( R_\varphi(v_0)|_W \in W^\vee \), so there is a \( w_0 \in W \) with

\[
R_\varphi(v_0)|_W = R_\psi(v_0)|_W,
\]

i.e., for every \( w \in W \) we have

\[
\psi(w, v_0) = R_\psi(v_0)(w) = R_\varphi(v_0)|_W(w) = \varphi(w, v_0).
\]

Thus, we have

\[
\varphi(w, v_0) = \psi(w, v_0) = \varphi(w, v_0)
\]

where we have used \( \varphi_{W \times W} = \psi \). Subtracting we have that for every \( w \in W \)

\[
\varphi(w, v_0 - w_0) = 0,
\]

i.e., \( v_0 - w_0 \in W^\perp \). This allows us to write

\[
v_0 = w_0 + (v_0 - w_0) \in W + W^\perp.
\]

Thus, \( V = W \perp W^\perp \) as desired.
Now suppose that \( \varphi \) is nondegenerate. Let \( v_0 \in W^\perp \). Since \( \varphi \) is nondegenerate there exists \( v \in V \) with \( \varphi(v, v_0) \neq 0 \). Write \( v = w_1 + w_2 \) with \( w_1 \in W \) and \( w_2 \in W^\perp \). Then

\[
0 \neq \varphi(v, v_0) = \varphi(w_1 + w_2, v_0) = \varphi(w_1, v_0) + \varphi(w_2, v_0) = \varphi(w_2, v_0)
\]

where we have used that \( \varphi(w_1, v_0) = 0 \) because \( w_1 \in W \) and \( v_0 \in W^\perp \). Thus, \( R_\varphi(v_0)(w_2) \neq 0 \) and since \( v_0 \) was arbitrary, this gives \( R_\varphi|_{W^\perp} \) is injective. Since we are assuming \( V \) is finite dimensional, this is enough to conclude that \( R_\varphi|_{W^\perp} \) is an isomorphism, i.e., \( \varphi|_{W^\perp \times W^\perp} \) is nondegenerate.

**Example 5.2.19.** It is not in general the case that if \( \varphi \) is a non-degenerate on a vector space \( V \) than \( \varphi|_{W \times W} \) is non-degenerate for a subspace \( W \subset V \). For instance, let \( V = F^2 \) with standard basis \( \mathcal{E}_2 = \{e_1, e_2\} \). Set \( W = \text{span}_F \{e_1\} \).

Consider the bilinear form defined by the matrix \[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]. It is easy to see this is a non-degenerate form on \( V \), but restricted to \( W \) this form is degenerate.

**Lemma 5.2.20.** Let \( W \) be a subspace of \((V, \varphi)\) with \( V \) finite dimensional. Assume \( \varphi|_{W \times W} \) and \( \varphi|_{W^\perp \times W^\perp} \) are both nondegenerate. Then \( (W^\perp)^\perp = W \).  

**Proof.** We can write \( V = W \perp W^\perp \) and \( V = W^\perp \perp (W^\perp)^\perp \) by using that \( W \) and \( W^\perp \) are both subspaces of \( V \). This immediately gives that \( \dim_F W = \dim_F (W^\perp)^\perp \). It is now easy to see using the definition that \( W \subset (W^\perp)^\perp \). Since they have the same dimension, we must have equality.

We will later see how to prove the above theorem in the case that \( W \) is finite dimensional without the need to assume that \( V \) is finite dimensional as well. Our next step is to classify forms on finite dimensional vector spaces where possible. The following result is essential in the case of \( \varphi \) symmetric or Hermitian.

**Lemma 5.2.21.** Let \((V, \varphi)\) be an \( F \)-vector space and assume that \( \varphi \) is nondegenerate. If \( \text{char}(F) \neq 2 \), assume \( \varphi \) is symmetric or Hermitian. If \( \text{char}(F) = 2 \), assume \( \varphi \) is Hermitian. Then there is a vector \( v \in V \) with \( \varphi(v, v) \neq 0 \).

**Proof.** Let \( v_1 \in V \) with \( v_1 \neq 0 \). If \( \varphi(v_1, v_1) \neq 0 \) we are done, so assume \( \varphi(v_1, v_1) = 0 \). The fact that \( \varphi \) is nondegenerate gives a nonzero \( v_2 \in V \) so that \( \varphi(v_1, v_2) \neq 0 \). Set \( b = \varphi(v_1, v_2) \). If \( \varphi(v_2, v_2) \neq 0 \) we are done, so assume \( \varphi(v_2, v_2) = 0 \). Set \( v_3 = tv_1 + v_2 \) where \( t \in F \). Then we have

\[
\varphi(v_3, v_3) = \varphi(tv_1 + v_2, tv_1 + v_2) = t\overline{t}\varphi(v_1, v_1) + t\overline{t}\varphi(v_1, v_2) + \overline{t}t\varphi(v_2, v_1) + \varphi(v_2, v_2) = tb + \overline{t}b
\]
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

where \( t = t \) and \( b = b \) if \( \varphi \) is symmetric. Thus, if \( V \) is symmetric \( \varphi(v_3, v_3) = 2tb \)
so choose any \( t \neq 0 \) and we are done. Now suppose that \( \varphi \) is Hermitian so
\( \varphi(v_3, v_3) = tb + \bar{t}b \). Set \( t = 1/b \) so \( tb = 1 \). We claim \( \bar{t} = 1 \). To see this, observe
\( \bar{t} = \bar{t} \cdot \bar{t} = \bar{t} \cdot \bar{t} \), so \( \bar{t} = 1 \). Thus, \( \varphi(v_3, v_3) = 2 \neq 0 \) as long as \( \text{char}(F) \neq 2 \).

Now suppose that \( \text{char}(F) = 2 \). We want to set \( t = a/b \) where \( a \neq a, -a \). Using
that the conjugation is non-trivial, we know there exists \( j \in F \) so that \( \bar{j} = -j \) and every element of \( F \) can be written uniquely in the form \( a = \alpha + j\beta \) with \( \alpha, \beta \in F_0 \). If \( \beta \neq 0 \) then \( a \neq \bar{a} \) and if \( \alpha \neq 0 \) then \( a \neq -\bar{a} \). Thus, just pick \( a \) to
be any element with \( \alpha\beta \neq 0 \). Set \( t = a/b \) so that \( \varphi(v_3, v_3) = a + \bar{a} \neq 0 \). Thus,
we have the result in all cases.

Suppose we have a basis \( B = \{v_1, \ldots, v_n\} \) for \( (V, \varphi) \) so that \( V = \text{span}_F \{v_i\} \)
satisfies \( V = V_1 \perp \cdots \perp V_n \).

Observe this gives
\[
[\varphi]_B = \begin{pmatrix}
a_1 & & \\
& \ddots & \\
& & a_n
\end{pmatrix}
\]

where \( a_i = \varphi(v_i, v_i) \). This leads to the following definition.

**Definition 5.2.22.** Let \( \varphi \) be a symmetric bilinear form or a Hermitian form
on a finite dimensional vector space \( V \). We say \( \varphi \) is **diagonalizable** if there are
1-dimensional subspaces \( V_1, \ldots, V_n \) so that \( V = V_1 \perp \cdots \perp V_n \).

Let \( a \in F \) with \( a \neq 0 \). We can define a bilinear form on \( F \) by considering
\( a \in \text{Mat}_1(F) \), i.e., we define \( \varphi_a(x, y) = xay \). Here we get \( \varphi_a \) is symmetric for
free. Similarly, if \( F \) has nontrivial conjugation we can define a sesquilinear form
\( \varphi_a \) by setting \( \varphi_a(x, y) = xay \). This is not Hermitian for free. In particular, we have
\[
\varphi_a(x, y) = xay = \overline{ya} = \overline{a} \overline{y}.
\]
This is equal to \( \overline{\varphi_a}(y, x) \) if and only if \( \overline{a} = a \). Thus, we have \( \varphi_a \) is Hermitian if
and only if \( \overline{a} = a \). In either case, we write \( [a] \) for the space \( (F, \varphi_a) \). This set-up allows us to rephrase the definition of diagonalizable to say \( \varphi \) is diagonalizable
if there exist \( a_1, \ldots, a_n \) so that \( \varphi \) is isometric to \( [a_1] \perp \cdots \perp [a_n] \).

**Theorem 5.2.23.** Let \( (V, \varphi) \) be a finite dimensional vector space over a field
\( F \). If \( \text{char}(F) \neq 2 \) we assume \( \varphi \) is symmetric or Hermitian. If \( \text{char}(F) = 2 \), we
require \( \varphi \) to be Hermitian. Then \( \varphi \) is diagonalizable.

**Proof.** We immediately reduce to the case that \( \varphi \) is nondegenerate by recalling
that there is a subspace \( V_1 \) so that \( \varphi|_{V_1 \times V_1} \) is nondegenerate and \( V = \ker(\varphi) \perp V_1 \).
Thus, if \( \varphi|_{V_1 \times V_1} \) is diagonalizable, then certainly \( \varphi \) is as well since the
restriction to \( \ker(\varphi) \) is just 0.
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

CHAPTER 5.

We now assume \( \varphi \) is nondegenerate and induct on the dimension of \( V \). The case that \( \dim_F V = 1 \) is trivial. Assume the result is true for all vector spaces of dimension less than \( n \) and let \( \dim_F V = n \). We have a \( v_1 \in V \) so that \( \varphi(v_1, v_1) \neq 0 \) by Lemma 5.2.21. Set \( a_1 = \varphi(v_1, v_1) \) and let \( V_1 = \text{span}_F \{v_1\} \). Since \( \varphi|_{V_1} \) is nondegenerate we have \( V = V_1 \perp V_1^\perp \). However, this gives \( \dim_F V_1^\perp = n - 1 \) and Lemma 5.2.18 gives that \( \varphi_1 := \varphi|_{V_1^\perp \times V_1^\perp} \) is nondegenerate. We apply the induction hypothesis to \((V_1^\perp, \varphi_1)\) to obtain one dimensional subspaces \( V_2, \ldots, V_n \) so that \( V_1^\perp = V_2 \perp \cdots \perp V_n \). Thus, \( V = V_1 \perp V_2 \perp \cdots \perp V_n \) and we are done.

We can use this theorem to give our first classification theorem. The following theorem is usually stated strictly over algebraically closed fields, but we state it a little more generally as algebraically closed is not required.

**Theorem 5.2.24.** Let \((V, \varphi)\) be an \( n \)-dimensional \( F \)-vector space with \( \varphi \) a nondegenerate symmetric bilinear form. Suppose that \( F \) satisfies that \( \text{char}(F) \neq 2 \) and it is closed under square-roots, namely, given any \( a \in F \), one has \( f(x) = x^2 - a \in F[x] \) has a root in \( F \). (In particular, if \( F \) is algebraically closed this is certainly the case.) Then \( \varphi \) is isometric to \([1] \perp \cdots \perp [1]\). In particular, all nondegenerate symmetric bilinear forms over such a field are isometric.

**Proof.** The previous theorem gives one dimensional subspaces \( V_1, \ldots, V_n \) so that \( V = V_1 \perp \cdots \perp V_n \). Let \( \{v_i\} \) be a basis for \( V_i \) and set \( a_i = \varphi(v_i, v_i) \). We choose \( v_i \) so that \( a_i \neq 0 \). By our assumption on \( F \) we have an element \( b_i \in F \) so that \( b_i \) is a root of \( f(x) = x^2 - 1/a_i \), i.e., \( b_i^2 = 1/a_i \). Set \( B = \{b_1v_1, \ldots, b_nv_n\} \). Then we have

\[
\varphi(b_iv_i, b_iv_i) = b_i^2 \varphi(v_i, v_i) = b_i^2 a_i = 1.
\]

This gives the result.

In the case that one has \( \varphi \) is isometric to \([1] \perp \cdots \perp [1]\) with \( n \) copies of \([1]\), we will write \( \varphi \simeq n[1] \).

Recall that the isometry group of a symmetric bilinear form \( \varphi \) was denoted \( O(\varphi) \). Consider now the case that \((V, \varphi)\) is defined over a field \( F \) that is closed under square-roots. Then we have just seen there is a basis \( B \) so that \([\varphi]_B = 1_n \) where \( 1_n \) is the \( n \times n \) identity matrix. Using this basis we can represent the orthogonal group as

\[
O_n(\varphi) = \{M \in \text{GL}_n(F) : \ iM1_nM = 1_n\} = \{M \in \text{GL}_n(F) : \ iM = M^{-1}\}.
\]

We refer to matrices \( M \in O_n(F) \) as orthogonal matrices. One important point to note here is that we saw before that change of basis matrices for bilinear and sesquilinear forms correspond to isometries. Thus, in this case we see that
change of basis matrices for non-degenerate symmetric bilinear forms correspond
to invertible matrices $M$ satisfying $M^2 = M^{-1}$. This shows that changing bases
for such forms is equivalent to changing the basis of a linear transformation. In
other words, one can apply all the nice results of the previous chapter to the
matrices of non-degenerate symmetric bilinear forms.

Note that the proof of Theorem 5.2.24 does not work for $\varphi$ a Hermitian form
even if $F$ is algebraically closed because in order to scale the $v_i$, we would need
$bi$ so that $b_i b_i = a_i$. However, this equation cannot be set up as a polynomial
equation and so algebraically closed does not guarantee a solution to such an
equation exists. To push our classifications further we need to restrict the fields
of interest some. In general one needs an ordered field. However, to save the
trouble of introducing ordered fields we will consider our fields $F$ to satisfy
$\mathbb{Q} \subset F \subset \mathbb{C}$. In addition, if we wish to consider symmetric forms, we will
require $F \subset \mathbb{R}$. If we wish to consider Hermitian forms, we only require $F$ has
nontrivial conjugation. One important point to note here is that with our set-
up, if $\varphi$ is symmetric we certainly have $\varphi(v, v) \in \mathbb{R}$ for all $v \in V$. However, if $\varphi$
is Hermitian we have for any $v \in V$ that $\varphi(v, v) \in \mathbb{C}$ satisfies $\varphi(v, v) = \overline{\varphi(v, v)}$,
i.e., $\varphi(v, v) \in \mathbb{R}$ in this case as well. This allows us to make the following
definition.

**Definition 5.2.25.** Let $(V, \varphi)$ be an $F$-vector space so that if $\varphi$ is symmetric,
$\mathbb{Q} \subset F \subset \mathbb{R}$ and if $\varphi$ is Hermitian $\mathbb{Q} \subset F \subset \mathbb{C}$. We say $\varphi$ is *positive definite* if
$\varphi(v, v) > 0$ for all nonzero $v \in V$ and $\varphi$ is *negative definite* if $\varphi(v, v) < 0$
for all nonzero $v \in V$. We say $\varphi$ is *indefinite* if there are vectors $v, w \in V$ so that
$\varphi(v, v) > 0$ and $\varphi(w, w) < 0$.

**Definition 5.2.26.** We say a matrix $A \in \text{Mat}_n(F)$ is *positive definite* if the
associated form $\varphi$ is positive definite. Likewise, we say $A$ is *negative definite*
if the associated form is negative definite.

With these definitions we give the next classification theorem.

**Theorem 5.2.27** (Sylvester’s Law of Inertia). Let $(V, \varphi)$ be a finite dimensional
$\mathbb{R}$-vector space with $\varphi$ a nondegenerate symmetric bilinear form on $V$ or let
$(V, \varphi)$ be a $\mathbb{C}$-vector space with $\varphi$ a nondegenerate Hermitian form on $V$. Then
$\varphi$ is isometric to $p[1] \perp q[-1]$ for well-defined integers $p, q$ with $p + q = \dim_{\mathbb{R}} V$.

**Proof.** The proof that $\varphi$ is isometric to $p[1] \perp q[-1]$ follows the same argument
as used in the proof of Theorem 5.2.24. We begin with the case that $(V, \varphi)$
is a $\mathbb{R}$-vector space and $\varphi$ is a nondegenerate symmetric bilinear form. As in
the proof of Theorem 5.2.24, we have one-dimensional subspaces $V_1, \ldots, V_n$ so
that $V_i = \text{span}_\mathbb{R} \{ v_i \}$ and $V = V_1 \perp \cdots \perp V_n$. Write $a_i = \varphi(v_i, v_i)$. Let
$b_i = 1/\sqrt{|a_i|} \in \mathbb{R}$. Then $b_i^2 = 1/|a_i|$ and so

\[
\varphi(b_i v_i, b_i v_i) = b_i^2 \varphi(v_i, v_i) = \frac{a_i}{|a_i|} = \text{sign}(a_i)
\]
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

CHAPTER 5.

where \( \text{sign}(a_i) = 1 \) if \( a_i > 0 \) and \( \text{sign}(a_i) = -1 \) if \( a_i < 0 \). Thus, if we let \( p \) be the number of positive \( a_i \) and \( q \) the number of negative \( a_i \) we get \( \varphi \simeq p[1] \perp q[-1] \).

Now suppose \((V, \varphi)\) is a \( \mathbb{C}\)-vector space and \( \varphi \) is a nondegenerate Hermitian form. Given any \( v \in V \) and \( b \in \mathbb{C} \) we have \( \varphi(bv, bv) = \overline{b} \varphi(v,v) = |b|^2 \varphi(v,v) \). Proceeding as above, this shows that for each \( j \) we can scale \( \varphi(v_j, v_j) \) by any positive real number. Since \( \varphi(v_j, v_j) \in \mathbb{R} \) for all \( j \), we again have \( \varphi \simeq p[1] \perp q[-1] \) where \( p \) and \( q \) are defined as above.

It remains to show that \( p \) and \( q \) are invariants of \( \varphi \) and do not depend on the choice of basis. Let \( V_+ \) be the largest subspace of \( V \) so that the restriction of \( \varphi \) to \( V_+ \) is positive definite and \( V_- \) the largest subspace of \( V \) so that the restriction of \( \varphi \) to \( V_- \) is negative definite. Set \( p_0 = \dim_{\mathbb{R}} V_+ \) and \( q_0 = \dim_{\mathbb{R}} V_- \). Note that \( p_0 \) and \( q_0 \) are well-defined and do not depend on any choices. We now show \( p = p_0 \) and \( q = q_0 \).

Let \( B = \{v_1, \ldots, v_n\} \) be a basis of \( V \) so that \([\varphi]_B = p[1] \perp q[-1] \). Set \( B_+ = \{v_1, \ldots, v_p\} \) and \( B_- = \{v_{p+1}, \ldots, v_n\} \). Let \( W_+ = \text{span}_F B_+ \) and \( W_- = \text{span}_F B_- \). Then we have \( \varphi \) restricted to \( W_+ \) is positive definite so \( p \leq p_0 \) and similarly we obtain \( q \leq q_0 \). Note that this gives \( n = p + q \leq p_0 + q_0 \). Suppose \( p \neq p_0 \). Then \( \dim F V_+ + \dim F V_- > n \), so \( V_+ \cap V_- \neq \{0\} \). This is easily seen to be a contradiction, so \( p = p_0 \) and the same argument gives \( q = q_0 \), completing the proof.

One important point to note is all that was really used in the above proof was that given any positive number \(|a_i|\), one had \( 1/\sqrt{|a_i|} \) in \( F \). Thus, we do not need to use \( F = \mathbb{R} \) or \( F = \mathbb{C} \) in the theorem. For example, if \((V, \varphi)\) is an \( F \)-vector space with \( \varphi \) is a symmetric bilinear form and \( F \subset \mathbb{R} \), to apply the result to \((V, \varphi)\) we only require that \( \sqrt{|a|} \in F \) for every \( a \in F \). If the field does not contain all its positive square roots things are more difficult as we will see below.

**Definition 5.2.28.** Let \((V, \varphi), p, \) and \( q \) be as in the previous theorem. The **signature** of \( \varphi \) is \((p, q)\).

**Corollary 5.2.29.** A nondegenerate symmetric bilinear form on a finite dimensional \( \mathbb{R}\)-vector space or a nondegenerate Hermitian form on a finite dimensional \( \mathbb{C}\)-vector space is classified up to isometry by its signature. If \( \varphi \) is not required to be nondegenerate, it is classified by its signature and its rank.

**Proof.** This follows immediately from the previous theorem with the exception of the degenerate case. However, if we do not require \( \varphi \) to be nondegenerate, then write \( V = \ker(\varphi) \perp V_1 \) and apply Sylvester’s law to \( V_1 \) and we obtain the result.

If we allow \( \varphi \) to be degenerate, the previous result gives \( \varphi \simeq p[1] \perp q[-1] \perp r[0] \) for \( r = n - p - q \).

We again briefly return to the isometry groups. Let \( \varphi \) be a nondegenerate symmetric bilinear form on a real vector space \( V \) of signature \((p, q)\) so \( \varphi \simeq p[1] \perp q[-1] \perp r[0] \).
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

CHAPTER 5.

Let \( p, q \) be given by \( 1_{p,q} = \begin{pmatrix} 1_p & -1_q \end{pmatrix} \).

Then the isometry group of \( \varphi \) is given by

\[
O_{p,q}(\mathbb{R}) := O(\varphi) = \{ M \in \text{GL}_{p+q}(\mathbb{R}) : {}^tM1_{p,q}M = 1_{p,q} \}.
\]

In the case that \( p = n = \dim_{\mathbb{R}} V \) and \( q = 0 \) we recover the case given above for a positive definite symmetric bilinear form.

Let \( \varphi \) be a nondegenerate Hermitian form on a complex vector space \( V \) of signature \((p,q)\) so \( \varphi \simeq p[1] \perp q[-1] \). The isometry group in this case is given by

\[
U_{p,q}(\mathbb{C}) := U(\varphi) = \{ M \in \text{GL}_{p+q}(\mathbb{C}) : {}^tM1_{p,q}M = 1_{p,q} \}.
\]

One will also often see this group denoted at \( U(p,q) \) when \( \mathbb{C} \) is clear from context. In the case \( p = n = \dim_{\mathbb{C}} V \) and \( q = 0 \), we have

\[
U_n(\mathbb{C}) := U_{n,0}(\mathbb{C}) = \{ M \in \text{GL}_n(\mathbb{C}) : {}^tM1_nM = 1_n \}.
\]

We now briefly address the situation of a nondegenerate symmetric bilinear or Hermitian form on a vector space over a field \( \mathbb{Q} \subset F \subset \mathbb{C} \) that is not closed under taking square roots of positive numbers. (If \( F \) is not a subset of \( \mathbb{R} \), we realize the “positive” numbers in \( F \) as \( F \cap \mathbb{R}_{>0} \). Note by requiring \( F \subset \mathbb{C} \) this makes sense.) Let \( \varphi \) be such a form. Theorem 5.2.23 gives elements \( a_1, \ldots, a_n \) so that \( \varphi \simeq [a_1] \perp \cdots \perp [a_n] \). We know that each \( a_i \) can be scaled by any element \( b_i^2 \) for \( b_i \in F \). This gives a nice diagonal representation for such forms. Let \( F^x = F - \{0\} \) and set \( (F^x)^2 = \{ b^2 : b \in F^x \} \). The above discussion may lead one to believe that there is a bijection between nondegenerate symmetric bilinear forms on an \( n \)-dimensional \( F \)-vector space \( V \) and \( (F^x/(F^x)^2)^n \). However, the following example shows this does not work even in the case \( F = \mathbb{Q} \), namely, there is not a well-defined map from the collection of nondegenerate symmetric bilinear forms to \( (F^x/(F^x)^2)^n \). One can relate the study of such forms to the classification of quadratic forms, but one quickly sees it is a very difficult problem in this level of generality.

**Example 5.2.30.** Consider the symmetric bilinear form \( \varphi \) on \( \mathbb{Q}^2 \) given by

\[
\varphi(x,y) = {}^t_x \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} y.
\]

Let \( P = \begin{pmatrix} 1 & 3 \\ 1 & -2 \end{pmatrix} \). Then we have

\[
\begin{pmatrix} 5 & 0 \\ 0 & 30 \end{pmatrix} = {}^tP \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} P,
\]

and so we have \( \varphi \) is isometric to \( \begin{pmatrix} 5 & 0 \\ 0 & 30 \end{pmatrix} \), but the diagonal entries of this do not differ from 2 and 3 by squares.
5.2. SYMMETRIC, SKEW-SYMMETRIC, AND HERMITIAN FORMS

We next classify skew-symmetric forms.

**Theorem 5.2.31.** Let $(V, \varphi)$ be a finite dimensional vector space over a field $F$ and let $\varphi$ be a skew-symmetric form. Then $n = \dim_F V$ must be even and $\varphi$ is isometric to $n/2$ copies of

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

i.e., there is a basis $B$ so that

$$
[\varphi]_B = \begin{pmatrix}
0 & 1 & & \\
-1 & 0 & & \\
& & \ddots & \\
& & & 0 & 1 \\
& & & -1 & 0
\end{pmatrix}.
$$

**Proof.** We use induction on $n = \dim_F V$. Let $n = 1$ and $B = \{v_1\}$ be a basis for $V$. Since $\varphi$ is skew-symmetric we have $\varphi(v_1, v_1) = 0$, which gives $[\varphi]_B = [0]$. This contradicts $\varphi$ being nondegenerate. Since the $n = 1$ case is reasonably trivial, we also show the $n = 2$ case as it is more illustrative of the general case.

Let $v_1 \in V$ with $v_1 \neq 0$. Since $\varphi$ is nondegenerate there is a $w \in V$ so that $\varphi(w, v_1) \neq 0$. Since $\varphi$ is skew-symmetric $w$ cannot be a multiple of $v_1$ and so $\dim \text{span}_F\{v_1, w\} = 2$. Set $a = \varphi(w, v_1)$ and $v_2 = w/a$. Set $B_1 = \{v_1, v_2\}$ and $V_1 = \text{span}_F\{v_1, v_2\}$. Then we have

$$
[\varphi|_{V_1}]_{B_1} = \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix}.
$$

Now suppose the theorem is true for all vector spaces of dimension less than $n$. Let $v_1 \in V$ with $v_1 \neq 0$. As in the $n = 2$ case we construct a $v_2$ and $V_1 = \text{span}_F\{v_1, v_2\}$ so that

$$
[\varphi|_{V_1}]_{B_1} = \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix}.
$$

This shows $\varphi|_{V_1}$ is nondegenerate, so $V = V_1 \perp V_1^\perp$. We now apply the induction hypothesis to $V_1^\perp$. Suppose that $n$ is odd and $\varphi$ is a nondegenerate Hermitian form on $V$. Then $\varphi|_{V_1^\perp}$ is a nondegenerate Hermitian form on a vector space of dimension $n - 2$. Since $n - 2$ is odd, the induction hypothesis gives there are no nondegenerate Hermitian forms on a vector space of dimension $n - 2$ if it is odd. This contradiction shows there are no nondegenerate Hermitian forms on $V$ if $n$ is odd. Now suppose $n$ is even, then since $\dim_F V_1^\perp = n - 2$ is even we have a basis $B_2 = \{v_3, \ldots, v_n\}$ of $V_1^\perp$ with

$$
[\varphi|_{V_1^\perp}]_{B_2} = \begin{pmatrix}
0 & 1 & & \\
-1 & 0 & & \\
& & \ddots & \\
& & & 0 & 1 \\
& & & -1 & 0
\end{pmatrix}.
$$
where there are \((n - 2)/2\) copies of \(
abla\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\) on the diagonal. The basis \(\mathcal{B} = \{v_1, v_2, v_3, \ldots, v_n\}\) now gives the result.

**Exercise 5.2.32.** One often sees the isometry group for a skew-symmetric bilinear form represented by a different matrix. Namely, let \((V, \varphi)\) be an \(F\)-vector space of dimension \(2n\) with \(\varphi\) a skew-symmetric bilinear form.

1. Show that there is a basis \(\mathcal{B}_1\) so that

\[
[\varphi]_{\mathcal{B}_1} = \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix}.
\]

2. Show that there is a basis \(\mathcal{B}_2\) so that

\[
[\varphi]_{\mathcal{B}_2} = \begin{pmatrix} \overbrace{\ddots}^{n-1} & 1 \\ -1 & \ddots \\ \vdots & -1 \end{pmatrix}
\]

Using the previous exercise, we can write

\[
\text{Sp}_{2n}(F) = \left\{ M \in \text{GL}_{2n}(F) : {^t}M \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix} M = \begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix} \right\}.
\]

The last case for us to classify is the case of skew-Hermitian forms. Note that one can obtain the conclusion below simply by using the fact that if \(\varphi\) is skew-symmetric, then \(j\varphi\) is Hermitian and then applying the earlier results on Hermitian forms. However, we give a direct proof here to illustrate the methods one last time.

**Theorem 5.2.33.** Let \((V, \varphi)\) be an \(F\)-vector space with \(\varphi\) a nondegenerate skew-Hermitian form. Then \(\varphi\) is isometric to \([a_1] \perp \cdots \perp [a_n]\) for some \(a_i \in F\) with \(a_i \neq 0\), \(\overline{a_i} = -a_i\). (Note that \(\overline{a_i} = -a_i\) is equivalent to \(a_i = jb_i\) for some \(b_i \in F_0\).)

**Proof.** Our first step is to show there is a \(v \in V\) so that \(\varphi(v, v) \neq 0\). Let \(v_1 \in V\) be any nonzero element. If \(\varphi(v_1, v_1) \neq 0\), set \(v = v_1\). Otherwise, choose \(v_2 \in V\) so that \(\varphi(v_2, v_1) \neq 0\). Such a \(v_2\) exists because \(\varphi\) is nondegenerate. Set \(a = \varphi(v_1, v_2)\). If \(\varphi(v_2, v_2) \neq 0\), set \(v = v_2\). Otherwise, for any \(c \in F\) set \(v_3 = v_1 + cv_2\). We have

\[
\varphi(v_3, v_3) = \varphi(v_1, v_1) + \varphi(v_1, cv_2) + \varphi(cv_1, v_1) + \varphi(cv_2, cv_2) = c\varphi(v_1, v_2) + \overline{c}\varphi(v_2, v_1) = ac - \overline{ac}.
\]
5.3. THE ADJOINT MAP

Set \( c = j/a \). Then we have
\[
\varphi(v_3, v_3) = j - \overline{j} = 2j \neq 0.
\]

We now proceed by induction on the dimension of \( V \). If \( \dim_F V = 1 \), we construct \( v_3 \) as above and we have \( \varphi \) is isometric to \([2j]\). Now assume the result is true for all vector spaces of dimension less than \( n \). Let \( v_1 \in V \) be a vector so that \( \varphi(v_1, v_1) \neq 0 \) as above. Set \( V_1 = \text{span}_F \{ v_1 \} \). We have \( \varphi|_{V_1 \times V_1} \neq 0 \), so \( \varphi|_{V_1 \times V_1} \) is nondegenerate so we have \( V = V_1 \perp V_1^\perp \). We now apply the induction hypothesis to \( V_1^\perp \). This gives the result.

Exercise 5.2.34. Let \( (V, \varphi) \) be an \( \mathbb{C} \)-vector space and \( \varphi \) a nondegenerate skew-Hermitian form. Then \( \varphi \) is isometric to \([p[i] \perp q[-i]]\) for some well-defined integers \( p \) and \( q \) with \( p + q = \dim_{\mathbb{C}} V \).

5.3 The adjoint map

The last topic of this chapter is the adjoint map. Let \( (V, \varphi) \) and \( (W, \psi) \) be \( F \)-vector spaces with \( \varphi \) and \( \psi \) either both non-degenerate bilinear or both non-degenerate sesquilinear forms. (Throughout this section when given \( \varphi \) and \( \psi \), we always assume they are non-degenerate and are both bilinear or both sesquilinear.) Let \( T \in \text{Hom}_F(V, W) \). The adjoint map is a linear map \( T^* \in \text{Hom}_F(W, V) \) that behaves nicely with respect to \( \varphi \) and \( \psi \).

Definition 5.3.1. Let \( T \in \text{Hom}_F(V, W) \). The adjoint map is a map \( T^*: W \to V \) satisfying
\[
\varphi(v, T^*(w)) = \psi(T(v), w)
\]
for all \( v \in V, w \in W \).

It is important to note here that many books will use \( T^* \) to denote the dual map \( T^\vee \). Be careful not to confuse the adjoint \( T^* \) with the dual map! We will see why books use this notation below.

The first step is to show that an adjoint map actually exists when \( V \) and \( W \) are finite dimensional.

Proposition 5.3.2. Let \( (V, \varphi) \) and \( (W, \psi) \) be finite dimensional vector spaces with \( \varphi \) and \( \psi \) bilinear forms. Then there is an adjoint map \( T^* \in \text{Hom}_F(W, V) \).

Proof. We begin by showing there is a map \( T^* \) that satisfies the equation. We then show it is unique and linear. Recall that given any \( \Phi \in W^\vee \), we have \( T^\vee \circ \Phi \in V^\vee \) defined by \( T^\vee \circ \Phi(v) = \Phi(T(v)) \) for \( v \in V \). Fix \( w \in W \) so we have \( \varphi(\cdot, w) = R_\varphi(w) \in W^\vee \). Then \( T^\vee \circ R_\varphi(w) \in V^\vee \) and is given by \( (T^\vee \circ R_\varphi(w))(v) = R_\varphi(w)(T(v)) = \psi(T(v), w) \). We now use that \( \varphi \) is non-degenerate, i.e., \( R_\varphi : V \to V^\vee \) is an isomorphism to conclude that for each \( w \in W \) there exists a unique element \( z_w \in V \) so that \( R_\varphi(z_w) = T^\vee \circ R_\varphi(w) \).
5.3. THE ADJOINT MAP CHAPTER 5.

i.e., for each \( w \in W \) and \( v \in V \) we have

\[
\varphi(v, z_w) = R_\varphi(z_w)(v) \\
= (T^\vee \circ R_\varphi(w))(v) \\
= R_\psi(w)(T(v)) \\
= \psi(T(v), w).
\]

Thus, we have a map \( T^* : W \to V \) defined by sending \( w \) to \( z_w \). Moreover, from the construction it is clear the map is unique.

One could easily show this map is linear directly, but we present a different proof that is conceptually more useful for later results. Suppose that \( \varphi \) and \( \psi \) are bilinear. We just saw the defining formula for \( T^* \) can be rewritten as

\[
R_\varphi(T^*(w))(v) = T^\vee(R_\psi(w))(v).
\]

Thus, \( R_\varphi \circ T^* = T^\vee \circ R_\psi \). Since \( \varphi \) is assumed to be non-degenerate, \( R_\varphi \) is an isomorphism so \( R_\varphi^{-1} \) exists and we can write

\[
T^* = R_\varphi^{-1} \circ T^\vee \circ R_\psi.
\]

Since each term on the right hand side is linear, so is \( T^* \). This finishes the proof of the result in the case the forms are bilinear. \( \square \)

We immediately obtain the following version in coordinates.

**Corollary 5.3.3.** Let \( B \) be a basis of \((V, \varphi)\) and \( C \) be a basis of \((W, \psi)\). Set \( P = [\varphi]_B \) and \( Q = [\psi]_C \). Then

\[
[T^*]_C^B = P^{-1} [T]_B^C Q
\]

if \( \varphi \) and \( \psi \) are bilinear.

One now needs the same results for sesquilinear forms. If one proceeds as above the map \( T^* \) that one constructs is a map from \( \overline{W} \) to \( \overline{V} \); the claim is there is a map from \( W \) to \( V \). We remedy this by introducing the conjugate of a linear map. As this is much easier to see in coordinates, we do that case first. Suppose we are given a linear map \( T : \overline{V} \to \overline{W} \). Let \( B = \{v_1, \ldots, v_n\} \) be a basis of \( V \) and \( C = \{w_1, \ldots, w_m\} \) be a basis of \( W \). (Note these are also bases of \( V \) and \( W \).) To obtain the matrix of \( T \) with respect to \( B \) and \( C \) we write

\[
T(v_i) = a_{i1} \cdot w_1 + \cdots + a_{im} \cdot w_m.
\]

Thus, the matrix of \( T : \overline{V} \to \overline{W} \) is given by \( A = (a_{ij}) \in \text{Mat}_{m,n}(F) \) where we view the matrices as acting on \( F^n \). However, if we wish to view this as a linear map from \( V \) to \( W \) we have

\[
T(v_i) = \overline{a_{i1}} w_1 + \cdots + \overline{a_{im}} w_m,
\]

so the correct matrix in this case is \( \overline{A} = (\overline{a_{ij}}) \in \text{Mat}_{m,n}(F) \) where we view the matrix as acting on \( F^n \). Thus, given a linear transformation \( T : \overline{V} \to \overline{W} \)
with associated matrix $A = [T]_B^C$, we obtain a linear transformation $\overline{T}$ given in coordinates by $[\overline{T}]_B^C = \overline{A}$. We can now remedy the above situation by setting $T^* = R_{\overline{\varphi}}^{-1} \circ T^\bigcirc \circ R_{\overline{\psi}}$. This finishes the proof of Proposition 5.3.2 and gives the following result.

**Corollary 5.3.4.** Let $B$ be a basis of $(V, \varphi)$ and $C$ be a basis of $(W, \psi)$. Set $P = [\varphi]_B$ and $Q = [\psi]_C$. Then

$$[T^*]_B^C = P^{-1} \overline{[T]_B^C} Q$$

if $\varphi$ and $\psi$ are sesquilinear.

Note the above proof fails completely when $V$ and $W$ are not finite dimensional vector spaces. One no longer has that $R_{\varphi}$ and $R_{\psi}$ are isomorphisms. In fact, in the infinite dimensional case it often happens that the adjoint map does not exist! One particularly important case where one knows adjoints exist in the infinite dimensional case is for bounded linear maps on Hilbert spaces. This follows from the Riesz Representation Theorem. We do not expand on this here as it takes us too far afield.

One special case of the above corollaries is when $V = W$ and $\varphi = \psi$. In this case one has $P = Q$. We can take this further with the following definition and result. These concepts will be developed more fully in the next chapter.

**Definition 5.3.5.** Let $(V, \varphi)$ be a vector space. A basis $B = \{v_i\}$ is said to be orthogonal if $\varphi(v_i, v_j) = 0$ for all $v_i, v_j \in B$ with $i \neq j$. We say the basis is orthonormal if it is orthogonal and satisfies $\varphi(v_i, v_i) = 1$ for all $v_i \in B$.

In general there is no reason an arbitrary vector space should have an orthonormal basis. However, if one has an orthonormal basis one obtains some very nice properties. One immediate property is the following corollary.

**Corollary 5.3.6.** Let $(V, \varphi)$ and $(W, \psi)$ be finite dimensional vector spaces with orthonormal bases $B$ and $C$ respectively. Let $T \in \text{Hom}_F(V, W)$. Then

$$[T^*]_B^C = [T]_B^C$$

if $\varphi$ and $\psi$ are bilinear and

$$[T^*]_B^C = \overline{[T]_B^C}$$

if $\varphi$ and $\psi$ are sesquilinear.

**Proof.** This follows from Proposition 5.3.2 upon observing that $P$ and $Q$ are both the identity matrix due to the fact that $B$ and $C$ are orthonormal. \[\square\]

Suppose we have a finite dimensional vector space $(V, \varphi)$ with $\varphi$ a non-degenerate symmetric bilinear form. Suppose there are orthonormal bases $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_n\}$ of $V$. Let $T \in \text{Hom}_F(V, W)$ be defined by $T(v_i) = w_i$. Observe we have

$$\varphi(v_i, v_j) = \delta_{i,j}$$

$$= \varphi(w_i, w_j)$$

$$= \varphi(T(v_i), T(v_j))$$

122
where \( \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \). We immediately obtain that \( \varphi(v, \tilde{v}) = \varphi(T(v), T(\tilde{v})) \) for all \( v, \tilde{v} \in V \). In particular, we have \( \varphi(v, \tilde{v}) = \varphi(v, T^*(T(\tilde{v}))) \) for all \( v, w \in V \), i.e., \( \varphi(v, \tilde{v} - T^*(T(\tilde{v}))) = 0 \) for all \( v, \tilde{v} \in V \). Since \( \varphi \) is non-degenerate, this gives \( T^* \circ T = \text{id}_V \), i.e., if we set \( P = [T]_B^C \), then \( P \circ P = \text{id}_V \). Thus, if \( P \) is a change of basis matrix between two orthonormal bases then \( P \) is an orthogonal matrix. The same argument gives that if \( \varphi \) is a Hermitian form, then a change of basis matrix between two orthonormal bases is a unitary matrix.

**Exercise 5.3.7.** Let \( B = \{ v_1, \ldots, v_n \} \) be an orthonormal basis of a finite dimensional vector space \( (V, \varphi) \) with \( \varphi \) a non-degenerate symmetric bilinear form. Let \( T \in \text{Hom}_F(V, V) \) and let \( P = [T]_B^C \). Show that if \( P \) is an orthogonal matrix then \( C = \{ T(v_1), \ldots, T(v_n) \} \) is an orthonormal basis of \( V \). Prove the analogous result when \( \varphi \) is Hermitian as well.

**Exercise 5.3.8.** Let \( (U, \phi) \), \( (V, \varphi) \), and \( (W, \psi) \) be finite dimensional vector spaces. Prove the following elementary results about the adjoint map.

1. \( S, T \in \text{Hom}_F(V, W) \). Show that \((S + T)^* = S^* + T^* \).
2. \((cT)^* = cT^* \) for every \( c \in F \).
3. Let \( T \in \text{Hom}_F(U, V) \) and \( S \in \text{Hom}_F(V, W) \). Then \((S \circ T)^* = T^* \circ S^* \).
4. Let \( p(x) = a_n x^n + \cdots + a_1 x + a_0 \in F[x] \) and \( T \in \text{Hom}_F(V, W) \). Then \((p(T))^* = \overline{p}(T^*) \) where \( \overline{p}(x) = \overline{a_n} x^n + \cdots + \overline{a_1} x + \overline{a_0} \).

**Corollary 5.3.9.** Let \( (V, \varphi) \) and \( (W, \psi) \) be vector spaces with \( \varphi \) and \( \psi \) both symmetric, skew-symmetric, Hermitian, or skew-Hermitian. Then \((T^*)^* = T \).

**Proof.** We prove the Hermitian case and leave the others as exercises as the proofs are analogous. Set \( S = T^* \) so that

\[
\varphi(v, S(w)) = \psi(T(v), w)
\]

for all \( v \in V, w \in W \). We have

\[
\varphi(S(w), v) = \overline{\varphi(v, S(w))} \\
= \overline{\psi(T(v), w)} \\
= \psi(w, T(v))
\]

for all \( v \in V, w \in W \). This shows we must have \( T = S^* \) since the adjoint is unique and \( T \) satisfies the properties of being the adjoint of \( S \).

Note the above proof works for infinite dimensional spaces as well with the added assumption that the adjoint exists.
5.4 Problems

For these problems we assume $V$ is a finite dimensional vector space over a field $F$.

1. Let $V_1$ and $V_2$ be subspaces of $V$. Show that $V = V_1 \perp V_2$ if
   (i) $V = V_1 \oplus V_2$ and
   (ii) given any $v, v' \in V$, when we write $v = v_1 + v_2$ and $v' = v'_1 + v'_2$ for $v_i, v'_i \in V_i$ we have
       \[
       \varphi(v, v') = \varphi_1(v_1, v'_1) + \varphi_2(v_2, v'_2)
       \]
   where $\varphi_i = \varphi|_{V_i}$.

2. Let $\varphi$ be a bilinear form on $V$ and assume $\text{char}(F) \neq 2$. Prove that $B$ is skew-symmetric if and only if the diagonal function $V \to F$ given by $v \mapsto \varphi(v, v)$ is additive.

3. Let $D$ be an integer so that $\sqrt{D} \notin \mathbb{Z}$. Let $V = \mathbb{R}^2$.
   (a) Show that $\varphi_D(x, y) = \varphi \left( \begin{pmatrix} 1 & 0 \\ 0 & D \end{pmatrix} \right) y$ is a bilinear form on $V$.
   (b) Given two such integers $D_1, D_2$, give necessary and sufficient conditions for $\varphi_{D_1}$ to be isometric to $\varphi_{D_2}$?
   (c) Suppose now that $V = \mathbb{Q}^2$. Under what condition is $\varphi_{D_1}$ isometric to $\varphi_{D_2}$?

4. Let $V = \mathbb{R}^2$. Set $\varphi((x_1, y_1), (x_2, y_2)) = x_1 x_2$.
   (a) Show this is a bilinear form. Give a matrix representing this form. Is this form nondegenerate?
   (b) Let $W = \text{span}_\mathbb{R}(e_1)$ where $e_1$ is the standard basis element. Show that $V = W \perp W^\perp$.
   (c) Calculate $(W^\perp)^\perp$.

5. (a) Let $A \in \text{Mat}_n(\mathbb{C})$. Prove that there is a unique Hermitian matrix $H \in \text{Mat}_n(\mathbb{C})$ and a unique skew-Hermitian matrix $S \in \text{Mat}_n(\mathbb{C})$ so that $A = H + S$. This is referred to as the Hermitian decomposition of $A$.
   (b) Find the Hermitian decomposition of $A = \begin{pmatrix} 2i & 1 + 3i \\ -1 + 3i & -5i \end{pmatrix}$.
6. Let \( A \in \text{Mat}_n(\mathbb{C}) \) be a Hermitian matrix. Prove all the eigenvalues of \( A \) must be real.

7. Let \( V \) be a 3-dimensional \( \mathbb{Q} \)-vector space with basis \( B = \{v_1, v_2, v_3\} \). Define a symmetric bilinear form on \( V \) by setting \( \varphi(v_1, v_1) = 0, \varphi(v_1, v_2) = -2, \varphi(v_1, v_3) = 2, \varphi(v_2, v_2) = 2, \varphi(v_2, v_3) = -2, \varphi(v_3, v_3) = 3 \).

(a) Give the matrix \([\varphi]_B\).

(b) If possible, find a basis \( B' \) so that \([\varphi]_{B'}\) is diagonal and give \([\varphi]_{B'}\). If it is not possible to diagonalize \( \varphi \), give reasons.

(c) Is the symmetric bilinear form given in (b) isometric to the symmetric bilinear form given by
\[
\begin{pmatrix}
-2 & 0 & 0 \\
0 & 8 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]?

8. Let \( A \in \text{U}_n(\mathbb{C}) \). Prove all the eigenvalues of \( A \) must lie on the unit circle in \( \mathbb{C} \).

9. Let \( A \in \text{O}_2(\mathbb{R}) \). Prove that the first row of \( A \) has the form \((\cos \theta, \sin \theta)\) for some \( \theta \in [0, 2\pi) \). Given this first row, what are all possible second rows of \( A \)?

10. (a) Let \( V = \mathbb{F}_3^3 \) and let \( E_3 \) be the standard basis for \( V \). Let \( \varphi : V \to V \) be the symmetric bilinear form given by
\[
[\varphi]_{E_3} = \begin{pmatrix}
1 & 1 & 2 \\
1 & 3 & 1 \\
2 & 1 & 4
\end{pmatrix}.
\]
Find an orthogonal basis for \( V \) with respect to \( \varphi \). Can you make this an orthonormal basis?

(b) Give a finite collection of matrices so that every nondegenerate symmetric bilinear form on \( V \) is isometric to one of the forms listed. Give a short justification on why your list is complete.

11. Let \( \varphi \) be the standard inner product on \( \mathbb{R}^n \) from multivariable calculus class. Define \( T \in \text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n) \) by
\[
T(x_1, \ldots, x_n) = (0, x_1, \ldots, x_{n-1}).
\]
Find a formula for the adjoint map \( T^* \).

12. Let \( T \in \text{Hom}_F(V, W) \). Show that \( T \) is injective if and only if \( T^* \) is surjective.
Chapter 6

Inner product spaces and the spectral theorem

We now specialize some of the results of the previous chapter to situations where more can be said. We will require all our vector spaces \((V, \varphi)\) to satisfy that \(\varphi\) is positive definite. In particular, we will restrict to the case that \(\mathbb{Q} \subset F \subset \mathbb{R}\) in the case that \(\varphi\) is a symmetric bilinear form and \(F \subset \mathbb{C}\) if \(\varphi\) is Hermitian. One can restrict to \(F = \mathbb{R}\) or \(F = \mathbb{C}\) in this chapter and not much will be lost unless one really feels the need to work over other fields.

In this chapter we will recover some of the things one sees in undergraduate linear algebra such as the Gram-Schmidt process. We will end with the Spectral Theorem, which tells when one can choose a “nice” basis of a linear map that is also “nice” with respect to the bilinear form on the vector space.

6.1 Inner product spaces

We begin by defining the spaces that we will work with in this chapter.

**Definition 6.1.1.**

1. Let \((V, \varphi)\) be a vector space with \(\varphi\) a positive definite symmetric bilinear form. We say \(\varphi\) is an inner product on \(V\) and say \((V, \varphi)\) is an inner product space.

2. Let \((V, \varphi)\) be a vector space with \(\varphi\) a positive definite Hermitian form. We say \(\varphi\) is an inner product on \(V\) and say \((V, \varphi)\) is an inner product space.

Some of the most familiar examples given in the last chapter turn out to be inner products.

**Example 6.1.2.** Let \(A \in \text{Mat}_n(\mathbb{R})\) with \(^tA = A\) and \(A\) positive definite or \(A \in \text{Mat}_n(\mathbb{C})\) with \(^tA = \overline{A}\) and \(A\) positive definite. Then the form \(\varphi_A\) is an inner product. Note this example recovers the usual dot product on \(\mathbb{R}^n\) by setting \(A = 1_n\).
Example 6.1.3. Let \( V = C^0([0,1], \mathbb{R}) \) be the space of continuous functions from \([0,1]\) to \(\mathbb{R}\). Define
\[
\varphi(f,g) = \int_0^1 f(x)g(x)dx.
\]
We have seen before this is a symmetric bilinear form. Moreover, note that
\[
\varphi(f,f) = \int_0^1 f(x)^2dx > 0
\]
for all \( f \in V \) with \( f \neq 0 \). Thus, \( \varphi \) is an inner product.

Example 6.1.4. Let \( V = C^0([0,1], \mathbb{C}) \) be the vector space of continuous functions from \([0,1]\) to \(\mathbb{C}\). Define
\[
\varphi(f,g) = \int_0^1 f(x)\overline{g(x)}dx.
\]
As in the previous example, one has this is an inner product.

Exercise 6.1.5. Let \((V, \varphi)\) be an inner product space. Show that for any subspace \( W \subset V \) one has \( \varphi|_W \) is nondegenerate. Conversely, if \( \varphi \) is a form on \( V \) so that \( \varphi|_W \) is nondegenerate for each subspace \( W \subset V \), then either \( \varphi \) or \( -\varphi \) is an inner product on \( V \).

One of the nicest things about an inner product is that it allows us to define a norm on the vector space, i.e., we can define a notion of distance on \( V \).

Definition 6.1.6. Let \((V, \varphi)\) be an inner product space. The norm of a vector \( v \in V \), denoted \( ||v|| \), is defined by
\[
||v|| = \sqrt{\varphi(v,v)}.
\]
The previous definition recovers the definition of the length of a vector given in multivariable calculus by considering \( V = \mathbb{R}^n \) and \( \varphi \) the usual dot product.

The following lemma gives some of the basic properties of norms. These are all familiar properties from calculus class for the norm of a vector.

Lemma 6.1.7. Let \((V, \varphi)\) be an inner product space.

1. We have \( ||cv|| = |c||v| \) for all \( c \in F, \ v \in V \).
2. We have \( ||v|| \geq 0 \) and \( ||v|| = 0 \) if and only if \( v = 0 \).
3. (Cauchy-Schwartz inequality) We have \( |\varphi(v,w)| \leq ||v||||w|| \). We have equality if and only if \( \{v,w\} \) is linearly dependent.
4. (Triangle inequality) We have \( ||v + w|| \leq ||v|| + ||w|| \) for all \( v, w \in V \). This is an equality if and only if \( w = 0 \) or \( v = cw \) for \( c \in F_{\geq 0} \) where \( F_{\geq 0} = F \cap \mathbb{R}_{\geq 0} \).
Proof. The proofs of the first two statements are left as exercises. We now prove the Cauchy-Schwartz inequality. If \( \{v, w\} \) is linearly dependent, then up to reordering either \( w = 0 \) or \( w = cv \) for some \( c \in F \). In either case it is clear we have equality. Now assume \( \{v, w\} \) is linearly independent, i.e., for any \( c \in F \) we have \( u = v - cw \neq 0 \). Observe we have
\[
0 \leq ||u||^2 = \varphi(u, u) = ||v||^2 + \varphi(-cw, v) + \varphi(v, -cw) + |c|^2||w||^2
\]
for any \( c \in F \). Set \( c = \frac{\varphi(v, w)}{\varphi(w, w)} \). This gives
\[
0 \leq ||v||^2 - \frac{\varphi(v, w)\varphi(v, w)}{||w||^2} - \frac{\varphi(v, w)\varphi(v, w)}{||w||^2} + \frac{\varphi(v, w)^2}{\varphi(w, w)} ||w||^2
\]
i.e., \( 0 \leq ||v||^2 - \frac{|\varphi(v, w)||^2}{||w||^2} \). Thus, \( |\varphi(v, w)| \leq ||v||||w|| \) as claimed. Moreover, note that we get a strict inequality here since \( \{v, w\} \) is linearly independent, we cannot have equality.

We now turn our attention to the triangle inequality. Observe for any \( v, w \in V \) we have
\[
||v + w||^2 = \varphi(v + w, v + w)
= ||v||^2 + \varphi(v, v) + \varphi(w, v) + ||w||^2
= ||v||^2 + \varphi(v, w) + \overline{\varphi(v, w)} + ||w||^2
= ||v||^2 + 2\Re(\varphi(v, w)) + ||w||^2
\leq ||v||^2 + 2|\varphi(v, w)| + ||w||^2
\leq ||v||^2 + 2||v|| ||w|| + ||w||^2
= (||v|| + ||w||)^2.
\]
This gives the triangle inequality. It remains to deal with the case when \( ||v + w||^2 = (||v|| + ||w||)^2 \). If this is the case, then both inequalities above must be equalities. The second inequality is equality if and only if \( w = 0 \) (note the first inequality is also an equality if \( w = 0 \)) or if \( w \neq 0 \) and \( w = cv \) for some \( c \in F \). We have
\[
\varphi(v, w) + \varphi(w, v)\varphi(cw, w) + \varphi(w, cw) = (c + \pi)||w||^2.
\]
Thus the first inequality is an equality if and only if \( c \in F_{\geq 0} \).

The fact that we are restricting ourselves to characteristic 0 fields and \( \varphi \) being symmetric bilinear form or a Hermitian form immediately gives that there is a basis of \( V \) consisting of orthogonal vectors via Theorem 5.2.23. Moreover, if we take \( F \) to be \( \mathbb{R} \) or \( \mathbb{C} \) then we can combine the facts that the form is positive definite with Sylvester’s Law of Inertia to conclude that \( V \) has an orthonormal
basis as well. We now give a few easy lemmas that deal with orthogonality of vectors.

**Lemma 6.1.8.** Let $\mathcal{B} = \{v_i\}$ be an orthogonal set of nonzero vectors in an inner product space $(V, \varphi)$. If $v \in V$ can be written as $v = \sum_j c_j v_j$ for $c_j \in F$, then $c_j = \frac{\varphi(v, v_j)}{||v_j||^2}$. If $\mathcal{B}$ is orthonormal, then $c_j = \varphi(v, v_j)$.

*Proof.* Observe we have

$$
\varphi(v, v_j) = \varphi\left(\sum_i c_i v_i, v_j\right) = \sum_i c_i \varphi(v_i, v_j) = c_j \varphi(v_j, v_j).
$$

This gives $c_j = \frac{\varphi(v, v_j)}{\varphi(v_j, v_j)}$. If $\mathcal{B}$ happens to be orthonormal then $\varphi(v_j, v_j) = 1$.

We can use the previous result to talk about projection of vectors onto subspaces. Let $W \subset V$ be a subspace and let $v$ be any vector in $V$. We can project $v$ onto a unique vector $v_W$ in $W$ and onto a unique vector $v_{W^\perp}$ in $W^\perp$ so that $v = v_W + v_{W^\perp}$. To see this we just use that $V = W \perp W^\perp$. We can use the previous result to write down explicitly what $v_W$ and $v_{W^\perp}$ are in terms of an orthogonal basis. Let $\mathcal{B}_W = \{v_1, \ldots, v_m\}$ be an orthogonal basis of $W$ and $\mathcal{B} = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$ be an orthogonal basis of $V$. We have

$$
v = \sum_{j=1}^m \frac{\varphi(v, v_j)}{\varphi(v_j, v_j)} v_j + \sum_{j=m+1}^n \frac{\varphi(v, v_j)}{\varphi(v_j, v_j)} v_j.
$$

It is now clear that if we set

$$
v_W = \sum_{j=1}^m \frac{\varphi(v, v_j)}{\varphi(v_j, v_j)} v_j
$$

and

$$
v_{W^\perp} = \sum_{j=m+1}^n \frac{\varphi(v, v_j)}{\varphi(v_j, v_j)} v_j,
$$

then we have the claim. In this case one usually writes $\text{proj}_W v$ for $v_W$ and $\text{proj}_{W^\perp} v$ for $v_{W^\perp}$. If one takes $V = \mathbb{R}^n$ and $\varphi$ to be the usual dot product, this recovers vector projection studied in multivariable calculus.

**Lemma 6.1.9.** Let $W \subset V$ be a subspace and $v \in V$. Then for all $w \in W$ with $w \neq \text{proj}_W v$ we have

$$
||v - \text{proj}_W v|| < ||v - w||.
$$

129
Proof. Observe we have
\[ v - w = (v - \text{proj}_W v) + (\text{proj}_W v - w). \]
We have \( \text{proj}_W v - w \in W \) and \( (v - \text{proj}_W v) \in W^\perp \). Thus, the triangle inequality gives
\[ ||v - w||^2 = ||v - \text{proj}_W v||^2 + ||\text{proj}_W v - w||^2. \]
Since \( w \neq \text{proj}_W v \), we have \( ||\text{proj}_W v - w||^2 > 0 \), which gives the result. \( \square \)

Corollary 6.1.10. Let \( \mathcal{B} = \{v_i\} \) be a set of nonzero orthogonal vectors in an inner product space \((V, \varphi)\). Then \( \mathcal{B} \) is linearly independent.

Proof. Suppose there exists \( c_j \in F \) so that \( 0 = \sum_j c_j v_j \). The previous result gives \( c_j = \frac{\varphi(0, v_j)}{||v_j||^2} = 0 \). Thus, we have \( \mathcal{B} \) is linearly independent. \( \square \)

Lemma 6.1.11. Let \( \mathcal{B} = \{v_i\} \) be a set of nonzero orthogonal vectors in an inner product space \((V, \varphi)\). Let \( v \in V \) and suppose \( v \) can be written as \( v = \sum_j c_j v_j \). Then
\[ ||v||^2 = \sum_j |c_j|^2 ||v_j||^2. \]
If \( \mathcal{B} \) is orthonormal, then
\[ ||v||^2 = \sum_j |c_j|^2. \]

Proof. We have
\[ ||v||^2 = \varphi(v, v) \]
\[ = \varphi \left( \sum_i c_i v_i, \sum_j c_j v_j \right) \]
\[ = \sum_{i,j} c_i c_j \varphi(v_i, v_j) \]
\[ = \sum_j |c_j|^2 \varphi(v_j, v_j) \]
\[ = \sum_j |c_j|^2 ||v_j||^2. \]
This gives the result for \( \mathcal{B} \) orthogonal. The orthonormal case follows immediately upon observing \( ||v_j|| = 1 \) for all \( j \). \( \square \)

Corollary 6.1.12. Let \( \mathcal{B} = \{v_1, \ldots, v_n\} \) be a set of nonzero orthogonal vectors in \( V \). Then for any vector \( v \in V \) we have
\[ \sum_{j=1}^n \frac{\varphi(v, v)}{||v_j||^2} \leq ||v||^2 \]
with equality if and only if \( v = \sum_{j=1}^n \frac{\varphi(v, v_j)}{||v_j||^2} v_j \).
Proof. Set \( w = \sum_{j=1}^{n} \left( \frac{\varphi(v, v_j)}{||v_j||^2} \right) v_j \) and \( u = v - w \). Note,

\[
\varphi(w, v_i) = \sum_{j=1}^{n} \left( \frac{\varphi(v, v_j)}{||v_j||^2} \right) \varphi(v_j, v_i) = \varphi(v, v_i),
\]

and so \( \varphi(u, v_i) = \varphi(v, v_i) - \varphi(w, v_i) = 0 \) for \( i = 1, \ldots, n \). Thus,

\[
||v||^2 = \varphi(v, v) = \varphi(u + w, u + w) = \varphi(u, u) + \varphi(u, w) + \varphi(w, u) + \varphi(w, w) = ||u||^2 + ||w||^2
\]

since \( \varphi(u, w) = 0 \) as \( \varphi(u, v_i) = 0 \) for all \( i \). Hence,

\[
\sum_{j=1}^{n} \frac{||\varphi(v, v_j)||^2}{||v_j||^2} = ||w||^2 \\
\leq ||v||^2,
\]

as claimed. We obtain equality if and only if \( u = 0 \), i.e., \( v \) is of the form

\[
\sum_{j=1}^{n} \left( \frac{\varphi(v, v_j)}{||v_j||^2} \right) v_j.
\]

Our next step is to prove the traditional Gram-Schmidt process from undergraduate linear algebra. Upon completing this, we give a brief outline of how one can do this in a more general setting.

**Theorem 6.1.13.** (Gram-Schmidt) Let \( W \) be a finite dimensional subspace of \( V \) with \( \dim_F W = k \). Let \( B = \{v_1, \ldots, v_k\} \) be a basis of \( W \). Then there is an orthogonal basis \( C = \{w_1, \ldots, w_k\} \) of \( W \) so that \( \text{span}_F \{v_1, \ldots, v_l\} = \text{span}_F \{w_1, \ldots, w_l\} \) for \( l = 1, \ldots, k \). Moreover, if \( F \) contains \( \mathbb{R} \) we can choose \( C \) to be orthonormal.

**Proof.** The fact that an orthogonal basis exists follows Theorem 5.2.23. If \( \mathbb{R} \subset F \) we obtain an orthonormal basis via Sylvester’s Law of Inertia and the fact that \( \varphi \) is positive definite. To construct the basis we work inductively. Set \( x_1 = v_1 \) and for \( j \geq 1 \) we set

\[
x_i = v_i - \sum_{j<i} \frac{\varphi(v_i, v_j)}{||x_j||^2} x_j.
\]

One can easily check this gives an orthogonal basis that satisfies \( \text{span}_F \{v_1, \ldots, v_l\} = \text{span}_F \{x_1, \ldots, x_l\} \) for \( l = 1, \ldots, k \). If \( \mathbb{R} \subset F \), we set \( w_j = x_j/||x_j|| \) and obtain the result. \( \square \)
Let \((V, \varphi)\) be a real vector space with \(\varphi\) symmetric. (We do not require that \(\varphi\) be positive definite here!) Let \(\mathcal{B} = \{v_1, \ldots, v_r\}\) be a basis of \(\ker \varphi\). We can write \(V = \ker \varphi \perp W\) and \(\varphi|_W\) is non-degenerate. Let \(W^+\) be the maximal subspace so that \(\varphi|_{W^+}\) is positive definite and \(W^-\) the maximal subspace so that \(\varphi|_{W^-}\) is negative definite. Then we have \(V = \ker \varphi \perp W^+ \perp W^-\). Since \(\varphi|_{W^+}\) is positive definite, we can apply Gram-Schmidt to get the desired basis of \(W^+\). Similarly, since \(\varphi|_{W^-}\) is negative definite, we can apply Gram-Schmidt to \(-\varphi\) to get the desired basis of \(W^-\). This gives the matrix \[
[\varphi]_C = \begin{pmatrix}
0_r & 1_p \\
-1_q & 0_q
\end{pmatrix}.
\]

Note the difficulty in this case is in calculating \(\ker \varphi\), \(W^+\), and \(W^-\). This issue is not present for positive definite forms.

**Proposition 6.1.14.** Let \((V, \varphi)\) and \((W, \psi)\) be finite dimensional inner product spaces and let \(T \in \text{Hom}_F(V, W)\). Then we have:

1. \(\ker T^* = (\text{Im } T)^\perp\)
2. \(\ker T = (\text{Im } T^*)^\perp\)
3. \(\text{Im } T^* = (\ker T)^\perp\)
4. \(\text{Im } T = (\ker T^*)^\perp\)

**Proof.**

1) Let \(w \in W\). We have \(w \in \ker T^*\) if and only if \(T^*(w) = 0\). Since \(\varphi\) is nondegenerate, \(T^*(w) = 0\) if and only if \(\varphi(v, T^*(w)) = 0\) for all \(v \in V\). However, this is equivalent to \(\psi(T(v), w) = 0\) for all \(v \in V\) by the definition of the adjoint map. Finally, this is equivalent to \(w \in (\text{Im } T)^\perp\).

2) This follows exactly as 1) upon interchanging \(T\) and \(T^*\).

3) Observe we have \(V = \ker T \perp (\ker T)^\perp = (\text{Im } T^*)^\perp \perp (\ker T)^\perp\) by 2). Moreover, we have \(V = \text{Im } T^* \perp (\text{Im } T^*)^\perp\). Combining these gives \(\text{Im } T^* = (\ker T)^\perp\).

4) This follows exactly as in 3) upon interchanging \(T\) and \(T^*\).

**Lemma 6.1.15.** Let \((V, \varphi)\) be a finite dimensional inner product space and let \(T \in \text{Hom}_F(V, V)\). Then we have

\[
\dim_F \ker T = \dim_F \ker T^*.
\]

**Proof.** The previous proposition gives that \(\text{Im } T^* = (\ker T)^\perp\). Observe we have

\[
\dim_F V = \dim_F \ker T + \dim_F (\ker T)^\perp = \dim_F \ker T + \dim_F \text{Im } T^*.
\]

On the other hand, we also have \(\dim_F V = \dim_F \ker T^* + \dim_F \text{Im } T^*\). Equating these and canceling \(\dim_F \text{Im } T^*\) gives the result.

---

132
Example 6.1.16. Suppose one is given a collection of points in $\mathbb{R}^2$, say $p_1 = (x_1, y_1), \ldots, p_n = (x_n, y_n)$. We can consider the line $y = c_1 x + c_0$ that “best fits” these points. If the points all were on the line we would have a consistent set of linear equations

$$
\begin{align*}
y_1 &= c_1 x_1 + c_0 \\
y_2 &= c_1 x_2 + c_0 \\
& \vdots \\
y_n &= c_1 x_n + c_0.
\end{align*}
$$

This can be represented in matrix form by say there is a solution to the equation $Xc = Y$ where

$$
X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \text{ and } Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.
$$

In general the points will not be collinear, so we do not have a solution to this equation. However, we can ask for the values of $c_0$ and $c_1$ that minimize the distance $||Xc - Y||$. We call the solution of this minimization problem the least square regression line $y = c_0 + c_1 x$.

We now rephrase this a bit. As $X \in \text{Mat}_{n,2}(\mathbb{R})$, we view $X$ as a linear map from $\mathbb{R}^2$ to $\mathbb{R}^n$. The goal is to find a vector $Xc \in \text{Im}(X)$ that is as close as possible to $Y$. We know from Lemma 6.1.9 that this is given by the projection of $Y$ onto the subspace $\text{Im}(X)$. Thus, $Xc = \text{proj}_{\text{Im}(X)}(Y)$. As above, we have $Xc - Y = \text{proj}_{\text{Im}(X)} Y - Y$ is orthogonal to $\text{Im}(X)$, i.e., $Xc - Y \in \text{Im}(X)^\perp = \ker(X^*) = \ker(^t X)$. Thus,

$$
0 = (^t X)(Xc - Y)
$$

i.e., we have

$$
(^t X)c = ^t Y.
$$

So finding the best fit line boils down to solving this system of linear equations.

We can generalize the previous example as follows.

Example 6.1.17. Let $A \in \text{Mat}_{m,n}(\mathbb{C})$ with $m \geq n$ and $b \in \mathbb{C}^m$. In this case one is interested in solving the equation $Ax = b$ for some $x \in \mathbb{C}^n$. Of course, in general one will not be able to find a solution so one has to do a best approximation, i.e., find the vector $x$ that minimizes the length of the error vector $Ax - b$. Assume that $A$ has rank $n$, i.e., it has $n$ linearly independent columns. We will see there is a unique $x$ that minimizes the error vector.

The analysis goes essentially exactly as in the previous example. The first thing to show is that since $A$ has rank $n$, $A^*A$ is invertible where we recall the adjoint $A^*$ is given here by the conjugate transpose of $A$. Fix $z \in \mathbb{C}^n$ and observe that since $z$ is arbitrary, if we show that $A^*Az = 0$ implies that $z = 0$, we have $A^*A$ is invertible since it is necessarily injective and maps $\mathbb{C}^n$ to $\mathbb{C}^n$. 

133
Suppose that $A^*Az = 0$. Then we have

\[
\|Az\| = \varphi(Az, Az) \\
= (Az)^*Az \\
= z^*A^*Az \\
= 0.
\]

Thus, we must have $Az = 0$. If $z$ is not 0, we can write the equation $Az = 0$ as an equation in the column vectors of $A$, giving a linear dependence among them unless $z = 0$. Thus, we have $z = 0$ and so $A^*A$ is invertible. Using this, we have $x = (A^*A)^{-1}A^*b$ as in the previous example.

It remains to show this $x$ minimizes this error vector and it is the unique such vector. This follows from what we did in the previous example if one follows the same argument and uniqueness follows from the uniqueness of the projection vector, but we give a direct proof here for clarity. Let $y \in \mathbb{C}^n$. We have

\[
\varphi(Ax - b, A(y - x)) = (y - x)^*(A^*Ax - A^*b) \\
= 0.
\]

Thus, the vectors $Ax - b$ and $A(y - x)$ are orthogonal. We have via the Pythagorean theorem

\[
\|Ay - b\|^2 = \|A(y - x) + (Ax - b)\|^2 \\
= \|A(y - x)\|^2 + \|Ax - b\|^2 \\
\geq \|Ax - b\|^2.
\]

This shows of all vectors, $x$ minimizes $\|Ax - b\|^2$. Moreover, we have equality above if and only if $\|A(y - x)\| = 0$, i.e., $x = y$. Thus, the solution is unique.

We end this section with a couple of elementary results on the minimal and characteristic polynomials as well as the Jordan canonical form of $T^*$.

**Corollary 6.1.18.** Let $(V, \varphi)$ be a finite dimensional inner product space over $F$ and $T \in \text{Hom}_F(V, V)$. Suppose $c_T(x)$ factors into linear terms over $F$, i.e., $T$ has Jordan canonical form over $F$. Then $T^*$ has Jordan canonical form over $F$ and the Jordan canonical form of $T^*$ is given by conjugating the diagonals of the Jordan canonical form of $T$.

**Proof.** Assuming the roots of $c_T(x)$ lie in $F$, the dimensions of the spaces $E_k^*(T) = \ker(T - \lambda \text{id})^k$ for $\lambda$ an eigenvalue of $T$ determine the Jordan canonical form of $T$. (Note we include the linear map $T$ in the notation to avoid confusion.) Recall from Exercise 5.3.8 that we have $((T - \lambda \text{id})^k)^* = (T^* - \overline{\lambda} \text{id})^k$. We
now apply the previous lemma to obtain
\[ \dim_F E^k_\lambda(T^*) = \dim_F \ker(T - \lambda \text{id})^k \]
\[ = \dim_F \ker((T - \lambda \text{id})^k)^* \]
\[ = \dim_F \ker(T - \lambda \text{id})^k \]
\[ = \dim_F E^k_\lambda(T). \]

Since the dimensions of these spaces are equal, we have the Jordan canonical form of \( T^* \) exists and is obtained by conjugating the diagonals of the Jordan canonical form of \( T \).

\[ \square \]

**Corollary 6.1.19.** Let \((V, \varphi)\) be a finite dimensional inner product space. Let \( T \in \text{Hom}_F(V, V) \). Then

1. \( m_{T^*}(x) = m_T(x) \);
2. \( c_{T^*}(x) = c_T(x) \).

**Proof.** Observe that Exercise 5.3.8 gives that \( m_{T^*}(x) = m_T(x) \) immediately. For the second result it is enough to work over \( \mathbb{C} \), and then one can apply the previous result. \( \square \)

## 6.2 The spectral theorem

In this final section we give one of the more important results of the course. The spectral theorem tells us when we can choose a nice basis for a linear map that is also nice with respect to the inner products, i.e., orthonormal. Unfortunately this theorem does not apply to all linear maps, so we begin by defining the linear maps we will be considering. Throughout this section we fix \((V, \varphi)\) to be an inner product space.

**Definition 6.2.1.** Let \( T \in \text{Hom}_F(V, V) \). We say \( T \) is **normal** if \( T^* \) exists and \( T \circ T^* = T^* \circ T \). We say \( T \) is **self-adjoint** if \( T^* \) exists and \( T = T^* \).

**Lemma 6.2.2.** Let \( T \in \text{Hom}_F(V, V) \). Then \( \varphi(T(v), T(w)) = \varphi(v, w) \) for all \( v, w \in V \) if and only if \( ||T(v)|| = ||v|| \) for all \( v \in V \). Furthermore, If \( ||T(v)|| = ||v|| \) for all \( v \in V \), then \( T \) is an injection. If \( V \) is finite dimensional this gives an isomorphism.

**Proof.** Suppose \( \varphi(T(v), T(w)) = \varphi(v, w) \) for all \( v, w \in V \). This gives
\[ ||T(v)||^2 = \varphi(T(v), T(v)) = \varphi(v, v) = ||v||^2, \]
i.e., \( ||T(v)|| = ||v|| \) for all \( v \in V \).

Conversely, suppose \( ||T(v)|| = ||v|| \) for all \( v \in V \). Then we have the following identities:
1. If \( \varphi \) is a symmetric bilinear form, then
\[
\varphi(v, w) = \frac{1}{4}||v + w||^2 - \frac{1}{4}||v - w||^2.
\]

2. If \( \varphi \) is a Hermitian form, then
\[
\varphi(v, w) = \frac{1}{4}||v + w||^2 - \frac{i}{4}||v + iw||^2 - \frac{1}{4}||v - w||^2 - \frac{i}{4}||v - iw||^2.
\]

These two identities are enough to give the result. For example, if \( \varphi \) is a symmetric bilinear form we have
\[
\varphi(T(v), T(w)) = \frac{1}{4}||T(v) + T(w)||^2 - \frac{1}{4}||T(v) - T(w)||^2
\]
\[
= \frac{1}{4}||T(v + w)||^2 - \frac{1}{4}||T(v - w)||^2
\]
\[
= \frac{1}{4}||v + w||^2 - \frac{1}{4}||v - w||^2
\]
\[
= \varphi(v, w)
\]
for all \( v, w \in V \).

**Corollary 6.2.3.** Let \( T \in \text{Hom}_F(V, V) \) be an isometry. Then \( T \) has an adjoint and \( T^* = T^{-1} \).

**Proof.** Note that by definition if \( T \) is an isometry then \( T \) is an isomorphism, which implies \( T^{-1} \) exists. Now we calculate
\[
\varphi(v, T^{-1}(w)) = \varphi(T(v), T(T^{-1}(w)))
\]
\[
= \varphi(T(v), w)
\]
for all \( v, w \in V \). This fact, combined with the uniqueness of the adjoint map gives \( T^* = T^{-1} \).

**Corollary 6.2.4.** Let \( T \in \text{Hom}_F(V, V) \).

1. If \( T \) is self-adjoint, then \( T \) is normal.

2. If \( T \) is an isometry, then \( T \) is normal.

**Proof.** If \( T \) is self-adjoint, the only thing we need to check for \( T \) to be normal is that \( T \) commutes with its adjoint. However, self-adjoint gives \( T = T^* \), so this is obvious. If \( T \) is an isometry we have just seen that \( T^* \) exists and is equal to \( T^{-1} \). Again, the fact that \( T \) commutes with its adjoint is clear.

Recall that for a nondegenerate symmetric bilinear form the isometry group is denoted \( \text{O}(\varphi) \) and elements of this group are said to be *orthogonal*. If \( \varphi \) is a nondegenerate Hermitian form we denoted the isometry group by \( \text{U}(\varphi) \) and elements of this group are said to be *unitary*. We say a matrix \( P \in \text{GL}_n(F) \) is *orthogonal* if \( P = P^{-1} \) and we say \( P \) is *unitary* if \( P = P^{-1} \). In the case \( F = \mathbb{R} \) or \( F = \mathbb{C} \), the statement \( P \) is orthogonal is equivalent to \( P \in \text{O}_n(\mathbb{R}) \) and \( P \) being unitary is equivalent to \( P \in \text{U}_n(\mathbb{C}) \).
Corollary 6.2.5. Let $T \in \text{Hom}_F(V, V)$. Let $\mathcal{C}$ be an orthonormal basis and $M = [T]_\mathcal{C}$. Then

1. If $(V, \varphi)$ is a real vector space, then
   
   (a) if $T$ is self-adjoint, then $M$ is symmetric ($^tM = M$).
   
   (b) if $T$ is orthogonal, then $M$ is orthogonal.

2. If $(V, \varphi)$ is a complex vector space. Then
   
   (a) if $T$ is self-adjoint, then $M$ is Hermitian ($^tM = \overline{M}$).
   
   (b) if $T$ is unitary, then $M$ is unitary.

Proof. Recall that for $\mathcal{C}$ an orthonormal basis we have

$$[T^*]_\mathcal{C} = (^t[T])_\mathcal{C}$$

via Corollary 5.3.6. If $T$ is self-adjoint, this gives

$$M = [T]_\mathcal{C} = [T^*]_\mathcal{C} = ^tM,$$

i.e., $M$ is symmetric. The statement that $T$ being self-adjoint in the complex case implies $M$ is Hermitian follows along the exact same lines. The statements in regards to orthogonal and unitary were given before when discussing isometry groups in the previous chapter.

One should also note the reverse direction in the above result. For example, if $M$ is a symmetric matrix, then the associated map $T_A$ is certainly self-adjoint.

The following example is the fundamental example to keep in mind when working with normal and self-adjoint linear maps.

Example 6.2.6. Let $\lambda_1, \ldots, \lambda_m \in F$. Let $W_1, \ldots, W_m$ be nonzero subspaces of $V$ such that $V = W_1 \perp \cdots \perp W_m$. Define $T \in \text{Hom}_F(V, V)$ as follows. Let $\mathcal{B}_i = \{w_i^1, \ldots, w_i^k\}$ be a basis for $W_i$. Set $T(w_i^j) = \lambda_i w_i^j$ for $j = 1, \ldots, k$. This gives $W_i$ is the $\lambda_i$-eigenspace for $T$, i.e., $T(v) = \lambda_i v$ for all $v \in W_i$.

For any $v \in V$ we can write $v = v_1 + \cdots + v_m$ for some $v_i \in W_i$. We have

$$T(v) = \lambda_1 v_1 + \cdots + \lambda_m v_m$$

and

$$T^*(v) = \overline{\lambda_1} v_1 + \cdots + \overline{\lambda_m} v_m.$$ 

Thus

$$(T^* \circ T)(v) = |\lambda_1|^2 v_1 + \cdots + |\lambda_m|^2 v_m = (T \circ T^*)(v).$$

This gives that $T$ is a normal map. Moreover, we see immediately that $T$ is self-adjoint if and only if $\lambda_i = \overline{\lambda_i}$ for all $i$.

In the following proposition we collect some of the results we’ve already seen, as well as give some new ones that will be of use.
Proposition 6.2.7. Let $T \in \text{Hom}_F(V,V)$ be normal. Then $T^*$ is normal. Furthermore,

1. We have $p(T)$ is normal for all $p(x) \in F[x]$. If $T$ is self-adjoint, then $p(T)$ is self-adjoint.

2. We have $||T(v)|| = ||T^*(v)||$ for all $v$ and $\ker T = \ker T^*$.

3. We have $\ker T = (\text{Im} T)^\perp$ and $\ker T^* = (\text{Im} T^*)^\perp$.

4. If $T^2(v) = 0$, then $T(v) = 0$.

5. If $v \in V$ is an eigenvector of $T$ with eigenvalue $\lambda$, then $v$ is an eigenvector of $T^*$ with eigenvalue $\overline{\lambda}$.

6. Eigenspaces of distinct eigenvalues are orthogonal.

Proof. We have already seen that for $\varphi$ symmetric or Hermitian that if $T$ has an adjoint, so does $T^*$ and $(T^*)^* = T$. Since $T$ is assumed to be normal it has an adjoint and

$$T^*(T^*)^* = T^*T = TT^* = (T^*)^*T^*.$$ 

Thus, $T^*$ is normal.

We have already seen 1) and 5).

For 2) we have

$$||T(v)||^2 = \varphi(T(v),T(v)) = \varphi(v,T^*T(v)) = \varphi(v,TT^*(v)) = \varphi(v,(T^*)^*T^*(v)) = \varphi(T^*(v),T^*(v)) = ||T^*(v)||^2.$$

The statement about kernels now follows immediately.

We now prove 3). Note we have just seen that $v \in \ker T$ if and only if $v \in \ker T^*$. We have $v \in \ker T = \ker T^*$ is equivalent to $\varphi(T^*(v),w) = 0$ for every $w \in V$. However, since $\varphi(T^*(v),w) = \varphi(v,T(w))$, this gives that $v \in \ker T$ if and only if $\varphi(v,T(w)) = 0$ for every $w \in V$, i.e., if and only if $v \in (\text{Im} T)^\perp$. One gets the other equality by switching $T$ and $T^*$.

Assume $T^2(v) = 0$ and let $w = T(v)$. Note that $T(w) = T^2(v) = 0$, so $w \in \ker T$. However, $w \in \text{Im} T$ as well, so we have $w \in \ker T \cap \text{Im} T = 0$ by 3). This gives part 4).
We now prove 6). Let \( v_1 \) be an eigenvector of \( T \) with eigenvalue \( \lambda_1 \) and \( v_2 \) with eigenvalue \( \lambda_2 \) and assume \( \lambda_1 \neq \lambda_2 \). Set \( S = T - \lambda_1 I \), so \( S(v_1) = 0 \). We have

\[
0 = \varphi(S(v_1), v_2) \\
= \varphi(v_1, S^*(v_2)) \\
= \varphi(v_1, (T^* - \lambda_1)(v_2)) \\
= \varphi(v_1, (\overline{\lambda_2} - \lambda_1)v_2) \\
= (\lambda_1 - \lambda_2)\varphi(v_1, v_2).
\]

However, by assumption we have \( \lambda_2 - \lambda_1 \neq 0 \) so it must be that \( \varphi(v_1, v_2) = 0 \), i.e., \( v_1 \) and \( v_2 \) are orthogonal. \( \square \)

**Exercise 6.2.8.** Let \( (V, \varphi) \) be finite dimensional and \( T \in \text{Hom}_F(V, V) \) be normal. Show \( \text{Im} T = \text{Im} T^* \).

**Proposition 6.2.9.** Let \( (V, \varphi) \) be a finite dimensional inner product space and let \( T \in \text{Hom}_F(V, V) \) be normal. The minimal polynomial of \( T \) is a product of distinct irreducible factors. If \( V \) is a \( \mathbb{C} \)-vector space or \( V \) is an \( \mathbb{R} \)-vector space and \( T \) is self-adjoint, then every factor is linear.

**Proof.** Let \( p(x) \) be an irreducible factor of \( m_T(x) \). Suppose \( p^2 | m_T \). Thus, there is a vector \( v \) so that \( p^2(T)(v) = 0 \), but \( p(T)(v) \neq 0 \). Let \( S = p(T) \). Then \( S \) is normal and \( S^2(v) = 0 \), but \( S(v) \neq 0 \). This is a contradiction to 4) in the previous proposition, so \( p(x) \) exactly divides \( m_T(x) \).

For the second statement if \( V \) is a \( \mathbb{C} \) vector space there is nothing to prove. Assume \( V \) is an \( \mathbb{R} \)-vector space and \( T \) is self-adjoint. We can factor \( m_T(x) \) into linear and quadratic factors as this is true for any polynomial over \( \mathbb{R} \). Let \( p(x) = x^2 + bx + c \) be irreducible over \( \mathbb{R} \) and assume \( p(x) | m_T(x) \). Note that \( p(x) \) irreducible over \( \mathbb{R} \) gives \( b^2 - 4c < 0 \). Let \( v \in V \) with \( v \neq 0 \) such that \( p(T)(v) = 0 \). Write \( p(x) = (x + \frac{b}{2})^2 + d^2 \) with \( d = \sqrt{\frac{4c - b^2}{4}} \in \mathbb{R} \). Set \( S = T + \frac{b}{2} I \). Then \( (S^2 + d^2 I)(v) = 0 \), i.e., \( S^2 v = -d^2 v \). Since we are assuming \( T \) is self-adjoint we obtain \( S \) is self-adjoint because \( \frac{b}{2} \in \mathbb{R} \). Thus, we have

\[
0 < \varphi(S(v), S(v)) \\
= \varphi(S^*S(v), v) \\
= \varphi(S^2(v), v) \\
= -d^2 \varphi(v, v).
\]

This is clearly a contradiction, so it must be that \( p(x) \) is reducible over \( \mathbb{R} \). \( \square \)

**Theorem 6.2.10** (Spectral Theorem). 1. Let \( (V, \varphi) \) be a finite dimensional \( \mathbb{C} \)-vector space and let \( T \in \text{Hom}_F(V, V) \) be normal. Then \( V \) has an orthonormal basis that is an eigenbasis for \( T \).
2. Let $V$ be a finite dimensional $\mathbb{R}$-vector space and let $T \in \text{Hom}_F(V, V)$ be self-adjoint. Then there is an orthonormal basis of $V$ that is an eigenbasis for $T$.

Proof. Use the previous result to split $m_T(x)$ into distinct linear factors and let $\lambda_1, \ldots, \lambda_k$ be the roots of $m_T(x)$. Let $E_{\lambda_i}$ be the eigenspace associated to $\lambda_i$. We have $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$. But we saw eigenspaces of distinct eigenvalues are orthonormal. So $V = E_{\lambda_1} \perp \cdots \perp E_{\lambda_k}$. Let $C_i$ be an orthonormal basis of $E_{\lambda_i}$. Then $C = \bigcup_{i=1}^k C_i$ is the desired basis.

One can also write the Spectral theorem in the following form which will be useful in the next section. The proof is left as an exercise.

**Corollary 6.2.11.** Let $T \in \text{Hom}_F(V, V)$ be as in Theorem 6.2.10. Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of $T$. There are projection maps $\pi_1, \ldots, \pi_r$ so that

1. $T = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r$;
2. $\text{id} = \pi_1 + \cdots + \pi_r$;
3. $\pi_i \circ \pi_j = 0$ if $i \neq j$.

**Exercise 6.2.12.** Let $\lambda_1, \ldots, \lambda_r$ be the eigenvalues of $T$ and write

$$T = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r$$

as above. Show that

$$\pi_i = \prod_{j \neq i} \frac{1}{\lambda_i - \lambda_j} (T - \lambda_j \text{id}).$$

We can also rephrase the Spectral theorem in terms of matrices.

**Corollary 6.2.13.**

1. Let $A$ be a Hermitian matrix. Then there is a unitary matrix $P$ and a diagonal matrix $D$ such that $A = PD P^{-1} = PD^t P$.

2. Let $A$ be a real symmetric matrix. Then there is a real orthogonal matrix $P$ and real diagonal matrix $D$ such that $A = PD P^{-1} = PD^t P$.

Proof. Let $A \in \text{Mat}_n(\mathbb{R})$ be a symmetric matrix. Let $V = \mathbb{R}^n$ and let $E_n$ be the standard basis of $V$. We have that $A = [T_A]_{E_n}$. The fact that $A$ is symmetric immediately gives $T_A$ is a self-adjoint map. Thus, the spectral theorem implies there is an orthonormal basis $B$ of $V$ that is an eigenbasis for $T_A$, i.e., $[T_A]_{B} = D$ is a diagonal matrix. Let $Q$ be the change of basis matrix from $B$ to $E_n$. Then $Q$ is an orthogonal matrix and $QAQ^{-1} = D$. (To see why it is orthogonal, see the discussion preceding Exercise 5.3.7.) Thus, if we set $P = Q^{-1}$ we have the result in the real symmetric case. The Hermitian case follows along the same lines.
Example 6.2.14. Consider the matrix \( A = \begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix} \). We want to find a unitary matrix \( P \) and a diagonal matrix \( D \) so that \( A = PDP^{-1} \). Observe we have \( c_A(x) = x^2 - 5x + 4 = (x - 4)(x - 1) \), so the eigenvalues of \( A \) are \( \lambda_1 = 1 \) and \( \lambda_2 = 4 \). Thus, we must have \( D = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \). This gives that \( \mathbb{C}^2 = \ker(A - 1) \perp \ker(A - 41) \). One computes that \( \ker(A - 1) = \text{span}_\mathbb{C} \{ \left( \begin{array}{c} -1-i \\ 1 \end{array} \right) \} \) and \( \ker(A - 41) = \text{span}_\mathbb{C} \{ \left( \begin{array}{c} 1+i \\ 2 \end{array} \right) \} \). We now apply the Gram-Schmidt procedure to obtain an orthonormal basis given by \( w_1 = \left( \begin{array}{c} -1 \\ i \sqrt{3} \\ 1 \sqrt{3} \end{array} \right) \) and \( w_2 = \left( \begin{array}{c} 1 \\ -i \sqrt{6} \\ 2 \sqrt{6} \end{array} \right) \). Thus, we have \( P = \begin{pmatrix} -1-i/ \sqrt{3} & 1+i/ \sqrt{6} \\ 1-i/ \sqrt{3} & 1/ \sqrt{6} \end{pmatrix} \). One immediately checks that \( A = PDP^{-1} \) as desired.

Example 6.2.15. Consider the matrix \( A = \begin{pmatrix} 1 & 1 & 4 \\ 1 & 1 & 4 \\ 4 & 4 & -2 \end{pmatrix} \). We easily calculate that \( c_A(x) = x(x - 6)(x + 6) \). Write \( \lambda_1 = 0, \lambda_2 = 6, \lambda_3 = -6 \). Using the above exercise we have

\[
\begin{align*}
\pi_1 &= \frac{A - 6 \cdot 1_3}{-6} - A + 6 \cdot 1_3 = \begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\pi_2 &= \frac{A}{6} - A + 6 \cdot 1_3 = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \\
\pi_3 &= \frac{A}{-6} - A + 6 \cdot 1_3 = \begin{pmatrix} 1/6 & 1/6 & -1/3 \\ 1/6 & 1/6 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}.
\end{align*}
\]

Now given any \( v \in \mathbb{R}^3 \), we have \( \pi_i(v) \) lies in the \( \lambda_i \)-eigenspace. For example, we have

\[
\pi_3 \left( \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}
\]

and

\[
A \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ -6 \end{pmatrix} = -6 \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}.
\]

This allows one to quickly give an eigenbasis, and then one applies Gram-Schmidt to get an orthonormal eigenbasis to recover the orthogonal matrix \( P \) that diagonalizes \( A \).
6.2. THE SPECTRAL THEOREM

We saw in the previous section a way to use Gram-Schmidt to compute the signature. The following corollary allows us to compute the signature by counting eigenvalues.

**Corollary 6.2.16.** Let \((V, \varphi)\) be an inner product space. Set \(n = \dim_F V\). Let \(B\) be a basis for \(V\) and set \(A = [\varphi]_B\). Then

1. \(A\) has \(n\) real eigenvalues (counting multiplicity);
2. the signature of \(\varphi\) is \((r, s)\) where \(r\) is the number of positive eigenvalues and \(s\) is the number of negative eigenvalues of \(A\).

**Proof.** The first statement is left as a homework problem.

Let \(\varphi\) be a nondegenerate symmetric bilinear form. Then if we take any basis \(B\) of \(V\), we have \(A = [\varphi]_B\) is a symmetric matrix. We apply Corollary 6.2.13 to see that there exists an orthonormal basis \(C = \{v_1, \ldots, v_n\}\) so that \(D = [\varphi]_C\) is a diagonal matrix, i.e., there is an orthogonal matrix \(P\) so that \(A = PDP^{-1}\). The diagonal entries of \(D\) are precisely the eigenvalues of \(A\).

Upon reordering \(C\) we can assume that the first \(r\) diagonal entries of \(D\) are positive and the remaining \(s = n - r\) are negative. Let \(W_1 = \text{span}_F \{v_1, \ldots, v_r\}\) and \(W_2 = \text{span}_F \{v_{r+1}, \ldots, v_n\}\). We have \(V = W_1 \perp W_2\). Observe that \(\varphi|_{W_1}\) is positive definite and \(\varphi|_{W_2}\) is negative definite. Thus, the signature of \(\varphi\) is exactly \((\dim_F W_1, \dim_F W_2) = (r, s)\), as claimed.

Recall we saw before that if \(S\) and \(T\) are diagonalizable, they are simultaneously diagonalizable if and only if \(S\) and \(T\) commute. In fact, it was noted this result can be extended to a countable collection of diagonalizable maps. This, along with the spectral theorem, immediately gives the following result.

**Corollary 6.2.17.** Let \(\{T_i\}\) be a countable collection of normal or self-adjoint maps. Then \(\{T_i\}\) is simultaneously diagonalizable if and only if the \(T_i\) all commute.

We have the following elementary calculation of when an isometry with a minimal polynomial that splits completely is an isometry.

**Corollary 6.2.18.** Let \(T \in \text{Hom}_F(V, V)\) be normal. Suppose \(m_T(x)\) is a product of linear factors over \(F\). Then \(T\) is an isometry if and only if \(|\lambda| = 1\) for every eigenvalue \(\lambda\) of \(T\).

**Proof.** Let \(T\) be an isometry and let \(v\) be an eigenvector with eigenvalue \(\lambda\). Then

\[
\varphi(v, v) = \varphi(T(v), T(v)) = \varphi(\lambda v, \lambda v) = \lambda^2 \varphi(v, v) = |\lambda|^2 \varphi(v, v).
\]

Thus, \(|\lambda| = 1\).
Let $B = \{v_1, \ldots, v_n\}$ be a basis of orthogonal eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$. Assume $|\lambda_i| = 1$ for all $i$. Pick any $v \in V$ and write $v = \sum a_i v_i$ for some $a_i \in F$. We have

$$\varphi(T(v), T(v)) = \varphi\left(\sum a_i T(v_i), \sum a_j T(v_j)\right)$$

$$= \sum_{i,j} \varphi(a_i \lambda_i v_i, a_j \lambda_j v_j)$$

$$= \sum_{i,j} a_i \lambda_i a_j \lambda_j \varphi(v_i, v_j)$$

$$= \sum_j \lambda_j^2 \sum_{i,j} a_i a_j \varphi(v_i, v_j)$$

$$= \varphi\left(\sum_j a_j v_j, \sum_j a_j v_j\right)$$

$$= \varphi(v, v).$$

Thus, $T$ is an isometry as desired.

We now have the very nice result that if we have a normal map then it is diagonalizable with respect to an orthonormal basis. Of course, not all maps are normal so it is natural to ask if anything can be said in the more general case. We close this section with a couple of results in that direction.

**Theorem 6.2.19** (Schur’s Theorem). Let $(V, \varphi)$ be a finite dimensional inner product space and let $T \in \text{Hom}_F(V, V)$. Then $V$ has an orthonormal basis $C$ so that $[T]_C$ is upper triangular if and only if the minimal polynomial $m_T(x)$ is a product of linear factors.

**Proof.** Suppose there is an orthonormal basis $C$ such that $[T]_C$ is upper triangular. Then $c_T(x) = \prod (x - \lambda_i)$ for $\lambda_i$ the diagonal entries. Since $m_T(x) \mid c_T(x)$ we immediately obtain $m_T(x)$ splits into a product of linear factors.

Suppose $m_T(x)$ factors as a product of linear factors. Let $W$ be a $T$-invariant subspace. We claim $W^\perp$ is a $T^*$-invariant subspace. To see this, observe that for any $w \in W, y \in W^\perp$ we have

$$0 = \varphi(T(w), y) = \varphi(w, T^*(y)).$$

Thus, $T^*(y) \in W^\perp$, i.e., $W^\perp$ is $T^*$-invariant as claimed. We now use induction on the dimension of $V$. If $\dim_F V = 1$ the result is trivial. Suppose the result is true for any vector space of dimension less than $n$. Let $\dim_F V = n$. Note that since $m_T(x)$ splits into linear factors, so does $m_{T^*}(x)$ because $m_{T^*}(x) = m_T(x)$. Then $T^*$ has an eigenvector, say $v_n$. Scale this so $||v_n|| = 1$. Let $W = \text{span}_F \{v_n\}$. Using the fact that $W$ is $T^*$-invariant, $W^\perp$ is an $n - 1$ dimensional
subspace of $V$ that is a $(T^*)^*$-invariant subspace, i.e., a $T$-invariant subspace. Set $S = T |_{W^\perp}$. Now $m_S | m_T$ so $m_S$ splits into linear factors. Applying the induction hypothesis we get an orthonormal basis $C_1 = \{v_1, \ldots, v_{n-1}\}$ such that $[S]_{C_1}$ is upper triangular. Thus, setting $C = C_1 \cup \{v_n\}$, we have $[T]_C$ is upper triangular as desired.

In fact, it turns out we can characterize whether a map is normal by whether it can be diagonalized.

**Theorem 6.2.20.** Let $(V, \varphi)$ be a finite dimensional inner product space and $T \in \text{Hom}_F(V, V)$. Let $C$ be any orthonormal basis of $V$ with $[T]_C$ upper triangular. Then $T$ is normal if and only if $[T]_C$ is diagonal.

**Proof.** Certainly if $[T]_C$ is diagonal then $T$ is normal. So suppose $T$ is normal and set $E = [T]_C$. Let $C_1$ be a basis so that $[T]_{C_1} = D$ is diagonal. Such a basis exists by the Spectral Theorem. Set $P = [T]_{C_1}^C$, so $E = PDP^{-1}$. Since $C$ and $C_1$ are both orthonormal, we have $P$ is orthogonal in the real case and unitary in the Hermitian case, i.e., if $F = \mathbb{R}$, then $^tP = P^{-1}$ and if $F = \mathbb{C}$, then $^tP = \overline{P}^{-1}$. Using this, we have if $F = \mathbb{R}$ then

$$^tE = (^tPDP^{-1})$$
$$= (^tPD^tP)$$
$$= PD^tP$$
$$= PDP^{-1}$$
$$= E.$$ 

Since $E$ is assumed to be upper triangular, the only way it can equal its transpose is if it is actually diagonal. The same argument works in the Hermitian case as well.

### 6.3 Polar decomposition and the Singular Value Theorem

Recall that given a complex number $z$, one can write $z = x + iy$ for $x, y \in \mathbb{R}$ or $z = re^{i\theta}$ for $r \in \mathbb{R}_{\geq 0}$ and $\theta \in [0, 2\pi)$. One can ask if the same thing can be done for a matrix $A \in \text{Mat}_n(\mathbb{C})$. In the case of writing $z = x + iy$, the analogy was given in Problem 5 in Section 5.4 via the Hermitian decomposition of a matrix. To see this, recall that one showed $A = H + S$ for $H$ a Hermitian matrix and $S$ a skew-Hermitian matrix. Furthermore, recall that one can write $S = \overline{iH_2}$ for $H_2$ a Hermitian matrix, so this equation can be written as $A = H_1 + iH_2$ for $H_1$ Hermitian matrices. (One does not have that the $H_i$ have entries in $\mathbb{R}$ so don’t read too much into the analogy.) Naturally, we would like to give a version of the polar decomposition as well. That will be one of the main goals of this section. Once we have the polar decomposition of a complex matrix, we will use this to prove the singular value decomposition theorem.

144
6.3. POLAR DECOMPOSITION AND THE SINGULAR VALUE

THEOREM CHAPTER 6.

We require all the vector spaces to be finite dimensional inner product spaces in this section.

We begin with some more results on self-adjoint linear maps. First, observe that given a linear map \( T \in \text{Hom}_F(V, V) \), the map \( T^*T \) is always a self-adjoint linear map. This is because we have for any \( v, w \in V \)
\[
\varphi((T^*T)(v), w) = \varphi(T(v), T(w)) \\
= \varphi(v, (T^*T)(w)).
\]

Similarly, we have \( TT^* \) is self-adjoint. This allows us to apply the Spectral Theorem to these maps and conclude they are both diagonalizable. This will be very important for our applications. Moreover, since they are self-adjoint one has that the eigenvalues are all non-negative real numbers. Thus, we can write the eigenvalues of \( T^*T \) as \( \mu_1^2, \ldots, \mu_n^2 \) for some \( \mu_i > 0 \). Moreover, \( T^*T \) and \( TT^* \) actually have the same eigenvalues. (This is true in general. Given \( T : V \to W \) and \( S : W \to V, TS \) and \( ST \) have the same eigenvalues.)

**Definition 6.3.1.** Let \( T \in \text{Hom}_F(V, V) \). The positive square roots of the eigenvalues of \( T^*T \) are called the singular values of \( T \).

**Definition 6.3.2.** Let \( T \in \text{Hom}_F(V, V) \). We say \( T \) is positive if \( T \) is self-adjoint and \( \varphi(T(v), v) > 0 \) for all \( v \neq 0 \).

The positive linear maps act as our generalization of positive numbers. In particular, we know a complex number \( z \) is positive if and only if we can write \( z = w\overline{w} \) for some \( w \in \mathbb{C} \). We have the following result.

**Theorem 6.3.3.** Let \( T \in \text{Hom}_F(V, V) \). Then \( T \) is positive if and only if there is an invertible linear map \( S \in \text{Hom}_F(V, V) \) so that \( T = S^*S \).

**Proof.** First, suppose there exists such an \( S \). We have \( T^* = (S^*S)^* = S^*S = T \), so \( T \) is self-adjoint. Observe that \( \varphi(T(v), v) = \varphi(S^*S(v), v) = \varphi(S(v), S(v)) \geq 0 \). Since \( S \) is invertible, if \( v \neq 0 \) then \( S(v) \neq 0 \) and so \( \varphi(T(v), v) > 0 \) and so \( T \) is positive.

Now assume \( T \) is positive. We have via the homework problems that \( \psi(v, w) = \varphi(T(v), w) \) is an inner product on \( V \). Let \( B = \{v_1, \ldots, v_n\} \) be an orthonormal basis with respect to \( \varphi \) and \( C = \{w_1, \ldots, w_n\} \) an orthonormal basis with respect to \( \psi \). Then we have
\[
\psi(w_i, w_j) = \delta_{ij} = \varphi(v_i, v_j).
\]

Define \( S \) to be the unique linear map defined by \( S(w_j) = v_j \). This is invertible since \( B \) and \( C \) are bases. Observe that
\[
\psi(w_i, w_j) = \varphi(S(w_i), S(w_j)) = \varphi(v_i, v_j).
\]

Let \( w, w' \in V \) and write \( w = \sum a_iw_i \), \( w' = \sum b_iw_i \). Using that \( \psi \) is an inner product we see
\[
\psi(w, w') = \varphi(S(w), S(w')).
\]

The definition of \( \psi \) gives \( \varphi(T(w), w') = \psi(w, w') = \varphi(S(w), S(w')) = \varphi(S*S(w), w') \) for all \( w, w' \in V \). Thus, \( T = S^*S \).
6.3. POLAR DECOMPOSITION AND THE SINGULAR VALUE

THEOREM

If \( F = \mathbb{C} \) we can drop the requirement that \( T \) be self-adjoint from the definition of positive. This is because if \( \varphi(T(v), v) > 0 \) for all \( v \neq 0 \) in this case then \( T \) is automatically self-adjoint. This follows from the following result.

Lemma 6.3.4. Let \( V \) be a complex inner product space and \( T \in \text{Hom}_F(V,V) \). If \( \varphi(T(v), v) \in \mathbb{R} \) for all \( v \in V \), then \( T \) is self-adjoint.

Proof. Let \( v, w \in V \). Observe we have

\[
\varphi(T(v + w), v + w) = \varphi(T(v), v) + \varphi(T(v), w) + \varphi(T(w), v) + \varphi(T(w), w).
\]

We are assuming \( \varphi(T(v + w), v + w), \varphi(T(v), v), \text{ and } \varphi(T(w), w) \) are all real, so we must have \( \varphi(T(v), w) + \varphi(T(w), v) \) is real as well. Now run the same argument with the vector \( v + iw \) and we have

\[
\varphi(T(v + iw), v + iw) = \varphi(T(v), v) - i\varphi(T(v), w) + i\varphi(T(w), v) + \varphi(T(w), w).
\]

As above, this gives \(-i\varphi(T(v), w) + i\varphi(T(v), w)\) is real. Since these numbers are real, they are equal to their complex conjugates:

\[
\varphi(T(v), w) + \varphi(T(w), v) = \overline{\varphi(T(v), w) + \varphi(T(w), v)}
\]

\[
= \varphi(w,T(v)) + \varphi(v,T(w))
\]

and

\[
-i\varphi(T(v), w) + i\varphi(T(v), w) = \overline{-i\varphi(T(v), w) + i\varphi(T(v), w)}
\]

\[
= i\varphi(w,T(v)) - i\varphi(w,T(v)).
\]

We now multiply the second set of equations by \( i \) and add the result to the first set of equations and obtain

\[
2\varphi(T(v), w) = 2\varphi(w,T(v)),
\]

i.e., \( T \) is self-adjoint.

The previous result is not true for real inner product spaces as clearly we have \( \varphi(T(v), v) \in \mathbb{R} \) for any \( v \in V \) automatically. In general, if \( T \) is self-adjoint we have \( \varphi(T(v), v) = \varphi(v,T(v)) = \overline{\varphi(T(v), v)} \) so \( \varphi(T(v), v) \) is real. Thus, for a complex inner product space we see that \( T \) is self-adjoint if and only if \( \varphi(T(v), v) \in \mathbb{R} \) for all \( v \in V \) and \( T \) is positive if and only if \( \varphi(T(v), v) > 0 \) for all \( v \neq 0 \).

Theorem 6.3.5. Let \( T \in \text{Hom}_F(V,V) \) be a normal map on a complex inner product space. Then \( T \) is self-adjoint, positive, or unitary according to if its eigenvalues are real, positive, or absolute value 1.

Proof. The result on unitary maps was given in the homework of the previous chapter. The result on self-adjoint maps follows immediately from the fact that \( \varphi(T(v), v) \in \mathbb{R} \) if \( T \) is self-adjoint. It only remains to prove the result for positive
6.3. POLAR DECOMPOSITION AND THE SINGULAR VALUE THEOREM

maps. Using the Spectral Theorem write \( T = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r \). Then we have for \( v \in V \) that

\[
\varphi(T(v), v) = \varphi \left( \sum_{i=1}^{r} \lambda_i \pi_i(v), \sum_{j=1}^{r} \pi_j(v) \right) \\
= \sum_{i,j=1}^{r} \lambda_i \varphi(\pi_i(v), \pi_j(v)) \\
= \sum_{j=1}^{r} \lambda_j ||\pi_j(v)||^2.
\]

Thus, we see that \( \varphi(T(v), v) > 0 \) is satisfied if and only if \( \lambda_j > 0 \) for each \( j \).

**Definition 6.3.6.** Let \( T \in \text{Hom}_F(V, V) \). We say \( T \) is non-negative if \( T \) is self-adjoint and \( \varphi(T(v), v) \geq 0 \) for all \( v \in V \).

As above, if \( F = \mathbb{C} \) then we can drop the assumption that \( T \) is self-adjoint.

**Lemma 6.3.7.** Let \( T \in \text{Hom}_F(V, V) \) be a non-negative linear map. There is a unique non-negative linear map \( S \in \text{Hom}_F(V, V) \) so that \( T = S^2 \).

**Proof.** We can apply the Spectral Theorem to write \( T = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r \). Since \( T \) is assumed to be non-negative, we have \( \lambda_i \geq 0 \) for all \( i \). Set \( S = \sqrt{\lambda_1 \pi_1 + \cdots + \lambda_r \pi_r} \). It is clear we have \( S^2 = T \). Now let \( R \) be any other non-negative linear map with \( R^2 = T \). Let \( R = d_1 \pi_1 + \cdots + d_r \pi_r \) be the spectral resolution of \( R \). Since \( R \) is assumed to be non-negative, \( d_i \geq 0 \) for all \( i \). Since \( R^2 = T \) we have

\[
T = d_1^2 \pi_1 + \cdots + d_r^2 \pi_r.
\]

This decomposition satisfies the conditions of being a spectral resolution of \( T \), and so we must have \( d_i^2 = \lambda_i \) for each \( i \). Since \( d_i \geq 0 \), we get \( d_i = \sqrt{\lambda_i} \) for each \( i \) and so \( R = S \) as claimed.

**Exercise 6.3.8.** Given \( T \) and \( S \) as in the last theorem, the eigenvalues of \( S \) are given by \( \sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r} \) and for each \( i \) we have \( E_{\sqrt{\lambda_i}}(T) = E_{\lambda_i}(S) \).

**Exercise 6.3.9.** Prove that if \( T \in \text{Hom}_F(V, V) \) is a positive self-adjoint linear map, then for any \( m \geq 2 \) there is a unique positive self-adjoint linear map \( S \in \text{Hom}_F(V, V) \) so that \( T = S^m \).

We now state the Polar Decomposition Theorem. One should think of the non-negative linear map as the “\( r \)’” and the orthogonal/unitary map as the “\( e^{i\theta} \)”.

**Theorem 6.3.10.** Let \( T \in \text{Hom}_F(V, V) \). There are two non-negative linear maps \( S_1, S_2 \in \text{Hom}_F(V, V) \) and a linear map \( U \in \text{Hom}_F(V, V) \) (\( U \) is orthogonal if \( F = \mathbb{R} \) and unitary if \( F = \mathbb{C} \)) so that

\[
T = US_1 = S_2 U.
\]
Moreover, if \( \dim_F \text{Im}(T) = r \), then \( S_1 \) and \( S_2 \) have the same positive eigenvalues \( \mu_1, \ldots, \mu_r \), which are the singular values of \( T \). If \( T \) is invertible then \( U, S_1 \), and \( S_2 \) are unique. If \( T \) is normal, \( S_1 = S_2 \).

**Proof.** Suppose that we have \( T = US_1 \). Then \( T^* = (US_1)^* = S_1^*U^* = S_1U^* \) because \( S_1 \) is self-adjoint. Thus, \( T^*T = S_1U^*US_1 = S_1^2 \) because \( U \) is orthogonal or unitary. Thus, \( S_1 \) is uniquely determined by \( T \) because it is the unique square root of \( T^*T \) as given above. A similar argument works for \( S_2 \) except one gets it is the square-root of \( TT^* \).

We now need to prove the existence of \( U \) and \( S_1 \). Set \( S_1 \) to be the square-root of \( TT^* \). Similarly, define \( S_2 \) to be the square-root of \( TT^* \). This immediately gives that \( S_1 \) and \( S_2 \) are unique and have the same positive eigenvalues. We also have that if \( T \) is normal then \( S_1 = S_2 \). Moreover, Note that if \( T \) is invertible, \( S_1 \) is invertible as well because

\[
\varphi(S_1(v), S_1(v)) = \varphi(S_1^2(v), v) = \varphi(T^*T(v), v) = \varphi(T(v), T(v)).
\]

In this case, set \( U = TS_1^{-1} \). One defines \( S_2 \) similarly. It only remains to prove that \( U \) is orthogonal or unitary. We have \( U^* = (TS_1^{-1})^* = S_1^{-1}T^* \). Thus we have

\[
UU^* = TS_1^{-1}S_1^{-1}T^* = T(S_1^{-2})T^* = T(S_1^2)^{-1}T^* = T(T^*T)^{-1}T^* = TT^{-1}(T^*)^{-1}T^* = \text{id}.
\]

Thus, if \( T \) is invertible we have the result.

We now must deal with the case when \( T \) is not invertible. This is considerably more work to define \( U \). We begin by defining \( U \) on the image of \( S_1 \). Let \( w \in W = \text{Im}(S_1) \), say \( w = S_1(v) \). We want to define \( U \) so that \( US_1(v) = T(v) \). Thus, we set \( U(w) = T(v) \). We must check this is well-defined. Let \( w_1 = S_1(v_1) = S_1(v_2) \); we need to show that \( T(v_1) = T(v_2) \). We saw above that \( ||S_1(v)||^2 = ||T(v)||^2 \) for all \( v \). Set \( v = v_1 - v_2 \). Then we have \( S_1(v) = 0 \) if and only if \( T(v) = 0 \). This gives that \( U \) is well-defined. It now remains to define \( U \) on \( W^\perp \). Observe that \( T \) and \( S_1 \) have the same kernel, so the dimension of the image of \( S_1 \) is equal to the dimension of the image of \( T \). Thus, \( W^\perp \) has the same dimension as the orthogonal complement of the image of \( T \). Thus, we have an isomorphism of inner product spaces \( U_0 : W^\perp \rightarrow \text{Im}(T)^\perp \). Define \( U \) to be equal to \( U_0 \) on \( W^\perp \). We make this a bit clearer. Let \( v \in V \). We can uniquely write \( v = w_1 + w_2 \) with \( w_1 \in W \) and \( w_2 \in W^\perp \). Write \( w_1 = S_1(v_1) \). Define

\[
U(v) = T(v_1) + U_0(w_2).
\]
It is clear that $U$ is a well-defined linear map. We now check it is orthogonal/unitary. We have

$$\varphi(U(v), U(v)) = \varphi(T(v_1) + U_0(w_2), T(v_1) + U_0(w_2))$$

$$= \varphi(T(v_1), T(v_1)) + \varphi(U_0(w_2), U_0(w_2))$$

$$= \varphi(S_1(v_1), S_1(v_1)) + \varphi(w_2, w_2)$$

$$= \varphi(v, v)$$

where we have used that $||T(v_1)|| = ||S_1(v_1)||$ and $U_0$ is orthogonal/unitary because it is an isomorphism of inner product spaces. Thus, $U$ is orthogonal/unitary. This gives $US_1(v_1) = T(v_1)$ for each $v$.

One can run the same argument with $S_2$.

We now rephrase this in terms of matrices.

**Corollary 6.3.11.** 1. Let $A \in \text{Mat}_n(\mathbb{R})$. There is a matrix $U \in O_n(\mathbb{R})$ and a non-negative symmetric matrix $S$ so that

$$A = US.$$  

Moreover, if $A$ is invertible then $U$ and $S$ are unique.

2. Let $A \in \text{Mat}_n(\mathbb{C})$. There is a matrix $U \in U_n(\mathbb{C})$ and a non-negative Hermitian matrix $S$ so that

$$A = US.$$  

Moreover, if $A$ is invertible then $U$ and $S$ are unique.

We also obtain as an immediate corollary the Singular Value Decomposition Theorem (SVD).

**Theorem 6.3.12.** Let $T \in \text{Hom}_F(V, V)$ with $(V, \varphi)$ a inner product space of dimension $n$. There are orthonormal bases $B = \{u_1, \ldots, u_n\}$ and $C = \{v_1, \ldots, v_n\}$ such that if $r = \dim_F \text{Im}(T)$, we have

$$[T]_B^C = \begin{pmatrix}
\mu_1 \\
\vdots \\
\mu_n
\end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of $T$ and $\mu_{r+1} = \cdots = \mu_n = 0$. Moreover, $u_1, \ldots, u_n$ are the eigenvectors of $T^*T$, $v_1, \ldots, v_n$ are the eigenvectors of $TT^*$, and $T(u_i) = \mu_i v_i$ for $1 \leq i \leq n$.

We can rephrase the SVD in terms of matrices.

**Theorem 6.3.13.** 1. Let $A \in \text{Mat}_n(\mathbb{R})$. There are two matrices $U, V \in O_n(\mathbb{R})$ and a diagonal matrix $D$ so that $A = VD^tU$ where $D$ is a diagonal matrix with the positive square roots of the eigenvalues of $^tAA$ on the diagonal. The columns of $U$ are the eigenvectors of $^tAA$ and the columns of $V$ are the eigenvectors of $AA^t$. 

149
2. Let $A \in \text{Mat}_n(\mathbb{R})$. There are two matrices $U, V \in O_n(\mathbb{R})$ and a diagonal matrix $D$ so that $A = VD^tU$ where $D$ is a diagonal matrix with the positive square roots of the eigenvalues of $AA^*$ on the diagonal. The columns of $U$ are the eigenvectors of $AA^*$ and the columns of $V$ are the eigenvectors of $A^*A$.

We briefly indicate how to go from the polar decomposition to the SVD and vice versa. Suppose $A = U_1S$ with $U_1$ being orthogonal/unitary and $S$ positive symmetric/Hermitian. We can use the Spectral Theorem to write $S = U_2DU_2^*$ with $D$ a positive diagonal matrix and $U_2$ orthogonal/unitary. Thus,

$$A = U_1U_2DU_2^*.$$

The SVD is given by setting $V = U_1U_2$ and $U = U_2$. Now suppose we have $A = VDU^*$ is the SVD of $A$. Set $R = VU^*$ and $S = UDU^*$. Then we have $R$ is orthogonal/unitary, $S$ is positive symmetric/Hermitian and

$$RS = VU^*UDU^*$$
$$= VDU^*$$
$$= A.$$ 

We find the singular value decomposition of a matrix now.

**Example 6.3.14.** Let $A = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$. We want to find $U, V$ and $D$ so that $A = UD^tV$ with $U, V$ orthogonal and $D$ diagonal. First, we observe that the two equations we need to work with are

$${}^tAA = V {}^tDD^tV$$

and

$$AV = UD.$$

The first equation is just the diagonalization of $^tAA$. We have $^tAA = \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix}$. The eigenvalues of this are 20 and 80 as $c_{AA}(x) = (x - 20)(x - 80)$. To find $V$, we need to find an orthonormal basis for the eigenspaces of $^tAA$. We have

$$^tAA - 20 \cdot 1_2 = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix}.$$ 

A basis for the kernel of this space is given by $v_1 = \begin{pmatrix} -3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}$. Similarly, a basis for the eigenspace associated to the eigenvector 80 is given by $v_2 = \begin{pmatrix} 1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix}$. These form an orthonormal basis, so

$$V = \begin{pmatrix} -3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{pmatrix}.$$
The singular values of $A$ are given by $\sqrt{20}$ and $\sqrt{80}$, so $D = \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{pmatrix}$. We now just write

$$U = AVD^{-1} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$  

We now have all the components of the singular value decomposition of $A$.

One can give similar results for $A \in \text{Mat}_{m,n}(F)$, but we do not pursue those here. We end by stating (without proof) the following result that relates the eigenvalues of $A$ and the singular values of $A$.

**Theorem 6.3.15.** Let $A \in \text{Mat}_n(\mathbb{C})$ with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ and singular values $\mu_1, \ldots, \mu_n \in \mathbb{R}_{>0}$ listed so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and $\mu_1 \geq \cdots \geq \mu_n$, then

1. $|\lambda_1| \cdots |\lambda_n| = \mu_1 \cdots \mu_n$
2. $|\lambda_1| \cdots |\lambda_k| \leq \mu_1 \cdots \mu_k$ for $k = 1, \ldots, n - 1$. 

151
6.4 Problems

For these problems $V$ and $W$ are finite dimensional $F$-vector spaces.

1. Prove the two identities used in the proof of Lemma 6.2.2. Complete the proof in the case that $\varphi$ is a Hermitian form.

2. Prove the first statement in Corollary 6.2.16.

3. Let $V = \text{Mat}_2(\mathbb{R})$.
   (a) Show that the map $\varphi : V \times V \to \mathbb{R}$ given by $(A, B) \mapsto \text{Tr}(AB)$ is a bilinear form.
   (b) Determine the signature of $\varphi$.
   (c) Determine the signature of $\varphi$ on the subspace $\mathfrak{sl}_2(\mathbb{R})$.

4. Let $\varphi$ be a nondegenerate bilinear form on $V$.
   (a) Given any $\psi \in V^\vee$, show there is a unique vector $v \in V$ so that
   \[
   \psi(w) = \varphi(w, v)
   \]
   for every $w \in V$.
   (b) Find a polynomial $q \in P_2(\mathbb{R})$ so that
   \[
   p(1/2) = \int_0^1 p(t)q(t)dt
   \]
   for every $p \in P_2(\mathbb{R})$.

5. Let $V = P_2(\mathbb{R})$ and let $\varphi(f, g) = \int_{-1}^1 f(x)g(x)dx$. Find an orthonormal basis of $V$.

6. (a) Show that if $V$ is a finite dimensional vector space over $F$ with $F$ not of characteristic 2 and $\varphi$ is a symmetric bilinear form (not necessarily nondegenerate!) then $V$ has an orthogonal basis with respect to $\varphi$.
   (b) Let $V = \mathbb{Q}^3$. Let $\varphi$ be the bilinear form represented by
   \[
   \begin{pmatrix}
   1 & 2 & 1 \\
   2 & 1 & 2 \\
   1 & 2 & 0
   \end{pmatrix}
   \]
   in the standard basis. Find an orthogonal basis of $V$ with respect to $\varphi$. 

152
7. (a) Let \( \varphi \) be a bilinear form on a real vector space \( V \). Show there is a symmetric form \( \varphi_1 \) and a skew-symmetric form \( \varphi_2 \) so that \( \varphi = \varphi_1 + \varphi_2 \).

(b) Let \( A \in \text{Mat}_n(\mathbb{R}) \). Show that there is a unique symmetric matrix \( B \) and a unique skew-symmetric matrix \( C \) so that \( A = B + C \).

8. Let \( A = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix} \). Find the projection maps \( \pi_i \) associated to \( A \) via the Spectral Theorem. Give an orthogonal matrix \( P \) that diagonalizes \( A \).

9. Show that \( T \in \text{Hom}_V(F,F) \) is positive if and only if \( \psi(v,w) = \varphi(T(v),w) \) is an inner product.

10. Find a polar decomposition for the matrix \( \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \).

11. Provide a detailed proof of Theorem 6.3.12.

12. Compute the singular value decomposition and polar decomposition of the matrix \( A = \begin{pmatrix} 8 + i & -12 \\ 4 & -6 + i \end{pmatrix} \).
Chapter 7

Tensor products, exterior algebras, and the determinant

In this chapter we will introduce tensor products and exterior algebras and use the exterior algebra to provide a coordinate free definition of the determinant. The first time one sees a determinant it seems very unnatural. Viewing the determinant from the perspective of exterior algebras makes it a very natural object to consider. It takes considerable machinery to work up to the definition, but once it has been developed the familiar properties of the determinant essentially fall out for free. The benefit of this is that the machinery that one builds up is useful in many other contexts as well. Unfortunately, by restricting ourselves to tensor products over vector spaces we miss many of the more interesting and nontrivial properties one obtains by studying tensor products of modules.

7.1 Extension of scalars

In this section we discuss a particular example of tensor products that can be useful in many situations. Namely, given an $F$-vector space $V$ and a field $K$ that contains $F$, can we form a $K$-vector space that contains a copy of $V$? Before we can do this, we need to address a basic question about forming vector spaces with a given basis.

We have seen early on that given a vector space $V$ over a field $F$, we can always find a basis for $V$. Before we can define tensor products, we need to address the very natural question of if we are given a set $X$, can we find a vector space that has $X$ as a basis? In fact, one can construct such a vector space and moreover the vector space we construct satisfies a universal property.

**Theorem 7.1.1.** Let $F$ be a field and $X$ a set. There is an $F$-vector space $VS_F(X)$ that has $X$ as a basis. Moreover, $VS_F(X)$ satisfies the following uni-
versal property. If $W$ is any $F$-vector space and $t : X \to W$ is any map of sets, there is a unique $T \in \text{Hom}_F(\text{VS}_F(X), W)$ so that $T(x) = t(x)$ for all $x \in X$, i.e., the following diagram commutes:

\[
\begin{array}{ccc}
X & \xrightarrow{\text{incl.}} & \text{VS}_F(X) \\
| & & | \\
\downarrow^t & & \downarrow^T \\
W & & \\
\end{array}
\]

**Proof.** This result should not be surprising. In fact, the universal property essentially amounts to the fact that a linear map is determined by its image on a basis.

If $X = \emptyset$, set $\text{VS}_F(X) = \{0\}$ and we are done. If $X \neq \emptyset$, let

\[
\text{VS}_F(X) = \left\{ \sum_{i \in I} a_i x_i : a_i \in F, a_i = 0 \text{ for all but finitely many } i \right\}
\]

where $I$ is an indexing set of the same cardinality as $X$. We define

\[
\sum_{i \in I} a_i x_i + \sum_{i \in I} b_i x_i = \sum_{i \in I} (a_i + b_i) x_i
\]

and

\[
c \left( \sum_{i \in I} a_i x_i \right) = \sum_{i \in I} ca_i x_i.
\]

It is easy to see these both lie in $\text{VS}_F(X)$ and that $\text{VS}_F(X)$ is a vector space under this addition and scalar multiplication. By definition we have $X$ spans $\text{VS}_F(X)$ and is linearly independent by construction, so $X$ is a basis for $\text{VS}_F(X)$.

It remains to show the universal property. Let $t : X \to W$. We define $T : \text{VS}_F(X) \to W$ by

\[
T \left( \sum_{i \in I} a_i x_i \right) = \sum_{i \in I} a_i t(x_i).
\]

This gives a well-defined linear map because $X$ is a basis. It is unique because any linear map that agrees with $t$ on $X$ must also agree with $T$ on $\text{VS}_F(X)$. This gives the result.

It is important to note in the above result that we treat $X$ as strictly a set. It may be the case that $X$ is itself a vector space, but when we form $\text{VS}_F(X)$ we ignore any structure that $X$ may have.

**Example 7.1.2.** Let $F = \mathbb{R}$ and $X = \mathbb{R}$. We know that as an $\mathbb{R}$-vector space $\mathbb{R}$ is 1-dimensional with 1 as a basis. However, if we consider $\text{VS}_\mathbb{R}(\mathbb{R})$ as an $\mathbb{R}$-vector space, this is infinite dimensional with every element of $\mathbb{R}$ as a basis element. For example, an element of $\text{VS}_\mathbb{R}(\mathbb{R})$ is $2 \cdot 3 + 4 \cdot \pi$ where 3 and $\pi$ are considered as elements of $X$. This sum does not simplify in $\text{VS}_\mathbb{R}(\mathbb{R})$. 


7.1. EXTENSION OF SCALARS  

We now turn our attention to the main topic of this section, extension of scalars. Let \( V \) be an \( F \)-vector space and let \( K/F \) be an extension of fields, i.e., \( F \subset K \) and \( K \) is a field. We can naturally consider \( K \) as an \( F \)-vector space as well. When studying problems over the vector space \( V \), say looking for the Jordan canonical form of a linear transformation, the field \( F \) may not be big enough. In this case we would like to change the scalars of \( V \) so that we are working over a bigger field. We can construct such a vector space by forming products to check things for elements \( i \) for all but finitely many \( i \). This is a \( K \)-vector space, but it does not take into account the \( F \)-vector space structure of \( K \) or \( V \) at all. This means it is much too large to be useful for anything. We will cut this space down to something useful by taking the quotient by an appropriate subspace. We define a subspace \( \text{Rel}_K(X) \) of \( VS_K(X) \) by setting \( \text{Rel}_K(X) \) to be the \( K \)-span of the elements

1. \( (a_1 + a_2, v) - (a_1, v) - (a_2, v) \) for all \( a_1, a_2 \in K, \ v \in V \);
2. \( (a, v_1 + v_2) - (a, v_1) - (a, v_2) \) for all \( a \in K, \ v_1, v_2 \in V \);
3. \( (ca, v) - (a, cv) \) for all \( c \in F, a \in K \) and \( v \in V \);
4. \( a_1(a_2, v) - (a_1a_2, v) \) for all \( a_1, a_2 \in K, \ v \in V \).

Note the \( \text{Rel} \) stands for “relations” as we are using this quotient to identify some relations we are requiring be satisfied. We now consider the quotient space \( K \otimes_F V = VS_K(X)/\text{Rel}_K(X) \). The fact that \( K \otimes_F V \) is a vector space follows immediately because we have constructed it as the quotient of two vector spaces. Note the \( F \) subscript on the \( \otimes \) indicates the original field that \( K \) and \( V \) are vector spaces over. Given an element \( (a, v) \in VS_K(X) \), we denote the equivalence class \( (a, v) + \text{Rel}_K(X) \) by \( a \otimes v \). Observe from the definition of \( K \otimes_F V \) that we have

1. \( (a_1 + a_2) \otimes v = a_1 \otimes v + a_2 \otimes v \) for all \( a_1, a_2 \in K, \ v \in V \);
2. \( a \otimes (v_1 + v_2) = a \otimes v_1 + a \otimes v_2 \) for all \( a \in K, \ v_1, v_2 \in V \);
3. \( ca \otimes v = a \otimes cv \) for all \( c \in F, a \in K \) and \( v \in V \);
4. \( a_1(a_2 \otimes v) = (a_1a_2) \otimes v \) for all \( a_1, a_2 \in K, \ v \in V \).

It is very important to note that a typical element in \( K \otimes_F V \) is of the form \( \sum c_i(a_i \otimes v_i) \) where \( c_i \) is 0 for all but finitely many \( i \). Note we can combine the \( c_i \) and \( a_i \) so in this case a typical element is really just \( \sum a_i \otimes v_i \) where \( a_i = 0 \) for all but finitely many \( i \). It is a common mistake when working with tensor products to check things for elements \( a \otimes v \) without remembering this is not a typical element! One other important point is that since \( K \otimes_F V \) is a quotient, the elements are all equivalence classes. So one must be very careful to check things are well-defined when working with tensor products!
We have that $0 \otimes v = c \otimes 0 = 0 \otimes 0 = 0_{K \otimes_F V}$ for all $v \in V$, $c \in K$. First, note that since $0 \in K$ is also in $F$, we have $0 \otimes v = 0 \otimes 0v = 0 \otimes 0$ and $c \otimes 0 = 0c \otimes 0 = 0 \otimes 0$. Now we use uniqueness of the additive identity in a vector space and the fact that $0 \otimes 0$ is clearly an additive identity in $V$.

**Example 7.1.3.** Let $K = \mathbb{C}$, $F = \mathbb{R}$ and let $V$ be an $F$-vector space. We consider some elements in $\mathbb{C} \otimes_{\mathbb{R}} V$ and how they simplify. We have

$$i((2 + i) \otimes v) + 6 \otimes v = (-1 + 2i) \otimes v + 6 \otimes v = (5 + 2i) \otimes v.$$ 

The reason we were able to combine the terms is the term coming from $V$, namely $v$, was the same in both. Similarly, we have

$$2 \otimes v_1 + 2 \otimes v_2 = 2 \otimes (v_1 + v_2).$$

One other thing we can use to simplify is to bring a scalar that is in $F = \mathbb{R}$ across the tensor. So we have

$$2 \otimes v_1 + 7 \otimes v_2 = 1 \otimes 2v_1 + 1 \otimes 7v_2 = 1 \otimes (2v_1 + 7v_2).$$

However, if the first terms in the tensors lies in $K = \mathbb{C}$ and are unequal, and the second terms are not equal, the sum cannot be combined into one term. For instance, we have

$$(2 + i) \otimes v_1 + 3i \otimes v_2 = 2 \otimes v_1 + i \otimes v_1 + 3i \otimes v_2 = 1 \otimes 2v_1 + i \otimes v_1 + i \otimes 3v_2.$$ 

The only way the first two terms could be combined is if $v_1 = 0$, and for the second two one needs $v_1 = 3v_2$. Which means to combine this to one term the vectors must all be 0.

Since the original goal was to extend $V$ to be a vector space over $K$, it is important to check that there is actually a subspace of $K \otimes_F V$ that is isomorphic to $V$ as an $F$-vector space.

**Proposition 7.1.4.** Let $K/F$ be a field extension and $V$ an $F$-vector space. Then $K \otimes_F V$ contains a subspace isomorphic to $V$ as an $F$-vector space.

**Proof.** Let $B = \{v_i\}$ be an $F$-basis of $V$. Define $T \in \text{Hom}_F(V, K \otimes_F V)$ by setting $T(v_i) = 1 \otimes v_i$. Let $W$ be the image of $T$, i.e., $W$ is the $F$-span of $\{1 \otimes v_i\}$. By definition this gives $W$ is an $F$-subspace of $K \otimes_F V$. Note that this is not a $K$-subspace of $K \otimes_F V$. It is also clear that $T$ is a surjective linear map. It only remains to see that it is injective. Suppose $T(v) = 0$ for some $v \in V$. Then we have $1 \otimes v = 0 \otimes 0$. This is equivalent to $(1, v) - (0, 0) \in \text{Rel}_K(X)$. However, from the definition of $\text{Rel}_K(X)$ it is clear this is only the case if $v = 0$. Thus, $T$ is injective and so we have the result. 

\[\square\]
The $K$-vector space $K \otimes_F V$ is referred to as the \textit{extension of scalars of $V$ by $K$}. Thus, we have constructed a $K$-vector space that contains a copy of $V$ as an $F$-subspace. The following universal property shows that we have done as good as possible, i.e., the $K$-vector space $K \otimes_F V$ is the smallest $K$-vector space that contains $V$ as an $F$-subspace.

**Theorem 7.1.5.** Let $K/F$ be an extension of fields, $V$ an $F$-vector space, and \( \iota : V \to K \otimes_F V \) given by \( \iota(v) = 1 \otimes v \). Let $W$ be a $K$-vector space and \( t \in \text{Hom}_F(V, W) \). There is a unique \( T \in \text{Hom}_K(K \otimes_F V, W) \) so that \( t = T \circ \iota \), i.e., the following diagram commutes

\[
\begin{array}{ccc}
V & \xrightarrow{\iota} & K \otimes_F V \\
& \downarrow t & \\
& W, \quad T \\
\end{array}
\]

Conversely, if \( T \in \text{Hom}_K(K \otimes_F V, W) \) then \( T \circ \iota \in \text{Hom}_F(V, W) \).

**Proof.** Let \( t \in \text{Hom}_F(V, W) \). Consider the $K$-vector space $\text{VS}_K(K \times V)$. Since $W$ is a $K$-vector space, we have a map $(c, v) \mapsto ct(v)$. This extends to a map \( T : \text{VS}_K(K \times V) \to W \) by Theorem 7.1.1. Moreover, \( T \in \text{Hom}_K(\text{VS}_K(K \times V), W) \). It is now easy to check that \( T \) is 0 when restricted to $\text{Rel}_K(K \times V)$. Thus, we have \( T : K \otimes_F V \to W \) given by \( T(a \otimes v) = at(v) \).

Observe we have

\[
cT(a \otimes v) = c(at(v))
\]

\[
= (ca)t(v)
\]

\[
= T(c(a \otimes v))
\]

for all \( a, c \in K \) and \( v \in V \). It is also easy to see that \( T \) is additive, and so \( T \in \text{Hom}_K(K \otimes_F V, W) \). This gives the existence of \( T \) and that the diagram commutes.

We have that \( K \otimes_F V \) is given by the $K$-span of elements of the form \( 1 \otimes v \), so any $K$-linear map on $K \otimes_F V$ is determined by its image on these elements. Since \( T(1 \otimes v) = t(v) \), we get \( T \) is uniquely determined by \( t \). This gives the uniqueness statement.

It is now an easy exercise to check that for any \( T \in \text{Hom}_K(K \otimes_F V, W) \), one has \( T \circ \iota \in \text{Hom}_F(V, W) \).

**Example 7.1.6.** Let $K/F$ be an extension of fields. In this example we show that $K \otimes_F F \cong K$ as $K$-vector spaces. We have a natural inclusion map \( i : F \to K \). We write \( i : F \to K \otimes_F F \) as above. From the previous result we obtain a unique $K$-linear map \( T : K \otimes_F F \to K \) so that the following diagram commutes.
Thus, we see \( T(1 \otimes x) = x \). Moreover, since \( T \) is \( K \)-linear this completely determines \( T \) because for \( \sum a_i \otimes x_i \in K \otimes_F F \), we have
\[
T \left( \sum a_i \otimes x_i \right) = \sum T(a_i \otimes x_i) \\
= \sum T(a_i(1 \otimes x_i)) \\
= \sum a_i T(1 \otimes x_i).
\]

Define \( S : K \rightarrow K \otimes_F F \) by \( S(y) = y \otimes 1 \). We clearly have \( S \in \text{Hom}_K(K, K \otimes_F F) \) and we have
\[
S \circ T(y \otimes 1) = S(y) = y \otimes 1
\]
and
\[
T \circ S(y) = T(y \otimes 1) = y.
\]
Thus, we have \( T^{-1} = S \) and so \( K \otimes_F F \cong K \) as \( K \)-vector spaces.

**Example 7.1.7.** More generally, let \( K/F \) be an extension of fields and let \( V \) be an \( n \)-dimensional \( F \)-vector space. We claim that \( K \otimes_F V \cong K^n \) as \( K \)-vector spaces. We begin by using the universal property to obtain a \( K \)-linear map from \( K \otimes_F V \) to \( K^n \). Let \( B = \{v_1, \ldots, v_n\} \) be a basis of \( V \) and define an \( F \)-linear map \( t : V \rightarrow K^n \) by \( t(v_i) = e_i \) where \( e_i \) is the standard basis element of \( K^n \).
Using the universal property we obtain a \( K \)-linear map \( T : K \otimes_F V \rightarrow K^n \) given by \( T(1 \otimes v_i) = e_i \) for \( i = 1, \ldots, n \). Define a linear map \( S : K^n \rightarrow K \otimes_F V \) by \( S(e_i) = 1 \otimes v_i \). We know to define a linear map we only need to specify where it sends a basis, so this gives a well-defined \( K \)-linear map. Moreover, it is easy to see that \( S \) and \( T \) are inverse maps. Thus, \( K \otimes_F V \cong K^n \). Moreover, since \( S \) is an isomorphism and \( \{e_1, \ldots, e_n\} \) is a basis for \( K^n \), we see \( \{1 \otimes v_1, \ldots, 1 \otimes v_n\} \) is a basis of \( K \otimes_F V \).

### 7.2 Tensor products of vector spaces

We now turn our attention to tensor products of two \( F \)-vector spaces. One motivation for defining tensor products is it gives a way to form a new vector space where one has the “product” of elements from the original vector space. We began with a fairly simple case to help motivate the more general definition. Now that we have seen a particular example of tensor products in the previous section, namely, extension of scalars, we return to the more general situation of forming the tensor product of two \( F \)-vector spaces \( V \) and \( W \). Since the set-up is very similar to what was done in the previous section many of the details will be left to the reader.
Let $V$ and $W$ be $F$-vector spaces. (If we allow them to be $K$-vector spaces as well, we will recover what was done in the previous section.) Consider $X = V \times W = \{(v, w) : v \in V, w \in W\}$ and let $\text{VS}_F(V \times W)$ be the associated $F$-vector space as above. Let $\text{Rel}_F(V \times W)$ be the subspace of $\text{VS}_F(V \times W)$ given by the $F$-span of

1. $(v_1 + v_2, w) - (v_1, w) - (v_2, w)$ for all $v_1, v_2 \in V, w \in W$;
2. $(v, w_1 + w_2) - (v, w_1) - (v, w_2)$ for all $v \in V, w_1, w_2 \in W$;
3. $(cv, w) - (v, cw)$ for all $c \in F, v \in V, w \in W$;
4. $c(v, w) - (cv, w)$ for all $c \in F, v \in V, w \in W$.

Then, as above, we define $V \otimes_F W = \text{VS}_F(V \times W)/\text{Rel}_F(V \times W)$. As before, denote the equivalence class containing $(v, w)$ by $v \otimes w$.

We have $V \otimes_F W$ is an $F$-vector space with elements of the form $\sum_i c_i(v_i \otimes w_i)$. Note, we can represent any such element by combining $c_i$ with the $v_i \otimes w_i$, so elements can be represented in the form $\sum v_i \otimes w_i$ for $v_i \in V, w_i \in W$ with only finitely many terms being nonzero. Elements of the form $v \otimes w$ are called pure tensors.

We would like a similar universal property to the one we gave above for $K \otimes_F V$. However, this requires a new type of map. We defined bilinear maps $\text{Hom}_F(V; V; \text{F})$ in Chapter 5. We now extend that definition.

**Definition 7.2.1.** Let $V, W,$ and $U$ be $F$-vector spaces. We say a map $t : V \times W \to U$ is an $F$-bilinear map (or just bilinear map if $F$ is clear from context), and write $t \in \text{Hom}_V(V; W; U)$, if it satisfies

1. $t(cv_1 + v_2, w) = ct(v_1, w) + t(v_2, w)$ for all $c \in F, v_i \in V, w \in W$;
2. $t(v, cw_1 + w_2) = ct(v, w_1) + t(v, w_2)$ for all $c \in F, v \in V, w_i \in W$.

The point of bilinear maps as opposed to linear maps is they treat $V$ and $W$ as vector spaces, but they do not use the fact that we can define a vector space structure on $V \times W$. They keep $V$ and $W$ separate in terms of the algebraic structure; they are linear in each variable separately. This allows us to give the appropriate universal property for $V \otimes_F W$.

**Theorem 7.2.2.** Let $U, V,$ and $W$ be $F$-vector spaces. Define a map $\iota : V \times W \to V \otimes_F W$ by $\iota(v, w) = v \otimes w$. Then

1. $\iota \in \text{Hom}_F(V; W; V \otimes_F W)$, i.e., $\iota$ is $F$-bilinear;
2. if $T \in \text{Hom}_F(V \otimes_F W; U)$, then $T \circ \iota \in \text{Hom}_F(V; W; U)$;
3. if $t \in \text{Hom}_F(V; W; U)$, then there is a unique $T \in \text{Hom}_F(V \otimes_F W; U)$ so that $t = T \circ \iota$.

Equivalently, we can write the correspondence by saying we have a bijection between $\text{Hom}_F(V; W; U)$ and $\text{Hom}_F(V \otimes_F W; U)$ so that the following diagram commutes:
**Proof.** 1. This part follows immediately from the definition and properties of the tensors.

2. We will show that \( t = T \circ \iota \) is linear in the first variable; linear in the second variable is the same argument. Let \( v_1, v_2 \in V \), \( w \in W \), and \( c \in F \). We have

\[
t(cv_1 + v_2, w) = T \circ \iota((cv_1 + v_2) \otimes w) = T((cv_1 \otimes w + v_2 \otimes w) = cT(v_1 \otimes w) + T(v_2 \otimes w) = ct(v_1, w) + t(v_2, w).
\]

Thus, \( t \) is bilinear in the first variable.

3. Let \( t \in \text{Hom}_F(V, W; U) \). We have that \( t \) vanishes on the elements of \( \text{Rel}_F(V \times W) \) by the properties of being a bilinear map. Thus, we obtain a well-defined linear map \( T : V \otimes_F W \to U \) so that \( T(v \otimes w) = t(v, w) \). Thus, we have a unique map \( T \in \text{Hom}_F(V \otimes_F W, U) \) satisfying \( T \circ \iota(v, w) = T(v \otimes w) = t(v, w) \), as claimed.

**Exercise 7.2.3.** Show that \( \text{Hom}_F(V, W; U) \) is isomorphic to \( \text{Hom}_F(V \otimes_F W, U) \) as \( F \)-vector spaces.

We now illustrate how this universal property can be used to prove basic properties about tensor products. It is extremely powerful because it allows one to define a bilinear map on \( V \times W \) and obtain a linear map on \( V \otimes_F W \). The reason this is so nice is to define a map directly on \( V \otimes_F W \) one must also check the map is well-defined, which can be very tedious.

**Corollary 7.2.4.** Let \( V \) and \( W \) be \( F \)-vector spaces with bases \( B = \{ v_1, \ldots, v_m \} \) and \( C = \{ w_1, \ldots, w_n \} \) respectively. Then \( \{ v_i \otimes w_j \}_{1 \leq i \leq m, 1 \leq j \leq n} \) is a basis for \( V \otimes_F W \).

In particular, \( \dim_F(V \otimes_F W) = \dim_F V \dim_F W \).

**Proof.** We show this result by showing that \( V \otimes_F W \cong \text{Mat}_{m,n}(F) \cong F^{mn} \). Define a map \( t : V \times W \to F^{mn} \) by \( t((v_i, w_j)) = e_{i,j} \). First, observe this is enough to define a bilinear map. Given \( v \in V \) and \( w \in W \), write \( v = \sum_{i=1}^m a_i v_i \) and \( w = \sum_{j=1}^n b_j w_j \), and define \( t(v \otimes w) \) as follows:

\[
t(v \otimes w) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j e_{i,j}.
\]

This defines \( t \) as a bilinear map, and it is straightforward to verify that it satisfies the universal property of tensor products.
and \( w = \sum_{j=1}^{m} b_j w_j \). Then
\[
 t(v, w) = t \left( \sum_{i=1}^{m} a_i v_i, \sum_{j=1}^{n} b_j w_j \right)
 = \sum_{i=1}^{m} a_i t(v_i, \sum_{j=1}^{n} b_j w_j)
 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j t(v_i, w_j).
\]

Thus, just as it was enough to define a linear map on a basis, it is enough to specify the values of a bilinear map on elements \((v_i, w_j)\) for \(v_i \in B\) and \(w_j \in C\). We now apply the universal property to obtain an \(F\)-linear map \(T: V \otimes_F W \to \text{Mat}_{m,n}(F)\) that satisfies \(T(v_i \otimes w_j) = e_{i,j}\). We can define an \(F\)-linear map \(S: \text{Mat}_{m,n}(F) \to V \otimes_F W\) by \(S(e_{i,j}) = v_i \otimes w_j\): It is clear that \(S\) and \(T\) are inverse maps, so we obtain \(V \otimes_F W \cong \text{Mat}_{m,n}(F) \cong F^{mn}\). Moreover, since \(\{e_{i,j}\}\) forms a basis for \(\text{Mat}_{m,n}(F)\) and \(S\) is an isomorphism, we have that \(\{v_i \otimes w_j\}\) is a basis of \(V \otimes_F W\).

\[\square\]

**Example 7.2.5.** The vector space \(C \otimes_R C \cong \mathbb{R}^4\). A basis for \(C \otimes_R C\) is given by \(\{1 \otimes 1, 1 \otimes i, i \otimes 1, i \otimes i\}\).

The following results follow immediately from this corollary.

**Corollary 7.2.6.** Let \(U, V\) and \(W\) be \(F\)-vector spaces. We have
\begin{enumerate}
  \item \(V \otimes_F W \cong W \otimes_F V\);
  \item \((U \otimes_F V) \otimes_F W \cong U \otimes_F (V \otimes_F W)\).
\end{enumerate}

Given \(F\)-vector spaces \(U, V\) and \(W\), the next theorem follows immediately from Corollary 7.2.4. However, we give a direct proof. The benefits of this are it shows the spirit of how such things would be proven for modules, and it also allows us to easily see what the isomorphism is and show it is unique.

**Theorem 7.2.7.** Let \(U, V\) and \(W\) be \(F\)-vector spaces. There is a unique isomorphism
\[
 (U \oplus V) \otimes_F W \xrightarrow{\sim} (U \otimes_F W) \oplus (V \otimes_F W)
 (u, v) \otimes w \mapsto (u \otimes w, v \otimes w).
\]

**Proof.** We will once again make use of the universal property to define the appropriate maps. Define
\[
 (U \oplus V) \times W \longrightarrow (U \otimes_F W) \oplus (V \otimes_F W)
 ((u, v), w) \mapsto (u \otimes w, v \otimes w).
\]
It is easy to see this map is bilinear, so the universal property gives a unique linear map

\[ T : (U \oplus V) \otimes_F W \rightarrow (U \otimes_F W) \oplus (V \otimes_F W) \]

\[(u, v) \otimes w \mapsto (u \otimes w, v \otimes w). \]

It now remains to define an inverse map. We begin by defining maps

\[ U \times W \rightarrow (U \oplus V) \otimes_F W \]

\[(u, w) \mapsto (u, 0) \otimes w \]

and

\[ V \times W \rightarrow (U \oplus V) \otimes_F W \]

\[(v, w) \mapsto (0, v) \otimes w. \]

The universal property applied to each of these maps gives

\[ S_1 : U \otimes_F W \rightarrow (U \oplus V) \otimes_F W \]

\[ u \otimes w \mapsto (u, 0) \otimes w \]

and

\[ S_2 : V \otimes_F W \rightarrow (U \oplus V) \otimes_F W \]

\[ v \otimes w \mapsto (0, v) \otimes w. \]

Combining these gives a linear map

\[ S : (U \otimes_F W) \oplus (V \otimes_F W) \rightarrow (U \oplus V) \otimes_F W \]

\[(u \otimes w_1, v \otimes w_2) \mapsto (u, 0) \otimes w_1 + (0, v) \otimes w_2. \]

It now only remains to check that these are inverse maps. Since these are linear maps it is enough to check it on pure tensors. We have

\[ S \circ T((u, v) \otimes w) = S(u \otimes w, v \otimes w) \]

\[ = (u, 0) \otimes w + (0, v) \otimes w \]

\[ = (u, v) \otimes w \]

and

\[ T \circ S(u \otimes w_1, v \otimes w_2) = T((u, 0) \otimes w_1 + (0, v) \otimes w_2) \]

\[ = T((u, 0) \otimes w_1) + T((0, v) \otimes w_2) \]

\[ = (u \otimes w_1, 0 \otimes w_1) + (0 \otimes w_2, v \otimes w_2) \]

\[ = (u \otimes w_1, 0) + (0, v \otimes w_2) \]

\[ = (u \otimes w_1, v \otimes w_2). \]

Thus, we have the desired isomorphism. \qed
7.2. TENSOR PRODUCTS OF VECTOR SPACES

We can use the tensor product to give a coordinate-free construction of the trace map. We begin with the following lemma

Lemma 7.2.8. Let \( V \) be a finite dimensional \( F \) vector space. Then \( V \otimes V^\vee \cong \text{Hom}_F(V,V) \).

Proof. It is clear since the dimensions of the spaces are the same that they are isomorphic, but for our purposes we need to know the specific map giving the isomorphism. We define a map \( t \)

\[
V \times V^\vee \to \text{Hom}_F(V,V) \quad (v, \varphi) \mapsto (w \mapsto \varphi(w)v).
\]

It is easy to check this is a bilinear map, so the universal property gives a map \( T : V \otimes F V^\vee \to \text{Hom}_F(V,V) \) so that \( T(v \otimes \varphi)(w) = \varphi(w)v \). It remains to check this map is an isomorphism. Since the dimensions of the spaces are the same, it is enough to show that \( T \) is injective. Let \( \mathcal{B} = \{ v_1, \ldots, v_n \} \) be a basis for \( V \) and \( \{ v_1^\vee, \ldots, v_n^\vee \} \) a dual basis for \( V^\vee \). Suppose that \( T \left( \sum_{i=1}^n a_{i,j} (v_i \otimes v_j^\vee) \right) = 0 \) in \( \text{Hom}_F(V,V) \). Thus, for \( v_m \), we must have

\[
0 = T \left( \sum_{i,j=1}^n a_{i,j} (v_i \otimes v_j^\vee) \right) (v_m) \\
= \sum_{i,j=1}^n a_{i,j} v_j^\vee(v_m) v_i \\
= \sum_{i=1}^n a_{i,m} v_i.
\]

However, since \( \mathcal{B} \) is a basis, this gives \( a_{i,m} = 0 \) for all \( i = 1, \ldots, n \). Since \( m \) was arbitrary, this gives \( a_{i,j} = 0 \) for all \( i, j \), i.e., \( T \) is injective. Thus, we have \( T : V \otimes F V^\vee \to \text{Hom}_F(V,V) \) is an isomorphism. \( \square \)

In \( \text{Hom}_F(V,V) \) one can compose maps (recall this corresponds to matrix multiplication.) Since \( V \otimes F V^\vee \cong \text{Hom}_F(V,V) \), there is a well-defined map \( (V \otimes F V^\vee) \times (V \otimes F V^\vee) \to (V \otimes F V^\vee) \) that corresponds to the composition of linear maps on \( V \). We claim this map is given by sending \( (v \otimes \varphi) \times (w \otimes \psi) \) to \( \varphi(w)v \otimes \psi \). Denote this map by \( \Psi \). We need to show the following diagram commutes:

\[
\begin{array}{ccc}
(V \otimes F V^\vee) \times (V \otimes F V^\vee) & \xrightarrow{\Psi} & V \otimes V^\vee \\
\tau \times \tau & \downarrow & \\
\text{Hom}_F(V,V) \times \text{Hom}_F(V,V) & \xrightarrow{\text{comp}} & \text{Hom}_F(V,V).
\end{array}
\]

Let \( (v \otimes \varphi) \times (w \otimes \psi) \in (V \otimes F V^\vee) \times (V \otimes F V^\vee) \). We compute the image of this in \( \text{Hom}_F(V,V) \) by going in each direction of the diagram. If we first apply
7.2. TENSOR PRODUCTS OF VECTOR SPACES

CHAPTER 7.

\[ T \times T \] and then composition we obtain for each \( x \in V \)
\[
T(v \otimes \varphi) \circ T(w \otimes \psi)(x) = T(v \otimes \varphi)(T(w \otimes \psi)(x))
\]
\[
= T(v \otimes \varphi)(\psi(x)w)
\]
\[
= \psi(x)T(v \otimes \varphi)(w)
\]
\[
= \psi(x)\varphi(w)v.
\]

Going in the other direction we obtain
\[
T(\Psi((v \otimes \varphi) \times (w \otimes \psi)))(x) = T(\varphi(w)(v \otimes \psi))(x)
\]
\[
= \varphi(w)T(v \otimes \psi)(x)
\]
\[
= \varphi(w)\psi(x)v.
\]

Since the diagram commutes, we have that \( \Psi \) is the map corresponding to composition of functions.

We now apply these results to the trace map. Let \( T \in \text{Hom}_F(V, V) \). The trace map from undergraduate linear algebra is defined relative to a choice of basis. Let \( B = \{v_1, \ldots, v_n\} \) be a basis of \( V \) and \( A = (a_{ij}) = [T]_B \) the associated matrix. Let \( B^\vee = \{v^\vee_1, \ldots, v^\vee_n\} \) be the dual basis. It is straightforward to check that
\[
a_{ij} = v^\vee_i(T(v_j))
\]
for all \( 1 \leq i, j \leq n \). In particular, if we use \( \text{Tr} \) to denote the familiar trace from undergraduate linear algebra, then
\[
\text{Tr}(A) = \sum_{i=1}^{n} v^\vee_i(T(v_j)).
\]

This definition depends upon first choosing coordinates on \( V \). We now give a coordinate free definition of the trace and show it agrees with this definition upon choosing a basis. Consider the linear map \( V \otimes_F V^\vee \to F \) induced from the bilinear map \( V \times V^\vee \to F \) given by \( (v, \varphi) \mapsto \varphi(v) \). Since \( V \otimes_F V^\vee \cong \text{Hom}_F(V, V) \), this gives a map \( \text{tr} : \text{Hom}_F(V, V) \to F \). Note that this map does not depend upon any choice of basis. We have that the elements \( v_k \otimes v^\vee_l \) form a basis for \( V \otimes_F V^\vee \), and so the elements \( T(v_k \otimes v^\vee_l) \) form a basis of \( \text{Hom}_F(V, V) \).

We compute \( \text{Tr} \) on these elements with respect to the basis \( B \).

\[
\text{Tr}(T(v_k \otimes v^\vee_l)) = \sum_{i=1}^{n} v^\vee_i(T(v_k \otimes v^\vee_l))(v_i)
\]
\[
= \sum_{i=1}^{n} v^\vee_i(v^\vee_l(v_i)v_k)
\]
\[
= \sum_{i=1}^{n} v^\vee_i(v_i)v^\vee_l(v_k)
\]
\[
= v^\vee_l(v_k)
\]
\[
= \begin{cases} 
1 & k = l \\
0 & k \neq l 
\end{cases}
\]
On the other hand, we have
\[
\text{tr}(v_k \otimes v_l^\vee) = v_l^\vee(v_k) = \begin{cases} 1 & k = l \\ 0 & k \neq l. \end{cases}
\]

Since these two maps agree on basis elements, they are the same. We will use \(\text{Tr}\) to denote the trace map on \(V \otimes V^\vee\) and \(\text{Hom}_F(V, V)\) as well since we have seen they specialize to the undergraduate definition upon choosing coordinates. This coordinate-free definition allows us to easily prove basic properties of the trace map. For instance, the construction we constructed the coordinate-free trace to be a linear map, so for \(c \in F\) and \(A, B \in \text{Mat}_n(F)\), we have
\[
\text{Tr}(cA + B) = c\text{Tr}(A) + \text{Tr}(B).
\]

**Corollary 7.2.9.** Let \(A, B \in \text{Mat}_n(F)\). Then \(\text{Tr}(AB) = \text{Tr}(BA)\).

**Proof.** We prove this using the coordinate-free definition; the result in terms of matrices follows immediately upon choosing coordinates. Observe that the maps
\[
\text{Hom}_F(V, V) \times \text{Hom}_F(V, V) \to F \\
(A, B) \mapsto \text{Tr}(AB)
\]
and
\[
\text{Hom}_F(V, V) \times \text{Hom}_F(V, V) \to F \\
(A, B) \mapsto \text{Tr}(BA)
\]
are both bilinear forms. Thus, it is enough to show they agree on pure tensors \(v \otimes \varphi \in V \otimes V^\vee\). Let \(v \otimes \varphi\) and \(w \otimes \psi\) be two such pure tensors. We must show that
\[
\text{Tr}(T(v \otimes \varphi) \circ T(w \otimes \psi)) = \text{Tr}(T(w \otimes \psi) \circ T(v \otimes \varphi)).
\]
Recall we showed above that composition of the linear maps \(T(v \otimes \varphi) \circ T(w \otimes \psi)\) corresponds under the isomorphism identifying \(V \otimes F V^\vee\) with \(\text{Hom}_F(V, V)\) to \(\varphi(w)v \otimes \psi\). Using this, we have
\[
\text{Tr}(T(v \otimes \varphi) \circ T(w \otimes \psi)) = \text{Tr}(\varphi(w)v \otimes \psi) \\
= \varphi(w)\text{Tr}(v \otimes \psi) \\
= \varphi(w)\psi(v)
\]
and
\[
\text{Tr}(T(w \otimes \psi) \circ T(v \otimes \varphi)) = \text{Tr}(\psi(v)w \otimes \varphi) \\
= \psi(v)\text{Tr}(w \otimes \varphi) \\
= \psi(v)\varphi(w).
\]
Thus, we have the result. \(\square\)
Before we can introduce the exterior product of a vector space, we need to consider multilinear maps and tensor products of finitely many vector spaces. In particular, we saw above that given $F$-vector spaces $U, V$ and $W$, we have $U \otimes_F (V \otimes_F W) \cong (U \otimes_F V) \otimes_F W$ as $F$-vector spaces. Thus, it makes sense to just write $U \otimes_F V \otimes_F W$. By induction, given $F$-vector spaces $V_1, \ldots, V_n$, it makes sense to write $V \otimes_F V_2 \otimes_F \cdots \otimes_F V_n$. We will be particularly interested in the case when $V_i = V$ for all $i$. In this case we write $V \otimes^n V$. If there is any chance of confusion we write $V \otimes^n_F V$.

We now define multilinear maps; these are the appropriate maps to consider in this context.

Definition 7.2.10. Let $V_1, \ldots, V_n$ and $W$ be $F$-vector spaces. A map 
\[ t : V_1 \times \cdots \times V_n \to W \]
is said to be multilinear if $t$ is linear in each variable separately. We denote the set of multilinear maps by $\text{Hom}_F(V_1, \ldots, V_n; W)$.

Exercise 7.2.11. Show that $\text{Hom}_F(V_1, \ldots, V_n; W)$ is an $F$-vector space.

For tensor products of several vector spaces we have a universal property as well; bilinear maps are just replaced by multilinear maps.

Theorem 7.2.12. Let $V_1, \ldots, V_n$ be $F$-vector spaces. Define 
\[ t : V_1 \times \cdots \times V_n \to V_1 \otimes_F \cdots \otimes_F V_n \]
\[ (v_1, \ldots, v_n) \mapsto v_1 \otimes \cdots \otimes v_n. \]

Then we have
1. For every $T \in \text{Hom}_F(V_1 \otimes_F \cdots \otimes_F V_n; W)$ the map $T \circ t \in \text{Hom}_F(V_1, \ldots, V_n; W)$.
2. For every $t \in \text{Hom}_F(V_1, \ldots, V_n; W)$ there is a unique $T \in \text{Hom}_F(V_1 \otimes_F \cdots \otimes_F V_n, W)$ so that $t = T \circ t$.

The proof of this theorem is left as an exercise as it follows from the same type of arguments used in the case $n = 2$. Note the theorem can be restated by saying there is a bijection between $\text{Hom}_F(V_1, \ldots, V_n; W)$ and $\text{Hom}_F(V_1 \otimes_F \cdots \otimes_F V_n, W)$ so that the following diagram commutes

\[ \begin{array}{ccc}
V_1 \times \cdots \times V_n & \xrightarrow{t} & V_1 \otimes_F \cdots \otimes_F V_n \\
& \downarrow{T} & \\
& W. & \\
\end{array} \]

Corollary 7.2.13. Let $V_1, \ldots, V_k$ be $F$-vector spaces of dimension $n_1, \ldots, n_k$. Let $B_i = \{ v^i_1, \ldots, v^i_{n_i} \}$ be a basis of $V_i$. Show that $\{ v^1_1 \otimes \cdots \otimes v^k_{n_k} \}$ is a basis for $V_1 \otimes_F \cdots \otimes_F V_k$. In particular, show that
\[ \dim_F(V_1 \otimes_F \cdots \otimes_F V_k) = \prod_{j=1}^k \dim_F V_i. \]

Proof. See homework problems.
7.3 Alternating forms, exterior powers, and the determinant

In this section we will define exterior powers of a vector space and see how studying these gives a very natural definition of the determinant.

Let $V$ be a finite dimensional $F$-vector space and let $k$ be a positive integer. We begin by defining the $k^{th}$ exterior power of $V$.

**Definition 7.3.1.** Let $A_k(V)$ of $V^\otimes k$ be the subspace spanned by $v_1 \otimes \cdots \otimes v_k$ where $v_i = v_j$ for some $i \neq j$. The $k^{th}$ exterior power of $V$ is the quotient space $\Lambda^k(V) = V^\otimes k / A_k(V)$.

We denote elements of $\Lambda^k(V)$ by $v_1 \wedge \cdots \wedge v_k = v_1 \otimes \cdots \otimes v_k + A_k(V)$. Note that we have $v_1 \wedge \cdots \wedge v_k = 0$ if $v_i = v_j$ for any $i \neq j$. Thus, for $v, w \in V$ we have

$$0 = (v + w) \wedge (v + w) = v \wedge v + v \wedge w + w \wedge v + w \wedge w = v \wedge w + w \wedge v.$$

This gives that in $\Lambda^2(V)$ we have $v \wedge w = -w \wedge v$. This can be used for higher degree exterior powers as well. Namely, we have

$$v_1 \wedge \cdots \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_k = -v_1 \wedge \cdots v_{i+1} \wedge v_i \wedge \cdots \wedge v_k.$$

This will be very useful when proving the universal property for exterior powers.

Before we can state that, we need to define the appropriate maps.

**Definition 7.3.2.** Let $V$ and $W$ be $F$-vector spaces and let $t \in \text{Hom}_F(V, \ldots, V; W)$. We say $t$ is alternating if $t(v_1, \ldots, v_k) = 0$ whenever $v_i = v_{i+1}$ for some $1 \leq i \leq k - 1$. We denote the set of alternating maps by $\text{Alt}_F^k(V; W)$. We set $\text{Alt}_F^0(V; W) = F$.

Note that in the case that $W = F$ and $k = 2$, an alternating map is just a skew-symmetric bilinear form.

**Exercise 7.3.3.**

1. Show that $\text{Alt}_F^k(V; W)$ is an $F$-subspace of $\text{Hom}_F(V, \ldots, V; W)$.
2. Show that $\text{Alt}_F^k(V; F) = \text{Hom}_F(V, F) = V^\vee$.
3. If $\dim_F V = n$, show $\text{Alt}_F^k(V; F) = 0$ for all $k > n$.

**Theorem 7.3.4.** Let $V$ and $W$ be $F$-vector spaces and $k$ a positive integer. Define a map $\iota : V \times \cdots \times V \to \Lambda^k(V)$ by $\iota(v_1, \ldots, v_k) = v_1 \wedge \cdots \wedge v_k$. Then

1. $\iota \in \text{Alt}_F^k(V; \Lambda^k(V))$, i.e., $\iota$ is alternating;
2. if $T \in \text{Hom}_F(\Lambda^k(V), W)$, then $T \circ \iota \in \text{Alt}_F^k(V \times \cdots \times V; W)$;
3. if $t \in \text{Alt}_F^k(V \times \cdots \times V; W)$, then there is a unique $T \in \text{Hom}_F(\Lambda^k(V), W)$ so that $t = T \circ \iota$.
Equivalently, we can write the correspondence by saying we have a bijection between $\text{Alt}^k(V \times \cdots \times V; W)$ and $\text{Hom}_F(\Lambda^k(V), W)$ so that the following diagram commutes:

$$
\begin{array}{ccc}
V \times \cdots \times V & \xrightarrow{\iota} & \text{Alt}^k(V) \\
& \downarrow T & \\
& W. & \\
\end{array}
$$

Proof. This essentially follows from Theorem 7.2.12. To see $\iota$ is alternating, observe that it is the composition of the multilinear map used in the universal property of the tensor product composed with projection onto the quotient. This gives $\iota$ is multilinear. It vanishes on elements of the form $(v_1, \ldots, v_k)$ with $v_i = v_j$ for $i \neq j$ by the definition of the exterior power. This gives the first statement.

We now show $t = T \circ \iota$ is alternating. Let $c \in F$ and $v_1, \ldots, v_k, v'_1 \in V$. We have

$$
t(av_1 + v'_1, v_2, \ldots, v_k) = T((av_1 + v'_1) \wedge v_2 \wedge \cdots \wedge v_k)
$$

$$
= T(av_1 \wedge v_2 \wedge \cdots \wedge v_k + v'_1 \wedge v_2 \wedge \cdots \wedge v_k)
$$

$$
= aT(v_1 \wedge v_2 \wedge \cdots \wedge v_k) + T(v'_1 \wedge v_2 \wedge \cdots \wedge v_k)
$$

$$
= at(v_1 \wedge v_2 \wedge \cdots \wedge v_k) + t(v'_1 \wedge v_2 \wedge \cdots \wedge v_k).
$$

This shows $t$ is multilinear. To see it is also alternating, one just uses $\iota$ is alternating.

Suppose now we have $t \in \text{Alt}^k(V \times \cdots \times V; W)$. Since the alternating forms are a subset of the multilinear forms, we obtain a linear map $S : V^\otimes k \to W$ so that $S(v_1 \otimes \cdots \otimes v_k) = t(v_1, \ldots, v_k)$. As $\Lambda^k(V)$ is a quotient of $V^\otimes k$, we obtain a linear map $T : \Lambda^k(V) \to W$ given by composing $S$ with the projection map, i.e., $T(v_1 \cdots v_i) = t(v_1, \ldots, v_i)$. It is now easy to check this map is unique because it agrees with $t$ and that the diagram above commutes. 

Example 7.3.5. Let $V$ be an $F$-vector space with $\dim_F V = 1$. Thus, given any nonzero $v \in V$, the set \{v\} forms a basis for $V$. Given any positive integer $k$, we can write elements of $\Lambda^k(V)$ as finite sums of elements of the form $a_1 v \wedge \cdots \wedge a_k v$. However, we know $a_1 v \wedge \cdots \wedge a_k v = (a_1 \cdots a_k) v \wedge \cdots \wedge v$ and since $v \wedge v = 0$, we have $a_1 v \wedge \cdots \wedge a_k v = 0$ for $k \geq 2$. Thus, we have $\Lambda^0(V) = F$, $\Lambda^1(V) \cong V$, and $\Lambda^k(V) = 0$ for $k \geq 2$.

Example 7.3.6. Let $V$ be a 2-dimensional $F$-vector space and let $B = \{v_1, v_2\}$ be a basis. Let $k$ be a positive integer. Elements of $\Lambda^k(V)$ consist of finite sums of the form $(a_1 v_1 + b_1 v_2) \wedge \cdots \wedge (a_k v_1 + b_k v_2)$ for $a_i, b_i \in F$. If $k \geq 3$, this sum is 0 because each term will have a term $v_i \wedge v_j \wedge v_1$, and since the dimension is...
2 one of these terms will be a repeat and so 0. If \( k = 2 \), a typical element can be written in the form
\[
(av_1 + bv_2) \wedge (cv_1 + dv_2) = adv_1 \wedge v_2 + bc v_2 \wedge v_1
= (ad - bc)v_1 \wedge v_2.
\]
Thus, in this case we have \( \Lambda^0(V) = F, \Lambda^1(V) \cong V \), and \( \Lambda^2(V) \cong F(v_1 \wedge v_2) \cong F \).

**Theorem 7.3.7.** Let \( V \) be an \( n \)-dimensional \( F \)-vector space. Let \( B = \{v_1, \ldots, v_n\} \) be a basis of \( V \). For \( k \leq n \) the vectors \( \{v_{i_1} \wedge \cdots \wedge v_{i_k} : 1 \leq i_1 \leq \cdots \leq i_k \leq n\} \) form a basis of \( \Lambda^k(V) \). For \( k > n \) we have \( \Lambda^k(V) = 0 \). In particular, \( \dim_F \Lambda^k(V) = \binom{n}{k} \).

**Proof.** This follows easily from the fact that \( \{v_{i_1} \otimes \cdots \otimes v_{i_k} : 1 \leq i_j \leq n\} \) is a basis for \( V^\otimes k \). Since \( \Lambda^k(V) \) is a quotient of \( V^\otimes k \) by \( \mathcal{A}_k(V) \), to find a basis it only amounts to finding a basis of \( \mathcal{A}_k(V) \). However, we clearly have any element \( v_{i_1} \otimes \cdots \otimes v_{i_k} \) with \( i_{j_1} = i_{j_2} \) for some \( j_1 \neq j_2 \) lies in \( \mathcal{A}_k(V) \). Moreover, we know the elements \( v_{i_1} \wedge \cdots \wedge v_{i_k} \) can be reordered by introducing a negative sign. This gives the result. \( \square \)

**Exercise 7.3.8.** Prove the above theorem directly from Theorem 7.3.4.

As was observed in the previous examples for particular cases, this theorem shows in general that for an \( n \)-dimensional \( F \)-vector space one has \( \Lambda^n(V) \) is of dimension 1 over \( F \) and has as a basis \( v_1 \wedge \cdots \wedge v_n \) for \( \{v_1, \ldots, v_n\} \) a basis of \( V \). This is the key point in defining the determinant of a linear map.

Let \( T \in \text{Hom}_F(V,W) \) with \( V \) and \( W \) finite dimensional vector spaces. The map \( T \) induces a map
\[
T^\otimes k : V^\otimes k \to V^\otimes k \quad v_1 \otimes \cdots \otimes v_k \mapsto T(v_1) \otimes \cdots \otimes T(v_k).
\]
It is easy to see that the generators of \( \mathcal{A}_k(V) \) are sent to generators of \( \mathcal{A}_k(W) \), so this descends to a map
\[
\Lambda^k(T) : \Lambda^k(V) \to \Lambda^k(W) \quad v_1 \wedge \cdots \wedge v_k \mapsto T(v_1) \wedge \cdots \wedge T(v_k).
\]

We now restrict to the case that \( V = W \) and \( \dim_F V = n \). Since \( \Lambda^n(V) \) is 1-dimensional over \( F \), we have \( \Lambda^n(T)(v) = (\det(T))v \) for all \( v \in \Lambda^n(V) \). Given a matrix \( A \in \text{Mat}_n(F) \), we define the determinant of \( A \) to be the determinant of the associated linear map \( T_A \).
7.3. ALTERNATING FORMS, EXTERIOR POWERS, AND THE DETERMINANT

One should note that it is clear from this definition of the determinant that there is no dependence on the choice of a basis of $V$ as none was used in defining the determinant.

**Lemma 7.3.10.** Let $S, T \in \text{Hom}_F(V, V)$. Then $\det(T \circ S) = \det(T) \det(S)$.

**Proof.** Let $v_1 \wedge \cdots \wedge v_n \in \Lambda^n(V)$ be any nonzero element (so a basis). We have

\[
\det(T \circ S)v_1 \wedge \cdots \wedge v_n = \Lambda^n(T \circ S)(v_1 \wedge \cdots \wedge v_n) \\
= T \circ S(v_1) \wedge \cdots \wedge T \circ S(v_n) \\
= T(S(v_1)) \wedge \cdots \wedge T(S(v_n)) \\
= \Lambda^n(T)(S(v_1) \wedge \cdots \wedge S(v_n)) \\
= \Lambda^n(T)(\Lambda^n(S)(v_1 \wedge \cdots \wedge v_n)) \\
= \Lambda^n(T)(\det(S)v_1 \wedge \cdots \wedge v_n) \\
= \det(S)\Lambda^n(T)(v_1 \wedge \cdots \wedge v_n) \\
= \det(S)\det(T)v_1 \wedge \cdots \wedge v_n.
\]

This gives the result. \qed

Of course, for this to be useful we want to show this is the same determinant that was defined in undergraduate linear algebra class. Before showing this in general we check the case $V$ has dimension 2.

**Example 7.3.11.** Let $V = F^2$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and let $T_A$ be the associated linear map. Thus, we have

\[
\Lambda^2(T)(e_1 \wedge e_2) = T(e_1) \wedge T(e_2) \\
= (ae_1 + ce_2) \wedge (be_1 + de_2) \\
= abe_1 \wedge e_1 + ade_1 \wedge e_2 + cbe_2 \wedge e_1 + cde_2 \wedge e_2 \\
= ade_1 \wedge e_2 + bce_2 \wedge e_1 \\
= (ad - bc)e_1 \wedge e_2.
\]

Thus, we see $\det(A) = ad - bc$, as one expects from undergraduate linear algebra.

**Exercise 7.3.12.** Let $A \in \text{Mat}_3(F)$. Show that the definition of $\det(A)$ given here matches the definition from undergraduate linear algebra.

We now want to prove in general that given a matrix $A \in \text{Mat}_n(F)$ one has $\det(A)$ is the same as the value from undergraduate linear algebra. Note that we can view $A$ as an element of $F^n \times \cdots \times F^n$ with each column of $A$ a vector in $F^n$. Thus, we can view

\[
\det : F^n \times \cdots \times F^n \to F.
\]

**Theorem 7.3.13.** The determinant function is in $\text{Alt}^n(F^n, F)$ and satisfies $\det(1_n) = 1$. 

171
Proof. We begin by checking bilinearity in the first variable; the other variables follow from the same argument. Write

\[ w_1 = a_{11}e_1 + \cdots + a_{n1}e_n \]
\[ \vdots \]
\[ w_n = a_{n1}e_1 + \cdots + a_{nn}e_n \]

and \( w = b_{11}e_1 + \cdots + b_{n1}e_n \) for \( e_1, \ldots, e_n \) the standard basis of \( F^n \). We want to show that for any \( c \in F \) we have

\[ \det(w_1 + cw, w_2, \ldots, w_n) = \det(w_1, w_2, \ldots, w_n) + c \det(w, w_2, \ldots, w_n). \]

To do this, we need to translate this into a statement about linear maps. Define \( T_1 : F^n \to F^n \) by \( T_1(e_i) = w_i \), so

\[ [T_1]e_n = A_1 = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}; \]

\( T_2 : F^n \to F^n \) by \( T_2(e_1) = w \) and \( T_2(e_j) = w_j \) for \( 2 \leq j \leq n \) so

\[ [T_2]e_n = A_2 = \begin{pmatrix} b_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}; \]

\( T_3 : F^n \to F^n \) by \( T_3(e_1) = w_1 + cw \) and \( T_3(e_j) = w_j \) for \( 2 \leq j \leq n \) so

\[ [T_3]e_n = A_3 = \begin{pmatrix} a_{11} + cb_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} + cb_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}. \]

Now observe we have

\[ \det(w_1, w_2, \ldots, w_n) = \det(T_1) \]
\[ \det(w, w_2, \ldots, w_n) = \det(T_2) \]
\[ \det(w_1 + cw, w_2, \ldots, w_n) = \det(T_3). \]

Thus, we just need to show that \( \det(T_3) = \det(T_1) + c \det(T_2) \). We have

\[
\det(T_3)e_1 \wedge \cdots \wedge e_n = \Lambda^n(T_3)(e_1 \wedge e_2 \wedge \cdots \wedge e_n) \\
= T_3(e_1) \wedge T(e_2) \wedge \cdots \wedge T(e_n) \\
= (w_1 + cw) \wedge w_2 \wedge \cdots \wedge w_n \\
= w_1 \wedge \cdots \wedge w_n + cw \wedge w_2 \wedge \cdots \wedge w_n \\
= T_1(e_1) \wedge \cdots \wedge T_1(e_n) + cT_2(e_1) \wedge \cdots \wedge T_2(e_n) \\
= \Lambda^n(T_1)(e_1 \wedge \cdots \wedge e_n) + c\Lambda^n(T_2)(e_1 \wedge \cdots \wedge e_n) \\
= \det(T_1)(e_1 \wedge \cdots \wedge e_n) + c\det(T_2)(e_1 \wedge \cdots \wedge e_n) \\
= (\det(T_1) + c\det(T_2))(e_1 \wedge \cdots \wedge e_n).
\]
This gives \( \det \in \text{Hom}_F(F^n,\ldots,F^n;F) \). Next we show that \( \det \) is alternating. Suppose that \( w_i = w_j \) for some \( i \neq j \). Then we have
\[
\det(w_1,\ldots,w_n) = \Lambda^n(T_1)(e_1 \wedge \cdots \wedge e_n)
= T_1(e_1) \wedge \cdots \wedge T_1(e_n)
= w_1 \wedge \cdots \wedge w_n
= 0.
\]
Thus,
\[
\det(w_1,\ldots,w_n) = \det(T_1) = 0.
\]
This gives \( \det \in \text{Alt}^n(F^n;F) \). It is easy to see that \( \det 1^n = 1 \).

Not only is \( \det \) in \( \text{Alt}^n(F^n;F) \) satisfying \( \det(1^n) = 1 \), it is the only such map. This takes a little work to prove, but it is the key to seeing this is actually the same map as the undergraduate version. This will follow immediately from observing the undergraduate definition certainly takes \( 1^n \) to 1 and is in \( \text{Alt}^n(F^n;F) \). To see the undergraduate definition is in \( \text{Alt}^n(F^n;F) \), one only needs to recall that the determinant is preserved under elementary column operations. We need a few result before proving the uniqueness.

Let \( n \) be a positive integer and let \( S_n \) denote the collection of bijections from \( \{1,\ldots,n\} \) to itself. It is easy to see this is a group under composition of functions. Given a tuple \( (x_1,\ldots,x_n) \), we define
\[
\sigma(x_1,\ldots,x_n) = (x_{\sigma(1)},\ldots,x_{\sigma(n)}).
\]
Note this tuple could be a tuple of anything: variables, vectors, numbers, etc. Define
\[
\Delta = \prod_{1 \leq i < j \leq n} (x_i - x_j).
\]
For example, if \( n = 3 \) we have
\[
\Delta = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3).
\]
We have
\[
\sigma\Delta = \prod_{1 \leq i < j \leq n} (x_{\sigma(i)} - x_{\sigma(j)}).
\]
Since \( \sigma \) just permutes the values of \( \{1,\ldots,n\} \), we have \( \sigma\Delta = \pm\Delta \) where the \( \pm \) depends on \( \sigma \). We define \( \text{sign}(\sigma) \in \{\pm 1\} \) by
\[
\text{sign}(\sigma)\Delta = \sigma\Delta.
\]

Lemma 7.3.14. Let \( t \in \text{Alt}^k(V,F) \). Then
\[
t(v_1,\ldots,v_i,v_{i+1},\ldots,v_k) = -t(v_1,\ldots,v_{i-1},v_{i+1},v_i,v_{i+2},\ldots,v_k).
\]
7.3. ALTERNATING FORMS, EXTERIOR POWERS, AND THE DETERMINANT

Proof. Define

$$\psi(x, y) = t(v_1, \ldots, v_{i-1}, x, y, v_{i+1}, \ldots, v_n)$$

for fixed $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$. It is enough to show that $\psi(x, y) = -\psi(y, x)$. Since $t \in \text{Alt}^k(V; F)$ we have

$$\psi(x + y, x + y) = 0.$$  

Expanding this and noting that $\psi(x, x) = \psi(y, y) = 0$ we obtain

$$\psi(x, y) = -\psi(y, x).$$

We leave the proof of the following lemma to the homework problems.

Lemma 7.3.15. Let $t \in \text{Alt}^k(V; F)$.

1. For each $\sigma \in S_k$,

$$t(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \text{sign}(\sigma) t(v_1, \ldots, v_k).$$

2. If $v_i = v_j$ for any $i \neq j$, then $t(v_1, \ldots, v_k) = 0$.

3. If $v_i$ is replaced by $v_i + cv_j$ for any $i \neq j$, $c \in F$, then the value of $t$ is unchanged.

The following proposition is the key step needed to show that det is unique.

Proposition 7.3.16. Let $t \in \text{Alt}^k(V; F)$. Assume for some $v_1, \ldots, v_n \in V$, $w_1, \ldots, w_n \in V$, and $a_{ij} \in F$ we have

$$w_1 = a_{11}v_1 + \cdots + a_{1n}v_n$$

$$\vdots$$

$$w_n = a_{n1}v_1 + \cdots + a_{nn}v_n.$$  

Then

$$t(w_1, \ldots, w_n) = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} t(v_1, \ldots, v_n).$$

Proof. We expand $t(w_1, \ldots, w_n)$ using multilinearity. Note that any of the terms with equal entries vanish since $t$ is alternating so we are left with terms of the form

$$a_{i_1} \cdots a_{i_n} t(v_{i_1}, \ldots, v_{i_n})$$

where the $i_1, \ldots, i_n$ run over $1, \ldots, n$ and are all distinct. Such tuples are in bijective correspondence with $S_n$, so we can write each such term as

$$a_{\sigma(1)} \cdots a_{\sigma(n)} t(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$$

for a unique $\sigma \in S_n$. This gives the result.  

\hfill \Box
Corollary 7.3.17. The determinant function defined above is the unique map in $\text{Alt}^n(F^n; F)$ that satisfies $\det(1_n) = 1$.

Proof. We have already seen that $\det$ satisfies the conditions; it only remains to show it is the unique function that does so. Let $E_n = \{e_1, \ldots, e_n\}$ be the standard basis of $F^n$. As above, we can identify elements of $\text{Mat}_n(F)$ with $F^n \times \cdots \times F^n$ via column vectors. Thus, the identity matrix is identified with $(e_1, \ldots, e_n)$. Let $A \in \text{Mat}_n(F)$ with column vectors $v_1, \ldots, v_n$, i.e.,

$$v_1 = a_{11}e_1 + \cdots + a_{1n}e_n$$
$$\vdots$$
$$v_n = a_{n1}e_1 + \cdots + a_{nn}e_n.$$

We now apply the previous proposition to see

$$\det(A) = \det(v_1, \ldots, v_n)$$
$$= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} \det(e_1, \ldots, e_n)$$
$$= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n}.$$

Moreover, given any $t \in \text{Alt}^n(F^n; F)$ with $t(1_n) = 1$ we see

$$t(v_1, \ldots, v_n) = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n}.$$

Thus, we have the result. \qed
7.4 Problems

For these problems $V$ and $W$ are finite dimensional $F$-vector spaces.

1. Let $V$ and $W$ be $F$-vector spaces. Prove using the universal property that $V \otimes_F W \cong W \otimes_F V$.

2. Let $V, W,$ and $U$ be $F$-vector spaces. Prove that the space of bilinear forms $\text{Hom}_F(V, W; U)$ is an $F$-vector space. Moreover, prove that $\text{Hom}_F(V, W; U) \cong \text{Hom}_F(V \otimes_F W, U)$.

3. Let $V_1, V_2, W_1,$ and $W_2$ be $F$-vector spaces. Let $T_1 \in \text{Hom}_F(V_1, W_1)$ and $T_2 \in \text{Hom}_F(V_2, W_2)$. Prove there is a unique $F$-linear map $T_1 \otimes T_2$ from $V_1 \otimes_F V_2$ to $W_1 \otimes_F W_2$ satisfying $(T_1 \otimes T_2)(v_1 \otimes v_2) = T_1(v_1) \otimes T_2(v_2)$.

4. Let $V_1 = W_1 = \mathbb{R}^3$ both with basis $A = \{x_1, x_2, x_3\}$ and $V_2 = W_2 = \mathbb{R}^2$ with basis $B = \{y_1, y_2\}$. Let $T_1(ax_1 + bx_2 + cx_3) = cx_1 + 2ax_2 - 3bx_3$ and $T_2(ay_1 + by_2) = (a + 3b)y_1 + (4b - 2a)y_2$. Let $\mathcal{C}$ be the basis for $V_1 \otimes_{\mathbb{R}} V_2$ and $W_1 \otimes_{\mathbb{R}} W_2$ given by

$$\mathcal{C} = \{x_1 \otimes y_1, x_1 \otimes y_2, x_2 \otimes y_1, x_2 \otimes y_2, x_3 \otimes y_1, x_3 \otimes y_2\}.$$

Compute the matrix $[T_1 \otimes T_2]_{\mathcal{C}}^\mathcal{C}$.

5. Let $V = P_5(\mathbb{Q})$ and $W = \mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Give a basis for $V \otimes_{\mathbb{Q}} W$.

6. Show that the cross product, defined in multivariable calculus, is a bilinear form from $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$.

7. (a) Let $\varphi \in V^\vee$ and $\psi \in W^\vee$. Define a map

$$B_{\varphi, \psi} : V \times W \to F$$

$$(v, w) \mapsto \varphi(v)\psi(w).$$

Show that $B_{\varphi, \psi}$ is a bilinear form.

(b) Prove that there is a natural isomorphism between $(V \otimes_F W)^\vee$ and $V^\vee \otimes_F W^\vee$. (Note a natural isomorphism means it does not depend on a choice of basis.)

8. Let $V$ and $W$ be $F$-vector spaces. Prove that $\text{Alt}^k(V, W)$ is an $F$-subspace of $\text{Hom}_F(V, \ldots, V; W)$.
9. Use the definition to compute the determinant of a 3 by 3 matrix over a field $F$. Check your result agrees with the definition that was given in undergraduate linear algebra.

10. Let $V$ be an $F$-vector space. Show that $\Lambda^k(V)^\vee \cong \text{Alt}^k(V)$ as $F$-vector spaces.

11. Let $T \in \text{Hom}_F(V,W)$. Give a detailed proof that $T$ induces a well-defined linear map $\Lambda^k(T) : \Lambda^k(V) \to \Lambda^k(W)$ given by

$$\Lambda^k(T)(v_1 \wedge \cdots \wedge v_k) = T(v_1) \wedge \cdots \wedge T(v_k)$$

for each $k \geq 0$.

12. Use the definition to prove that if $A \in \text{GL}_n(F)$, then $\det(A^{-1}) = \det(A)^{-1}$.

13. Prove that $v \wedge v_1 \wedge v_2 \wedge \cdots \wedge v_k = (-1)^k(v_1 \wedge v_2 \wedge \cdots \wedge v_k \wedge v)$.
Appendix A

Groups, rings, and fields: a quick recap

It is expected that students reading these notes have some experience with abstract algebra. For instance, an undergraduate course in abstract algebra is more than sufficient. A good source for those beginning in abstract algebra is [2] and a more advanced resource is [1]. We do not attempt to give a comprehensive treatment here; we present only the basic definitions and some examples to refresh the reader’s memory.

**Definition A.0.1.** Let $G$ be a nonempty set and $\star : G \times G \to G$ a binary operation. We call $(G,\star)$ a **group** if it satisfies the following properties:

- for all $g_1, g_2, \text{ and } g_3 \in G$ we have $(g_1 \star g_2) \star g_3 = g_1 \star (g_2 \star g_3)$;
- there exists an element $e_G \in G$, referred to as the identity element of $G$, so that $g \star e = g = e \star g$ for all $g \in G$;
- for each $g \in G$ there exists an element $g^{-1}$, referred to as the inverse of $g$, so that $g \star g^{-1} = e_G = g^{-1} \star g$.

We say a group is **abelian** if for all $g_1, g_2 \in G$ we have $g_1 \star g_2 = g_2 \star g_1$.

Often the operation $\star$ will be clear from context so we will write our groups as $G$ instead of $(G,\star)$.

**Exercise A.0.2.** Let $G$ be a group. Show that inverses are unique, i.e., if given $g \in G$ there exists $h \in G$ so that $g \star h = e_G = h \star g$, then $h = g^{-1}$.

**Definition A.0.3.** Let $(G,\star)$ be a group and $H \subseteq G$. We say $H$ is a **subgroup** of $(G,\star)$ if $(H,\star)$ is a group. One generally denotes $H$ is a subgroup of $G$ by $H \leq G$.

**Example A.0.4.** Let $G = \mathbb{C}$ and $\star = +$. It is easy to check that this gives an abelian group. In fact, it is easy to see that $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$. Note the operation is addition for all of these.
Example A.0.5. Let $G = \mathbb{Z}$ and let $\ast$ be multiplication. This does not give a group as $2 \in \mathbb{Z}$ but the multiplicative inverse of 2, namely, $1/2$ is not in $\mathbb{Z}$.

Example A.0.6. Let $G = \mathbb{R}^\times = \{x \in \mathbb{R} : x \neq 0\}$ with $\ast$ being multiplication. This gives an abelian group.

Example A.0.7. Let $X$ be a set of $n$ elements. Let $S_n$ denote the set of bijections from $X$ to $X$, i.e., the permutations of $X$. This forms a group under composition of functions.

Example A.0.8. Let $\text{Mat}_n(\mathbb{R})$ denote the $n$ by $n$ matrices with entries in $\mathbb{R}$. This forms a group under matrix addition but not under matrix multiplication as the 0 matrix, the matrix with all 0’s as entries, does not have a multiplicative inverse.

Example A.0.9. Let $\text{GL}_n(\mathbb{R})$ be the subset of $\text{Mat}_n(\mathbb{R})$ consisting of invertible matrices. This gives a group under matrix multiplication. Let $\text{SL}_n(\mathbb{R})$ be the subset of $\text{GL}_n(\mathbb{R})$ consisting of matrices with determinant 1. This is a subgroup of $\text{GL}_n(\mathbb{R})$.

Example A.0.10. Let $n$ be a positive integer and consider the set $H = n\mathbb{Z} = \{nm : m \in \mathbb{Z}\}$. Note that $H \leq (\mathbb{Z}, +)$. We can form a quotient group in this case as follows. We define an equivalence relation on $\mathbb{Z}$ by setting $a \sim b$ if $n|(a - b)$. One should check this is in fact an equivalence relation. This partitions $\mathbb{Z}$ into equivalence classes. For example, given $a \in \mathbb{Z}$, the equivalence class containing $a$ is given by $[a]_n = \{b \in \mathbb{Z} : n|(a - b)\}$. We write $a \equiv b \pmod{n}$ if $a$ and $b$ lie in the same equivalence class, i.e., if $[a]_n = [b]_n$. Note that for each $a \in \mathbb{Z}$, there is an element $r \in \{0, 1, \ldots, n - 1\}$ so that $[a]_n = [r]_n$. This follows immediately from the division algorithm: write $a = nq + r$ with $0 \leq r \leq n - 1$. Write $\mathbb{Z}/n\mathbb{Z} = \{[0]_n, [1]_n, \ldots, [n - 1]_n\}$.

We define addition on this set by setting $[a]_n + [b]_n = [a + b]_n$. One can easily check this is well-defined and makes $\mathbb{Z}/n\mathbb{Z}$ into an abelian group with $n$ elements. We drop the subscript $n$ on the equivalence classes when it is clear from context.

Consider the case that $n = 4$. We have $\mathbb{Z}/4\mathbb{Z} = \{[0], [1], [2], [3]\}$. An addition table is given here:

<table>
<thead>
<tr>
<th></th>
<th>[0]</th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0]</td>
<td>[0]</td>
<td>[1]</td>
<td>[2]</td>
<td>[3]</td>
</tr>
<tr>
<td>[1]</td>
<td>[1]</td>
<td>[2]</td>
<td>[3]</td>
<td>[0]</td>
</tr>
<tr>
<td>[2]</td>
<td>[2]</td>
<td>[3]</td>
<td>[0]</td>
<td>[1]</td>
</tr>
<tr>
<td>[3]</td>
<td>[3]</td>
<td>[0]</td>
<td>[1]</td>
<td>[2]</td>
</tr>
</tbody>
</table>

Example A.0.11. Let $n \in \mathbb{Z}$ and consider the subset of $\mathbb{Z}/n\mathbb{Z}$ consisting of those $[a]$ so that $\gcd(a, n) = 1$. We denote this set by $(\mathbb{Z}/n\mathbb{Z})^\times$. Recall that if $\gcd(a, n) = 1$ then there exists $x, y \in \mathbb{Z}$ so that $ax + ny = 1$. This means that $n \mid (ax - 1)$, i.e., $ax \equiv 1 \pmod{n}$. If $[a], [b] \in (\mathbb{Z}/n\mathbb{Z})^\times$, then $[ab] \in (\mathbb{Z}/n\mathbb{Z})^\times$ as well. One can now check that if we set $[a] \ast [b] = [ab]$, then this makes $(\mathbb{Z}/n\mathbb{Z})^\times$ into an abelian group. Note that $(\mathbb{Z}/n\mathbb{Z})^\times$ is not a subgroup of $\mathbb{Z}/n\mathbb{Z}$ even though it is a subset; they do not have the same operations.
Exercise A.0.12. (a) How many elements are in \((\mathbb{Z}/n\mathbb{Z})^\times\)?
(b) Give multiplication tables for \((\mathbb{Z}/4\mathbb{Z})^\times\) and \((\mathbb{Z}/5\mathbb{Z})^\times\). Do you notice anything interesting about these tables?

As is always the case, the relevant maps to study are the ones that preserve the structure being studied. For groups these are group homomorphisms.

Definition A.0.13. Let \((G, \star_G)\) and \((H, \star_H)\) be groups. A map \(\phi : G \rightarrow H\) is a group homomorphism if it satisfies \(\phi(g_1 \star_G g_2) = \phi(g_1) \cdot \phi(g_2)\) for all \(g_1, g_2 \in G\).

We define the kernel of the map \(\phi\) by
\[
\ker \phi = \{ g \in G : \phi(g) = e_H \}.
\]

Exercise A.0.14. (a) Show that \(\ker \phi \leq G\).
(b) Show that if \(\phi : G \rightarrow H\) is a group homomorphism, then \(\phi\) is injective if and only if \(\ker \phi = \{e_G\}\).

Example A.0.15. Let \(n \in \mathbb{Z}\) be a positive integer and define \(\phi : \mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}\) by \(\phi(a) = [a]_n\). This is easily seen to be a surjective group homomorphism. Moreover, one has that \(\ker \phi = n\mathbb{Z}\).

Example A.0.16. Let \(\phi : \mathbb{R} \rightarrow \mathbb{C}^\times\) be defined by \(\phi(x) = e^{2\pi i x}\). Given \(x, y \in \mathbb{R}\) we have \(\phi(x + y) = e^{2\pi i (x+y)} = e^{2\pi i x}e^{2\pi iy} = \phi(x)\phi(y)\). Thus, \(\phi\) is a group homomorphism. The image of this group homomorphism is \(S^1 = \{z \in \mathbb{C} : |z| = 1\}\), i.e., the unit circle. The kernel is given by \(\mathbb{Z}\). For those that remember the first isomorphism theorem, this gives \(\mathbb{R}/\mathbb{Z} \cong S^1\).

Example A.0.17. The determinant map gives a group homomorphism from \(\text{GL}_n(\mathbb{R})\) to \(\mathbb{R}^\times\) with kernel \(\text{SL}_n(\mathbb{R})\).

As one may have observed, the first examples of groups given were familiar objects such as \(\mathbb{Z}\) and \(\mathbb{R}\) with only one operation considered. It is natural to consider both operations for such objects, which brings us to the definition of a ring.

Definition A.0.18. A ring is a nonempty set \(R\) with two binary operations, \(+\) and \(\cdot\), satisfying the following properties

- \(R\) is an abelian group under the operation \(+\) (the additive identity is denoted \(0_R\));
- there is an element \(1_R\) that satisfies \(r \cdot 1_R = r = 1_R \cdot r\) for all \(r \in R\);
- for all \(r_1, r_2, r_3 \in R\) one has \((r_1 \cdot r_2) \cdot r_3 = r_1 \cdot (r_2 \cdot r_3)\);
- for all \(r_1, r_2, r_3 \in R\) one has \((r_1 + r_2) \cdot r_3 = r_1 \cdot r_3 + r_2 \cdot r_3\);
- for all \(r_1, r_2, r_3 \in R\) one has \(r_1 \cdot (r_2 + r_3) = r_1 \cdot r_2 + r_1 \cdot r_3\).

We say \(R\) is commutative if \(r_1 \cdot r_2 = r_2 \cdot r_1\) for all \(r_1, r_2 \in R\).
It is often the case that one writes \( r_1 \cdot r_2 \) as \( r_1 r_2 \) to ease notation. We will follow that convention here when there is no fear of confusion.

Note that what we have defined as a ring is often referred to as a \textit{ring with identity}. This is because in some cases one would like to consider rings without requiring that there be an element \( 1_R \). We will not be interested in such rings so in these notes “ring” always means “ring with identity”.

\textbf{Example A.0.19.} Many of the examples given above are also rings. The groups \( \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \) and \( \mathbb{C} \) are all commutative rings when considered with the operations \( + \) and \( \cdot \).

\textbf{Example A.0.20.} Let \( R \) be a ring. The group \( \mathrm{Mat}_n(R) \) is a ring under the operations of matrix addition and multiplication. However, if \( n \geq 2 \) this is not a commutative ring.

\textbf{Example A.0.21.} Let \( R \) be a ring and let \( R[x] \) denote the set of polynomials with coefficients in \( R \). The set \( R[x] \) is a ring under polynomial addition and multiplication.

Given a ring \( R \), a \textit{subring} of \( R \) is a nonempty set \( S \) that is a ring under the same operations. However, it turns out that the concept of ideals is more important than the notion of subrings.

\textbf{Definition A.0.22.} Let \( R \) be a ring and \( I \subset R \) be a subring. We say \( I \) is an \textit{ideal} if \( ra \in I \) and \( ar \in I \) for all \( r \in R \) and \( a \in I \).

Note this is a stronger condition than being a subring. For example, \( \mathbb{Z} \) is a subring of \( \mathbb{R} \) but it is not an ideal.

\textbf{Example A.0.23.} Consider the ring \( \mathbb{Z} \). For any integer \( n \) we have \( n\mathbb{Z} \) is an ideal in \( \mathbb{Z} \). In fact, every ideal can be written in this form. Let \( I \) be an ideal in \( \mathbb{Z} \). If \( I = \mathbb{Z} \) then clearly we have \( I = 1\mathbb{Z} \). Suppose \( I \) is a proper subset of \( \mathbb{Z} \) and \( I \neq 0 \). Let \( I^+ \) be the subset of \( I \) consisting of positive elements. Since \( I \neq 0 \) this is a nonempty set because if \( m \in I \), either \( m \in I^+ \) or \( -m \in I^+ \). Since \( I^+ \) is a set of positive integers, it has a minimal element; call this minimal element \( n \). We claim \( I = n\mathbb{Z} \). Let \( m \in I \). Then we can write \( m = nq + r \) for some \( 0 \leq r \leq n - 1 \). Since \( m \in I \) and \( n \in I \), we have \( r = m - nq \in I \). However, this contradicts \( n \) being the least positive integer in \( I \) unless \( r = 0 \). Thus, \( m \in n\mathbb{Z} \) and so \( I = n\mathbb{Z} \). We call an ideal generated by a single element a \textit{principal ideal}. A ring where all the ideals are principal is called a \textit{principal ideal domain}. Thus, \( \mathbb{Z} \) is a principal ideal domain.

\textbf{Exercise A.0.24.} Let \( F \) be a field. Mimic the proof that \( \mathbb{Z} \) is a principal ideal domain to show \( F[x] \) is a principal ideal domain.

\textbf{Example A.0.25.} Consider the polynomial ring \( R[x] \) with \( R \) a ring. Let \( f(x) \in R[x] \) be any nonzero polynomial. We define an equivalence relation on \( R[x] \) much as we did on \( \mathbb{Z} \) when defining \( \mathbb{Z}/n\mathbb{Z} \), namely, we say \( g(x) \sim h(x) \) if \( f(x) \mid (g(x) - h(x)) \). We write \( g(x) \equiv h(x) \) (mod \( f(x) \)) if \( g(x) \equiv h(x) \). This
equivalence relation partitions \( R[x] \) into equivalence classes. Given \( g(x) \in R[x] \), we let \([g(x)]_{f(x)}\) denote the equivalence class containing \( g(x) \), i.e., \([g(x)]_{f(x)} = \{h(x) \in R[x] : h(x) \equiv g(x) \pmod{f(x)}\}\). We drop the subscript \( f(x) \) when it is clear from context. We denote the set of equivalence classes by \( R[x]/(f(x)) = \{[g(x)] : g(x) \in R[x]\} \). Note that \( R[x]/(f(x)) \) need not be a finite set. In this case one can use the division algorithm to see that one can always choose a representative \( r(x) \) for the equivalence class with \( \deg r(x) < \deg f(x) \). We define addition and multiplication on \( R[x]/(f(x)) \) by \([g(x)] + [h(x)] = [g(x) + h(x)]\) and \([g(x)] \cdot [h(x)] = [g(x)h(x)]\). This makes \( R[x]/(f(x)) \) into a ring. This ring is commutative if \( R \) is commutative.

**Exercise A.0.26.** Consider the ring \( \mathbb{R}[x]/(x^2 + 1) \). Show that one has
\[
\mathbb{R}[x]/(x^2 + 1) = \{ax + b : a, b \in \mathbb{R}\}.
\]
What is \([x]^2\) equal to in this ring? What familiar ring does this look like?

**Exercise A.0.27.** Let \( R \) be a ring and \( I \subset R \) an ideal. Define a relation \( \sim \) on \( R \) by setting \( a \sim b \) if \( a - b \in I \). Show this is an equivalence relation. For \( a \in R \), let \( a + I \) denote the equivalence class containing \( a \), i.e., \( a + I = \{a + i : i \in I\} \). (Note this is also sometimes denoted \([a]\).) Let \( R/I \) denote the set of these equivalence classes. Define \((a + I) + (b + I) = (a + b) + I\) and \((a + I)(b + I) = ab + I\). Show this makes \( R/I \) into a ring. Rectify this with the notation \( \mathbb{Z}/n\mathbb{Z} \) and \( R[x]/(f(x)) \) used above.

While rings are extremely important if one wants to consider the more general results of linear algebra that are phrased in terms of modules, for the main results in these notes we stick with vector spaces. Thus, we are mainly interested in fields.

**Definition A.0.28.** A commutative ring \( R \) is said to be a field if every nonzero element in \( R \) has a multiplicative inverse.

**Exercise A.0.29.** Show the only ideals in a field \( F \) are 0 and \( F \).

**Example A.0.30.** The rings \( \mathbb{Q}, \mathbb{R} \) and \( \mathbb{C} \) are all fields, but \( \mathbb{Z} \) is not a field.

**Example A.0.31.** Let \( n \) be a positive integer and recall the group \( \mathbb{Z}/n\mathbb{Z} \) with the operation given by \(+\). Given \([a],[b] \in \mathbb{Z}/n\mathbb{Z}\), we define \([a] \cdot [b] = [ab]\). One can easily check this makes \( \mathbb{Z}/n\mathbb{Z} \) into a commutative ring. Now consider the case when \( n = p \) is a prime. We still clearly have \( \mathbb{Z}/p\mathbb{Z} \) is a commutative ring, but in this case we have more. Let \([a] \in \mathbb{Z}/p\mathbb{Z} \) with \([a] \neq [0] \), i.e., \( p \nmid a \). Since \( p \) is a prime this means that \( \gcd(a,p) = 1 \), so there exists \( x, y \in \mathbb{Z} \) so that \( ax + by = 1 \), i.e., \( ax \equiv 1 \pmod{p} \). This means that \([a][x] = [1] \), i.e., \([a] \) has a multiplicative inverse. Thus, \( \mathbb{Z}/p\mathbb{Z} \) is a field. (Note this does not work for general \( n \); look back at your multiplication table for \( \mathbb{Z}/4\mathbb{Z} \).) When we wish to think of \( \mathbb{Z}/p\mathbb{Z} \) as a field we denote it by \( \mathbb{F}_p \).

**Exercise A.0.32.** Let \( F \) be a field and \( f(x) \in F[x] \) an irreducible polynomial. Show that \( F[x]/(f(x)) \) is a field. (Hint: mimic the proof that \( \mathbb{Z}/p\mathbb{Z} \) is a field.)
Appendix A. Groups, Rings, and Fields: A Quick Recap

Exercise A.0.33. Consider the ring $\mathbb{C}[x]/(x^2 + 1)$. Is this a field? Prove your answer.

Definition A.0.34. Let $R$ and $S$ be rings. A map $\phi : R \to S$ is a ring homomorphism if it satisfies

- $\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2)$ for all $r_1, r_2 \in R$;
- $\phi(r_1 \cdot r_2) = \phi(r_1) \cdot \phi(r_2)$ for all $r_1, r_2 \in R$.

We define the kernel of $\phi$ by

$$\ker \phi = \{ r \in R : \phi(r) = 0_S \}.$$

Exercise A.0.35. Let $\phi : R \to S$ be a ring homomorphism. Show $\ker \phi$ is an ideal of $R$.

Exercise A.0.36. Let $F$ be a field and $R$ a ring. Let $\phi : F \to R$ be a ring homomorphism and assume $\phi$ is not the zero map, i.e., there exists $x \in F$ so that $\phi(x) \neq 0_R$. Prove that $\phi$ is an injective map.

Example A.0.37. Let $S \subset R$ be a subring. Then inclusion map $S \hookrightarrow R$ is a ring homomorphism with a trivial kernel.

Example A.0.38. Let $R$ be a ring and $I \subset R$ an ideal. The natural projection map

$$R \to R/I$$

$$a \mapsto a + I$$

is a surjective ring homomorphism with kernel equal to $I$.

Example A.0.39. Define $\phi : \mathbb{R}[x] \to \mathbb{C}$ by $f(x) \mapsto f(i)$. This is a ring homomorphism with kernel $(x^2 + 1)$.

Exercise A.0.40. Define a map $\phi : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$ by $\phi([a]) = [4a]$. What is the kernel of this map? What is the image of this map?
Bibliography


