Linear Algebra: MAT 217
Lecture notes, Spring 2012

Michael Damron

compiled from lectures and exercises designed with Tasho Kaletha

Princeton University

Contents

1 Vector spaces **4**
- 1.1 Definitions . . . 4
- 1.2 Subspaces . . . 6
- 1.3 Linear independence and bases . . . 9
- 1.4 Exercises . . . 14

2 Linear transformations **16**
- 2.1 Definitions and basic properties . . . 16
- 2.2 Range and nullspace, one-to-one, onto . . . 17
- 2.3 Isomorphisms and $\mathcal{L}(V, W)$ . . . 20
- 2.4 Matrices and coordinates . . . 22
- 2.5 Exercises . . . 26

3 Dual spaces **32**
- 3.1 Definitions . . . 32
- 3.2 Annihilators . . . 33
- 3.3 Double dual . . . 34
- 3.4 Dual maps . . . 35
- 3.5 Exercises . . . 37

4 Determinants **39**
- 4.1 Permutations . . . 39
- 4.2 Determinants: existence and uniqueness . . . 41
- 4.3 Properties of determinants . . . 44
- 4.4 Exercises . . . 47

5 Eigenvalues **54**
- 5.1 Definitions and the characteristic polynomial . . . 54
- 5.2 Eigenspaces and the main diagonalizability theorem . . . 56
- 5.3 Exercises . . . 58

6 Jordan form **62**
- 6.1 Generalized eigenspaces . . . 62
- 6.2 Primary decomposition theorem . . . 63
- 6.3 Nilpotent operators . . . 66
- 6.4 Existence and uniqueness of Jordan form, Cayley-Hamilton . . . 70
- 6.5 Exercises . . . 72

7 Bilinear forms **79**
- 7.1 Definitions . . . 79
- 7.2 Symmetric bilinear forms . . . 81
- 7.3 Sesquilinear and Hermitian forms . . . 84

7.4 Exercises 85

8 Inner product spaces **88**
8.1 Definitions 88
8.2 Orthogonality 89
8.3 Adjoints 93
8.4 Spectral theory of self-adjoint operators 95
8.5 Normal and commuting operators 98
8.6 Exercises 99

1 Vector spaces

1.1 Definitions

We begin with the definition of a vector space. (Keep in mind vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$.)

Definition 1.1.1. *A vector space is a collection of two sets, V and F. The elements of F (usually we take $\mathbb{R}$ or $\mathbb{C}$) are called scalars and the elements of V are called vectors. For each $v, w \in V$, there is a vector sum, $v + w \in V$, with the following properties.*

0. *There is one (and only one) vector called $\vec{0}$ with the property*
$$v + \vec{0} = v \text{ for all } v \in V \ ;$$
1. *for each $v \in V$ there is one (and only one) vector called $-v$ with the property*
$$v + (-v) = \vec{0} \text{ for all } v \in V \ ;$$
2. *commutativity of vector sum:*
$$v + w = w + v \text{ for all } v, w \in V \ ;$$
3. *associativity of vector sum:*
$$(v + w) + z = v + (w + z) \text{ for all } v, w, z \in V \ .$$

Furthermore, for each $v \in V$ and $c \in F$ there is a scalar product $cv \in V$ with the following properties.

1. *For all $v \in V$,*
$$1v = v \ .$$
2. *For all $v \in V$ and $c, d \in F$,*
$$(cd)v = c(dv) \ .$$
3. *For all $c \in F, v, w \in V$,*
$$c(v + w) = cv + cw \ .$$
4. *For all $c, d \in F, v \in V$,*
$$(c + d)v = cv + dv \ .$$

Here are some examples.

1. $V = \mathbb{R}^n$, $F = \mathbb{R}$. Addition is given by
$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) = (v_1 + w_1, \ldots, v_n + w_n)$$
and scalar multiplication is given by
$$c(v_1, \ldots, v_n) = (cv_1, \ldots, cv_n) \ .$$

2. Polynomials: take V to be all polynomials of degree up to n with real coefficients

$$V = \{a_n x^n + \cdots + a_1 x + a_0 : a_i \in \mathbb{R} \text{ for all } i\}$$

and take $F = \mathbb{R}$. Note the similarity to $\mathbb{R}^n$.

3. Let S be any nonempty set and let V be the set of functions from S to $\mathbb{C}$. Set $F = \mathbb{C}$. If $f_1, f_2 \in V$ set $f_1 + f_2$ to be the function given by

$$(f_1 + f_2)(s) = f_1(s) + f_2(s) \text{ for all } s \in S$$

and if $c \in \mathbb{C}$ set cf_1 to be the function given by

$$(cf_1)(s) = c(f_1(s)) \ .$$

4. Let

$$V = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right\}$$

(or any other fixed object) with $F = \mathbb{C}$. Define

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

and

$$c \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \ .$$

In general F is allowed to be a *field.*

Definition 1.1.2. *A set F is called a field if for each $a, b \in F$ there is an element $ab \in F$ and one $a + b \in F$ such that the following hold.*

1. *For all $a, b, c \in F$ we have $(ab)c = a(bc)$ and $(a + b) + c = a + (b + c)$.*
2. *For all $a, b \in F$ we have $ab = ba$ and $a + b = b + a$.*
3. *There exists an element $0 \in F$ such that for all $a \in F$, we have $a + 0 = a$; furthermore there is a non-zero element $1 \in F$ such that for all $a \in F$, we have $1a = a$.*
4. *For each $a \in F$ there is an element $-a \in F$ such that $a + (-a) = 0$. If $a \neq 0$ there exists an element a^{-1} such that $aa^{-1} = 1$.*
5. *For all $a, b, c \in F$,*

$$a(b + c) = ab + ac \ .$$

Here are some general facts.

1. For all $c \in F$, $c\vec{0} = \vec{0}$.

Proof.

$$c\vec{0} = c(\vec{0} + \vec{0}) = c\vec{0} + c\vec{0}$$

$$\begin{aligned} c\vec{0} + (-(c\vec{0})) &= (c\vec{0} + c\vec{0}) + (-c(\vec{0})) \\ \vec{0} &= c\vec{0} + (c\vec{0} + (-(c\vec{0}))) \\ \vec{0} &= c\vec{0} + \vec{0} \\ \vec{0} &= c\vec{0}\,. \end{aligned}$$

□

Similarly one may prove that for all $v \in V$, $0v = \vec{0}$.

2. For all $v \in V$, $(-1)v = -v$.

Proof.

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

However $-v$ is the unique vector such that $v + (-v) = \vec{0}$. Therefore $(-1)v = -v$. □

1.2 Subspaces

Definition 1.2.1. *A subset $W \subseteq V$ of a vector space is called a subspace if (W, F) with the same operations is also a vector space.*

Many of the rules for vector spaces follow directly by "inheritance." For example, if $W \subseteq V$ then for all $v, w \in W$ we have $v + w = w + v$. We actually only need to check a few:

A. $\vec{0} \in W$.

B. For all $w \in W$ the vector $-w$ is also in W.

C. For all $w \in W$ and $c \in F$, $cw \in W$.

D. For all $v, w \in W$, $v + w \in W$.

Theorem 1.2.2. *$W \subseteq V$ is a subspace if and only if it is nonempty and for all $v, w \in W$ and $c \in F$ we have $cv + w \in W$.*

Proof. Suppose that W is a subspace and let $v, w \in W$, $c \in F$. By (C) we have $cv \in W$. By (D) we have $cv + w \in W$.

Conversely suppose that for all $v, w \in W$ and $c \in F$ we have $cv + w \in W$. Then we need to show A-D.

A. Since W is nonempty choose $w \in W$. Let $v = w$ and $c = -1$. This gives $\vec{0} = w - w \in W$.

B. Set $v = w$, $w = 0$ and $c = -1$.

C. Set $v = w$, $w = 0$ and $c \in F$.

D. Set $c = 1$.

□

Examples:

1. If V is a vector space then $\{0\}$ is a subspace.
2. Take $V = \mathbb{C}^n$. Let
$$W = \{(z_1, \ldots, z_n) : \sum_{i=1}^{n} z_i = 0\} \ .$$
Then W is a subspace. (Exercise.)
3. Let V be the set of 2×2 matrices with real entries.
$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} + \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{pmatrix}$$
$$c \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} = \begin{pmatrix} ca_1 & cb_1 \\ cc_1 & cd_1 \end{pmatrix} \ .$$
Then is W is a subspace, where
$$W = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a + d = 1 \right\} \ ?$$
4. In $\mathbb{R}^n$ all subspaces are hyperplanes through the origin.

Theorem 1.2.3. *Suppose that $\mathcal{C}$ is a non-empty collection of subspaces of V. Then the intersection*
$$\widetilde{W} = \cap_{W \in \mathcal{C}} W$$
is a subspace.

Proof. Let $c \in F$ and $v, w \in \widetilde{W}$. We need to show that (a) $\widetilde{W} \neq \varnothing$ and (b) $cv + w \in \widetilde{W}$. The first holds because each W is a subspace, so $0 \in W$. Next for all $W \in \mathcal{C}$ we have $v, w \in W$. Then $cv + w \in W$. Therefore $cv + w \in \widetilde{W}$. □

Definition 1.2.4. *If $S \subseteq V$ is a subset of vectors let $\mathcal{C}_S$ be the collection of all subspaces containing S. The span of S is defined as*
$$span(S) = \cap_{W \in \mathcal{C}_S} W \ .$$
Since $\mathcal{C}_S$ is non-empty, $span(S)$ is a subspace.

Question: What is $span(\{\})$?

Definition 1.2.5. *If*

$$v = a_1 w_1 + \cdots + a_k w_k$$

for scalars $a_i \in F$ and vectors $w_i \in V$ then we say that v is a linear combination of $\{w_1, \ldots, w_k\}$.

Theorem 1.2.6. *If $S \neq \varnothing$ then $span(S)$ is equal to the set of all finite linear combinations of elements of S.*

Proof. Set

$$\widetilde{S} = \text{ all finite l.c.'s of elements of } S \ .$$

We want to show that $\widetilde{S} = span(S)$. First we show that $\widetilde{S} \subseteq span(S)$. Let

$$a_1 s_1 + \cdots + a_k s_k \in \widetilde{S}$$

and let W be a subspace in $\mathcal{C}_S$. Since $s_i \in S$ for all i we have $s_i \in W$. By virtue of W being a subspace, $a_1 s_1 + \cdots + a_k s_k \in W$. Since this is true for all $W \in \mathcal{C}_S$ then $a_1 s_1 + \cdots + a_k s_k \in span(S)$. Therefore $\widetilde{S} \subseteq span(S)$.

In the other direction, the set $\widetilde{S}$ is itself a subspace and it contains S (exercise). Thus $\widetilde{S} \in \mathcal{C}_S$ and so

$$span(S) = \cap_{W \in \mathcal{C}_S} W \subseteq \widetilde{S} \ .$$

□

Question: If W_1 and W_2 are subspaces, is $W_1 \cup W_2$ a subspace? If it were, it would be the smallest subspace containing $W_1 \cup W_2$ and thus would be equal to $span(W_1 \cup W_2)$. But it is not!

Proposition 1.2.7. *If W_1 and W_2 are subspaces then $span(W_1 \cup W_2) = W_1 + W_2$, where*

$$W_1 + W_2 = \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\} \ .$$

Furthermore for each $n \geq 1$,

$$span\left(\cup_{k=1}^{n} W_k\right) = W_1 + \cdots + W_n \ .$$

Proof. A general element $w_1 + w_2$ in $W_1 + W_2$ is a linear combination of elements in $W_1 \cup W_2$ so it is in the span of $W_1 \cup W_2$. On the other hand if $a_1 v_1 + \cdots + a_n v_n$ is an element of the span then we can split the v_i's into vectors from W_1 and those from W_2. For instance, $v_1, \ldots, v_k \in W_1$ and $v_{k+1}, \ldots, v_n \in W_2$. Now $a_1 v_1 + \cdots + a_k v_k \in W_1$ and $a_{k+1} v_{k+1} + \cdots + a_n v_n \in W_2$ by the fact that these are subspaces. Thus this linear combination is equal to $w_1 + w_2$ for

$$w_1 = a_1 v_1 + \cdots + a_k v_k \text{ and } w_2 = a_{k+1} v_{k+1} + \cdots + a_n v_n \ .$$

The general case is an exercise. □

Remark.

1. *$Span(S)$ is the "smallest" subspace containing S in the following sense: if W is any subspace containing S then $span(S) \subseteq W$.*
2. *If $S \subseteq T$ then $Span(S) \subseteq Span(T)$.*
3. *If W is a subspace then $Span(W) = W$. Therefore $Span(Span(S)) = Span(S)$.*

1.3 Linear independence and bases

Now we move on to linear independence.

Definition 1.3.1. *We say that vectors $v_1, \ldots, v_k \in V$ are linearly independent if whenever*

$$a_1 v_1 + \cdots + a_k v_k = \vec{0} \text{ for scalars } a_i \in F$$

then $a_i = 0$ for all i. Otherwise we say they are linearly dependent.

Lemma 1.3.2. *Let $S = \{v_1, \ldots, v_n\}$ for $n \geq 1$. Then S is linearly dependent if and only if there exists $v \in S$ such that $v \in Span(S \setminus \{v\})$.*

Proof. Suppose first that S is linearly dependent and that $n \geq 2$. Then there exist scalars $a_1, \ldots, a_n \in F$ which are not all zero such that $a_1 v_1 + \cdots + a_n v_n = \vec{0}$. By reordering we may assume that $a_1 \neq 0$. Now

$$v_1 = -\frac{a_2}{a_1} v_2 + \ldots + \left(-\frac{a_n}{a_1}\right) v_n \ .$$

So $v_1 \in Span(S \setminus \{v_1\})$.

If S is linearly dependent and $n = 1$ then there exists a nonzero a_1 such that $a_1 v_1 = \vec{0}$, so $v_1 = \vec{0}$. Now $v_1 \in Span(\{\}) = Span(S \setminus \{v_1\})$.

Conversely, suppose there exists $v \in S$ such that $v \in span(S \setminus \{v\})$ and $n \geq 2$. By reordering we may suppose that $v = v_1$. Then there exist scalars $a_2, \ldots, a_n$ such that

$$v_1 = a_2 v_2 + \cdots + a_n v_n \ .$$

But now we have

$$(-1)v_1 + a_2 v_2 + \cdots + a_n v_n = \vec{0} \ .$$

Since this is a nontrivial linear combination, S is linearly dependent.

If $n = 1$ and $v_1 \in Span(S \setminus \{v_1\})$ then $v_1 \in Span(\{\}) = \{\vec{0}\}$, so that $v_1 = \vec{0}$. Now it is easy to see that S is linearly dependent.

□

Examples:

1. For two vectors $v_1, v_2 \in V$, they are linearly dependent if and only if one is a scalar multiple of the other. By reordering, we may suppose $v_1 = a v_2$.

2. For three vectors this is not true anymore. The vectors $(1,1)$, $(1,0)$ and $(0,1)$ in $\mathbb{R}^2$ are linearly dependent since

   $$(1,1) + (-1)(1,0) + (-1)(0,1) = (0,0) \ .$$

   However none of these is a scalar multiple of another.

3. $\{\vec{0}\}$ is linearly dependent:
$$1 \cdot \vec{0} = \vec{0} \ .$$

Proposition 1.3.3. *If S is a linearly independent set and T is a nonempty subset then T is linearly independent.*

Proof. Suppose that
$$a_1 t_1 + \cdots + a_n t_n = \vec{0}$$
for vectors $t_i \in T$ and scalars $a_i \in F$. Since each $t_i \in S$ this is a linear combination of elements of S. Since S is linearly independent all the coefficients must be zero. Thus $a_i = 0$ for all i. This was an arbitrary linear combination of elements of T so T is linearly independent. □

Corollary 1.3.4. *If S is linearly dependent and R contains S then R is linearly dependent. In particular, any set containing $\vec{0}$ is linearly dependent.*

Proposition 1.3.5. *If S is linearly independent and $v \in Span(S)$ then there exist unique vectors $v_1, \ldots, v_n$ and scalars $a_1, \ldots, a_n$ such that*
$$v = a_1 v_1 + \cdots + a_n v_n \ .$$

Proof. Suppose that S is linearly independent and there are two representations
$$v = a_1 v_1 + \cdots + a_n v_n \text{ and } v = b_1 w_1 + \cdots + b_k w_k \ .$$
Then split the vectors in $\{v_1, \ldots, v_n\} \cup \{w_1, \ldots, w_k\}$ into three sets: S_1 are those in the first but not the second, S_2 are those in the second but not the first, and S_3 are those in both.
$$\vec{0} = v - w = \sum_{s_j \in S_1} a_j s_j + \sum_{s_j \in S_2} b_j s_j + \sum_{s_j \in S_3} (a_j - b_j) s_j \ .$$
This is a linear combination of elements from S and by linear independence all coefficients are zero. Thus both representations used the same vectors (in S_3) and with the same coefficients and are thus the same. □

Lemma 1.3.6 (Steinitz exchange)**.** *Let $L = \{v_1, \ldots, v_k\}$ be an linearly independent set in a vector space V and let $S = \{w_1, \ldots, w_m\}$ be a spanning set; that is, $Span(S) = V$. Then*

1. *$k \leq m$ and*
2. *there exist $m - k$ vectors $s_1, \ldots, s_{m-k} \in S$ such that*
$$Span(\{v_1, \ldots, v_k, s_1, \ldots, s_{m-k}\}) = V \ .$$

Proof. We will prove this by induction on k. For $k = 0$ it is obviously true (using the fact that $\{\}$ is linearly independent). Suppose it is true for k and we will prove it for $k+1$. In other words, let $\{v_1, \ldots, v_{k+1}\}$ be a linearly independent set. Then by last lecture, since $\{v_1, \ldots, v_k\}$ is linearly independent, we find $k \leq m$ and vectors $s_1, \ldots, s_{m-k} \in S$ with

$$Span(\{v_1, \ldots, v_k, s_1, \ldots, s_{m-k}\}) = V \ .$$

Now since $v_{k+1} \in V$ we can find scalars $a_1, \ldots, a_k$ and $b_1, \ldots, b_{m-k}$ in F such that

$$v_{k+1} = a_1 v_1 + \cdots + a_k v_k + b_1 s_1 + \cdots b_{m-k} s_{m-k} \ . \tag{1}$$

We claim that not all of the b_i's are zero. If this were the case then we would have

$$v_{k+1} = a_1 v_1 + \cdots + a_k v_k$$

$$a_1 v_1 + \cdots + a_k v_k + (-1) v_{k+1} = \vec{0} \ ,$$

a contradiction to linear independence. Also this implies that $k \neq m$, since otherwise the linear combination (1) would contain no b_i's. Thus

$$k \leq m + 1 \ .$$

Suppose for example that $b_1 \neq 0$. Then we could write

$$b_1 s_1 = -a_1 v_1 + \cdots + (-a_k) v_k + v_{k+1} + (-b_2) s_2 + \cdots + (-b_{m-k}) s_{m-k}$$

$$s_1 = \left(-\frac{a_1}{b_1}\right) v_1 + \cdots + \left(-\frac{a_k}{b_1}\right) v_k + \frac{1}{b_1} v_{k+1} + \left(-\frac{b_2}{b_1}\right) s_2 + \cdots + \left(-\frac{b_{m-k}}{b_1}\right) s_{m-k} \ .$$

In other words,

$$s_1 \in Span(v_1, \ldots, v_k, v_{k+1}, s_2, \ldots, s_{m-k})$$

or said differently,

$$V = Span(v_1, \ldots, v_k, s_1, \ldots, s_{m-k}) \subseteq Span(v_1, \ldots, v_{k+1}, s_2, \ldots, s_{m-k}) \ .$$

This completes the proof. □

This lemma has loads of consequences.

Definition 1.3.7. *A set $B \subseteq V$ is called a basis for V if B is linearly independent and $Span(B) = V$.*

Corollary 1.3.8. *If B_1 and B_2 are bases for V then they have the same number of elements.*

Proof. If both sets are infinite, we are done. If B_1 is infinite and B_2 is finite then we may choose $|B_2| + 1$ elements of B_1 that will be linearly independent. Since B_2 spans V, this contradicts Steinitz. Similarly if B_1 is finite and B_2 is infinite.

Otherwise they are both finite. Since each spans V and is linearly independent, we apply Steinitz twice to get $|B_1| \leq |B_2|$ and $|B_2| \leq |B_1|$. □

Definition 1.3.9. *We define the dimension dim V to be the number of elements in a basis. By the above, this is well-defined.*

Remark. *Each nonzero element of V has a unique representation in terms of a basis.*

Theorem 1.3.10. *Let V be a nonzero vector space and suppose that V is finitely generated; that is, there is a finite set $S \subseteq V$ such that $V = Span(S)$. Then V has a finite basis: $dim\ V < \infty$.*

Proof. Let B be a minimal spanning subset of S. We claim that B is linearly independent. If not, then by a previous result, there is a vector $b \in B$ such that $b \in Span(B \setminus \{b\})$. It then follows that

$$V = Span(B) \subseteq Span(B \setminus \{b\}) \ ,$$

so $B \setminus \{b\}$ is a spanning set, a contradiction. □

Theorem 1.3.11 (1 subspace theorem)**.** *Let V be a finite dimensional vector space and W be a nonzero subspace of V. Then $dim\ W < \infty$. If $C = \{w_1, \ldots, w_k\}$ is a basis for W then there exists a basis B for V such that $C \subseteq B$.*

Proof. It is an exercise to show that $dim\ W < \infty$. Write n for the dimension of V and let B be a basis for V (with n elements). Since this is a spanning set and C is a linearly independent set, there exist vectors $b_1, \ldots, b_{n-k} \in B$ such that

$$B := C \cup \{b_1, \ldots, b_{n-k}\}$$

is a spanning set. Note that B is a spanning set with $n = dim\ V$ number of elements. Thus we will be done if we prove the following lemma. □

Lemma 1.3.12. *Let V be a vector space of dimension $n \geq 1$ and $S = \{v_1, \ldots, v_k\} \subseteq V$.*

1. *If $k < n$ then S cannot span V.*
2. *If $k > n$ then S cannot be linearly independent.*
3. *If $k = n$ then S is linearly independent if and only if S spans V.*

Proof. Let B be a basis of V. If S spans V, Steinitz gives that $|B| \leq |S|$. This proves 1. If S is linearly independent then again Steinitz gives $|S| \leq |B|$. This proves 2.

Suppose that $k = n$ and S is linearly independent. By Steinitz, we may add 0 vectors from the set B to S to make S span V. Thus $Span(S) = V$. Conversely, suppose that $Span(S) = V$. If S is linearly dependent then there exists $s \in S$ such that $s \in Span(S \setminus \{s\})$. Then $S \setminus \{s\}$ is a set of smaller cardinality than S and spans V. But now this contradicts Steinitz, using B as our linearly independent set and $S \setminus \{s\}$ as our spanning set. □

Corollary 1.3.13. *If V is a vector space and W is a subspace then $dim\ W \leq dim\ V$.*

Theorem 1.3.14. *Let W_1 and W_2 be subspaces of V (with $dim\ V < \infty$). Then*

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

Proof. Easy if either is zero. Otherwise we argue as follows. Let

$$\{v_1, \ldots, v_k\}$$

be a basis for $W_1 \cap W_2$. By the 1-subspace theorem, extend this to a basis of W_1:

$$\{v_1, \ldots, v_k, w_1, \ldots, w_{m_1}\}$$

and also extend it to a basis for W_2:

$$\{v_1, \ldots, v_k, \hat{w}_1, \ldots, \hat{w}_{m_2}\} \ .$$

We claim that

$$B := \{v_1, \ldots, v_k, w_1, \ldots, w_{m_1}, \hat{w}_1, \ldots, \hat{w}_{m_2}\}$$

is a basis for $W_1 + W_2$. It is not hard to see it is spanning.

To show linear independence, suppose

$$a_1 v_1 + \cdots + a_k v_k + b_1 w_1 + \cdots + b_{m_1} w_{m_1} + c_1 \hat{w}_1 + \cdots + c_{m_2} \hat{w}_{m_2} = \vec{0} \ .$$

Then

$$c_1 \hat{w}_1 + \cdots + c_{m_2} \hat{w}_{m_2} = (-a_1) v_1 + \cdots + (-a_k) v_k + (-b_1) w_1 + \cdots + (-b_{m_1}) w_{m_1} \in W_1 \ .$$

Also it is clearly in W_2. So it is in $W_1 \cap W_2$. So we can find scalars $\tilde{a}_1, \ldots, \tilde{a}_k$ such that

$$c_1 \hat{w}_1 + \cdots + c_{m_2} \hat{w}_{m_2} = \tilde{a}_1 v_1 + \cdots + \tilde{a}_k v_k$$

$$\tilde{a}_1 v_1 + \cdots + \tilde{a}_k v_k + (-c_1) \hat{w}_1 + \cdots + (-c_{m_2}) \hat{w}_{m_2} = \vec{0} \ .$$

This is a linear combination of basis elements, so all c_i's are zero. A similar argument gives all b_i's as zero. Thus we finally have

$$a_1 v_1 + \cdots + a_k v_k = \vec{0} \ .$$

But again this is a linear combination of basis elements so the a_i's are zero. □

Theorem 1.3.15 (2 subspace theorem)**.** *Let V be a finite dimensional vector space and W_1 and W_2 be nonzero subspaces. There exists a basis of V that contains bases for W_1 and W_2.*

Proof. The proof of the previous theorem shows that there is a basis for $W_1 + W_2$ which contains bases for W_1 and W_2. Extend this to a basis for V. □

1.4 Exercises

Notation:

1. If F is a field then define $s_1 : F \to F$ by $s_1(x) = x$. For integers $n \geq 2$ define $s_n : F \to F$ by $s_n(x) = x + s_{n-1}(x)$. Last, define the *characteristic* of F as

$$char(F) = \min\{n : s_n(1) = 0\} \ .$$

   If $s_n(1) \neq 0$ for all $n \geq 1$ then we set $char(F) = 0$.

Exercises:

1. **The finite field $\mathbb{F}_p$**: For $n \in \mathbb{N}$, let $\mathbb{Z}_n$ denote the set of integers mod n. That is, each element of $\mathbb{Z}/n\mathbb{Z}$ is a subset of $\mathbb{Z}$ of the form $d + n\mathbb{Z}$, where $d \in Z$. We define addition and multiplication on $\mathbb{Z}_n$ by

   - $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$.
   - $(a + n\mathbb{Z}) \cdot (b + n\mathbb{Z}) = (ab) + n\mathbb{Z}$.

   Show that these operations are well defined. That is, if $a', b' \in \mathbb{Z}$ are integers such that $a + n\mathbb{Z} = a' + n\mathbb{Z}$ and $b + n\mathbb{Z} = b' + n\mathbb{Z}$, then $(a' + n\mathbb{Z}) + (b' + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$ and $(a' + n\mathbb{Z}) \cdot (b' + n\mathbb{Z}) = (ab) + n\mathbb{Z}$. Moreover, show that these operations make $\mathbb{Z}_n$ into a field if and only if n is prime. In that case, one writes $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$.

2. **The finite field $\mathbb{F}_{p^n}$**: Let F be a finite field.

   (a) Show that $char(F)$ is a prime number (in particular non-zero).

   (b) Write p for the characteristic of F and define

$$\overline{F} = \{s_n(1) : n = 1, \ldots, p\} \ ,$$

   where s_n is the function given in the notation section. Show that $\overline{F}$ is a subfield of F, isomorphic to $\mathbb{F}_p$.

   (c) We can consider F as a vector space over $\overline{F}$. Vector addition and scalar multiplication are interpreted using the operations of F. Show that F has finite dimension.

   (d) Writing n for the dimension of F, show that

$$|F| = p^n \ .$$

3. Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$ (using addition and multiplication of real numbers). Does this vector space have finite dimension?

4. Recall the definition of direct sum: if W_1 and W_2 are subspaces of a vector space V then we write $W_1 \oplus W_2$ for the space $W_1 + W_2$ if $W_1 \cap W_2 = \{\vec{0}\}$. For $k \geq 3$ we write $W_1 \oplus \cdots \oplus W_k$ for the space $W_1 + \cdots + W_k$ if for each $i = 2, \ldots, k$, we have

$$W_i \cap (W_1 + \cdots + W_{i-1}) = \{\vec{0}\} \ .$$

Let $S = \{v_1, \ldots, v_n\}$ be a subset of nonzero vectors in a vector space V and for each $k = 1, \ldots, n$ write $W_k = Span(\{v_k\})$. Show that S is a basis for V if and only if

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

2 Linear transformations

2.1 Definitions and basic properties

We now move to linear transformations.

Definition 2.1.1. *Let V and W be vector spaces over the same field F. A function $T : V \to W$ is called a linear transformation if*

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

Remark. *We only need to check that for all $v_1, v_2 \in V$, $c \in F$, $T(cv_1 + v_2) = cT(v_1) + T(v_2)$.*

Examples:

1. Let $V = F^n$ and $W = F^m$, the vector spaces of n-tuples and m-tuples respectively. Any $m \times n$ matrix A defines a linear transformation $L_A : F^n \to F^m$ by
$$L_A \vec{v} = A \cdot \vec{v} \ .$$
2. Let V be a finite dimensional vector space (of dimension n) and fix an (ordered) basis
$$\beta = \{v_1, \ldots, v_n\}$$
of V. Define $T_\beta : V \to F^n$ by
$$T_\beta(v) = (a_1, \ldots, a_n) \ ,$$
where $v = a_1 v_1 + \cdots + a_n v_n$. Then T_β is linear. It is called the *coordinate map* for the ordered basis β.

Suppose that $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation and $\vec{x} \in \mathbb{F}^n$. Then writing $\vec{x} = (x_1, \ldots, x_n)$,

$$\begin{aligned} T(\vec{x}) &= T((x_1, \ldots, x_n)) \\ &= T(x_1(1, 0, \ldots, 0) + \cdots + x_n(0, \ldots, 0, 1)) \\ &= x_1 T((1, 0, \ldots, 0)) + x_n T((0, \ldots, 0, 1)) \ . \end{aligned}$$

Therefore we only need to know the values of T at the standard basis. This leads us to:

Theorem 2.1.2 (The slogan)**.** *Given V and W, vector spaces over F, let $\{v_1, \ldots, v_n\}$ be a basis for V. If $\{w_1, \ldots, w_n\}$ are any vectors in W, there exists exactly one linear transformation $T : V \to W$ such that*

$$T(v_i) = w_i \text{ for all } i = 1, \ldots, n \ .$$

Proof. We need to prove two things: (a) there is such a linear transformation and (b) there cannot be more than one. Motivated by the above, we first prove (a).

Each $v \in V$ has a unique representation

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

Define T by

$$T(v) = a_1 w_1 + \cdots + a_n w_n \ .$$

Note that by unique representations, T is well-defined. We claim that T is linear. Let $v, \widetilde{v} \in V$ and $c \in F$. We must show that $T(cv + \widetilde{v}) = cT(v) + T(\widetilde{v})$. If $v = \sum_{i=1}^n a_i v_i$ and $\widetilde{v} = \sum_{i=1}^n \widetilde{a}_i v_i$ then we claim that the unique representation of $cv + \widetilde{v}$ is

$$cv + \widetilde{v} = (ca_1 + \widetilde{a}_1)v_1 + \cdots + (ca_n + \widetilde{a}_n)v_n \ .$$

Therefore

$$\begin{aligned} T(cv + \widetilde{v}) &= (ca_1 + \widetilde{a}_1)w_1 + \cdots + (ca_n + \widetilde{a}_n)w_n \\ &= c(a_1 w_1 + \cdots + a_n w_n) + (\widetilde{a}_1 w_1 + \cdots + \widetilde{a}_n w_n) \\ &= cT(v) + T(\widetilde{v}) \ . \end{aligned}$$

Thus T is linear.

Now we show that T is unique. Suppose that T' is another linear transformation such that

$$T'(v_i) = w_i \text{ for all } i = 1, \ldots, n \ .$$

Then if $v \in V$ write $v = \sum_{i=1}^n a_i v_i$. We have

$$\begin{aligned} T'(v) &= T'(a_1 v_1 + \cdots + a_n v_n) \\ &= a_1 T'(v_1) + \cdots + a_n T'(v_n) \\ &= a_1 w_1 + \cdots + a_n w_n \\ &= T(v) \ . \end{aligned}$$

Since $T(v) = T'(v)$ for all v, by definition $T = T'$.

□

2.2 Range and nullspace, one-to-one, onto

Definition 2.2.1. *If $T : V \to W$ is linear we define the nullspace of T by*

$$N(T) = \{v \in V : T(v) = \vec{0}\} \ .$$

Here, $\vec{0}$ is the zero vector in W. We also define the range of T

$$R(T) = \{w \in W : \text{ there exists } v \in V \text{ s.t. } T(v) = w\} \ .$$

Proposition 2.2.2. *If $T : V \to W$ is linear then $N(T)$ is a subspace of V and $R(T)$ is a subspace of W.*

Proof. First note that $T(\vec{0}) = \vec{0}$. This holds from

$$\vec{0} = 0T(v) = T(0v) = T(\vec{0}) \ .$$

Therefore both spaces are nonempty.

Choose $v_1, v_2 \in N(T)$ and $c \in F$. Then

$$T(cv_1 + v_2) = cT(v_1) + T(v_2) = c\vec{0} + \vec{0} = \vec{0} \ ,$$

so that $cv_1 + v_2 \in N(T)$. Therefore $N(T)$ is a subspace of V. Also if $w_1, w_2 \in R(T)$ and $c \in F$ then we may find $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Now

$$T(cv_1 + v_2) = cT(v_1) + T(v_2) = cw_1 + w_2 \ ,$$

so that $cw_1 + w_2 \in R(T)$. Therefore $R(T)$ is a subspace of W.

□

Definition 2.2.3. *Let S and T be sets and $f : S \to T$ a function.*

1. *f is one-to-one if it maps distinct points to distinct points. In other words if $s_1, s_2 \in S$ such that $s_1 \neq s_2$ then $f(s_1) \neq f(s_2)$. Equivalently, whenever $f(s_1) = f(s_2)$ then $s_1 = s_2$.*

2. *f is onto if its range is equal to T. That is, for each $t \in T$ there exists $s \in S$ such that $f(s) = t$.*

Theorem 2.2.4. *Let $T : V \to W$ be linear. Then T is one-to-one if and only if $N(T) = \{\vec{0}\}$.*

Proof. Suppose T is one-to-one. We want to show that $N(T) = \{\vec{0}\}$. Clearly $\vec{0} \in N(T)$ as it is a subspace of V. If $v \in N(T)$ then we have

$$T(v) = \vec{0} = T(\vec{0}) \ .$$

Since T is one-to-one this implies that $v = \vec{0}$.

Suppose conversely that $N(T) = \{\vec{0}\}$. If $T(v_1) = T(v_2)$ then

$$\vec{0} = T(v_1) - T(v_2) = T(v_1 - v_2) \ ,$$

so that $v_1 - v_2 \in N(T)$. But the only vector in the nullspace is $\vec{0}$ so $v_1 - v_2 = \vec{0}$. This implies that $v_1 = v_2$ and T is one-to-one. □

We now want to give a theorem that characterizes one-to-one and onto linear maps in a different way.

Theorem 2.2.5. *Let $T : V \to W$ be linear.*

1. *T is one-to-one if and only if it maps linearly independent sets in V to linearly independent sets in W.*

2. *T is onto if and only if it maps spanning sets of V to spanning sets of W.*

Proof. Suppose that T is one-to-one and that S is a linearly independent set in V. We will show that $T(S)$, defined by

$$T(S) := \{T(s) : s \in S\} \ ,$$

is also linearly independent. If

$$a_1 T(s_1) + \cdots + a_k T(s_k) = \vec{0}$$

for some $s_i \in S$ and $a_i \in F$ then

$$T(a_1 s_1 + \cdots + a_k s_k) = \vec{0} \ .$$

Therefore $a_1 s_1 + \cdots + a_k s_k \in N(T)$. But T is one-to-one so $N(T) = \{\vec{0}\}$. This gives

$$a_1 s_1 + \cdots + a_k s_k = \vec{0} \ .$$

Linear independence of S gives the a_i's are zero. Thus $T(S)$ is linearly independent.

Suppose conversely that T maps linearly independent sets to linearly independent sets. If v is any nonzero vector in V then $\{v\}$ is linearly independent. Therefore so is $\{T(v)\}$. This implies that $T(v) \neq 0$. Therefore $N(T) = \{\vec{0}\}$ and so T is one-to-one.

If T maps spanning sets to spanning sets then let $w \in W$. Let S be a spanning set of V, so that consequently $T(S)$ spans W. If $w \in W$ we can write $w = a_1 T(s_1) + \cdots + a_k T(s_k)$ for $a_i \in F$ and $s_i \in S$, so

$$w = T(a_1 s_1 + \cdots + a_k s_k) \in R(T) \ ,$$

giving that T is onto.

For the converse suppose that T is onto and that S spans V. We claim that $T(S)$ spans W. To see this, let $w \in W$ and note there exists $v \in V$ such that $T(v) = w$. Write

$$v = a_1 s_1 + \cdots + a_k s_k \ ,$$

so $w = T(v) = a_1 T(s_1) + \cdots + a_k T(s_k) \in Span(T(S))$. Therefore $T(S)$ spans W. □

Corollary 2.2.6. *Let $T : V \to W$ be linear.*

1. *if V and W are finite dimensional, then T is an isomorphism (one-to-one and onto) if and only if T maps bases to bases.*

2. *If V is finite dimensional, then every basis of V is mapped to a spanning set of $R(T)$.*

3. *If V is finite dimensional, then T is one-to-one if and only if T maps bases of V to bases of $R(T)$.*

Theorem 2.2.7 (Rank-nullity theorem)**.** *Let $T : V \to W$ be linear and V of finite dimension. Then*

$$rank(T) + nullity(T) = dim\ V\ .$$

Proof. Let

$$\{v_1, \ldots, v_k\}$$

be a basis for $N(T)$. Extend it to a basis

$$\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$$

of V. Write $N' = Span\{v_{k+1}, \ldots, v_n\}$ and note that $V = N(T) \oplus N'$.

Lemma 2.2.8. *If $N(T)$ and N' are complementary subspaces (that is, $V = N(T) \oplus N'$) then T is one-to-one on N'.*

Proof. If $z_1, z_2 \in N'$ such that $T(z_1) = T(z_2)$ then $z_1 - z_2 \in N(T)$. But it is in N' so it is in $N' \cap N(T)$, which is only the zero vector. So $z_1 = z_2$. □

We may view T as a linear transformation only on N'; call it $T|_{N'}$; in other words, $T|_{N'}$ is a linear transformation from N' to W that acts exactly as T does. By the corollary, $\{T|_{N'}(v_{k+1}), \ldots, T|_{N'}(v_n)\}$ is a basis for $R(T|_{N'})$. Therefore $\{T(v_{k+1}), \ldots, T(v_n)\}$ is a basis for $R(T|_{N'})$. By part two of the corollary, $\{T(v_1), \ldots, T(v_n)\}$ spans $R(T)$. So

$$\begin{aligned} R(T) &= Span(\{T(v_1), \ldots, T(v_n)\}) \\ &= Span(\{T(v_{k+1}), \ldots, T(v_n)\}) \\ &= = R(T|_{N'})\ . \end{aligned}$$

The second equality follows because the vectors $T(v_1), \ldots, T(v_k)$ are all zero and do not contribute to the span (you can work this out as an exercise). Thus $\{T(v_{k+1}), \ldots, T(v_n)\}$ is a basis for $R(T)$ and

$$rank(T) + nullity(T) = (n - k) + k = n = dim\ V\ .$$

□

2.3 Isomorphisms and $\mathcal{L}(V, W)$

Definition 2.3.1. *If S and T are sets and $f : S \to T$ is a function then a function $g : T \to S$ is called an inverse function for f (written f^{-1}) if*

$$f(g(t)) = t \text{ and } g(f(s)) = s \text{ for all } t \in T, s \in S\ .$$

Fact: $f : S \to T$ has an inverse function if and only if f is one-to-one and onto. Furthermore the inverse is one-to-one and onto. (Explain this.)

Theorem 2.3.2. *If $T : V \to W$ is an isomorphism then the inverse map $T^{-1} : W \to V$ is an isomorphism.*

Proof. We have one-to-one and onto. We just need to show linear. Suppose that $w_1, w_2 \in W$ and $c \in F$. Then

$$T\left(T^{-1}(cw_1 + w_2)\right) = cw_1 + w_2$$

and

$$T\left(cT^{-1}(w_1) + T^{-1}(w_2)\right) = cT(T^{-1}(w_1)) + T(T^{-1}(w_2)) = cw_1 + w_2 \ .$$

However T is one-to-one, so

$$T^{-1}(cw_1 + w_2) = cT^{-1}(w_1) + T^{-1}(w_2) \ .$$

□

The proof of the next lemma is in the homework.

Lemma 2.3.3. *Let V and W be vector spaces with dim V = dim W. If $T : V \to W$ is linear then T is one-to-one if and only if T is onto.*

Example: The coordinate map is an isomorphism. For V of dimension n choose a basis

$$\beta = \{v_1, \ldots, v_n\}$$

and define $T_\beta : V \to F^n$ by $T(v) = (a_1, \ldots, a_n)$, where

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

Then T_β is linear (check). To show one-to-one and onto we only need to check one (since *dim* V = *dim* F^n. If $(a_1, \ldots, a_n) \in F^n$ then define $v = a_1 v_1 + \cdots + a_n v_n$. Now

$$T_\beta(v) = (a_1, \ldots, a_n) \ .$$

So T is onto.

The space of linear maps. Let V and W be vector spaces over the same field F. Define

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

We define addition and scalar multiplication as usual: for $T, U \in \mathcal{L}(V, W)$ and $c \in F$,

$$(T + U)(v) = T(v) + U(v) \text{ and } (cT)(v) = cT(v) \ .$$

This is a vector space (exercise).

Theorem 2.3.4. *If dim $V = n$ and dim $W = m$ then dim $\mathcal{L}(V, W) = mn$. Given bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ of V and W, the set $\{T_{i,j} : 1 \le i \le n, 1 \le j \le m\}$ is a basis for $\mathcal{L}(V, W)$, where*

$$T_{i,j}(v_k) = \begin{cases} w_j & \text{if } i = k \\ \vec{0} & \text{otherwise} \end{cases} \ .$$

Proof. First, to show linear independence, suppose that

$$\sum_{i,j} a_{i,j} T_{i,j} = 0_T \ ,$$

where the element on the right is the zero transformation. Then for each $k = 1, \ldots, n$, apply both sides to v_k:

$$\sum_{i,j} a_{i,j} T_{i,j}(v_k) = 0_T(v_k) = \vec{0} \ .$$

We then get

$$\vec{0} = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{i,j} T_{i,j}(v_k) = \sum_{j=1}^{m} a_{k,j} w_j \ .$$

But the w_j's form a basis, so all $a_{k,1}, \ldots, a_{k,m} = 0$. This is true for all k so the $T_{i,j}$'s are linearly independent.

To show spanning suppose that $T : V \to W$ is linear. Then for each $i = 1, \ldots, n$, the vector $T(v_i)$ is in W, so we can write it in terms of the w_j's:

$$T(v_i) = a_{i,1} w_1 + \cdots + a_{i,m} w_m \ .$$

Now define the transformation

$$\widetilde{T} = \sum_{i,j} a_{i,j} T_{i,j} \ .$$

We claim that this equals T. To see this, we must only check on the basis vectors. For some $k = 1, \ldots, n$,

$$T(v_k) = a_{k,1} w_1 + \cdots + a_{k,m} w_m \ .$$

However,

$$\begin{aligned} \widetilde{T}(v_k) &= \sum_{i,j} a_{i,j} T_{i,j}(v_k) = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{i,j} T_{i,j}(v_k) \\ &= \sum_{j=1}^{m} a_{k,j} w_j \\ &= a_{k,1} w_1 + \cdots + a_{k,m} w_m \ . \end{aligned}$$

□

2.4 Matrices and coordinates

Let $T : V \to W$ be a linear transformation and

$$\beta = \{v_1, \ldots, v_n\} \text{ and } \gamma = \{w_1, \ldots, w_m\}$$

bases for V and W, respectively.

We now build a matrix, which we label $[T]_\gamma^\beta$. (Using the **column convention**.)

1. Since $T(v_1) \in W$, we can write
$$T(v_1) = a_{1,1}w_1 + \cdots + a_{m,1}w_m \ .$$
2. Put the entries $a_{1,1}, \ldots, a_{m,n}$ into the first column of $[T]_\gamma^\beta$.
3. Repeat for $k = 1, \ldots, n$, writing
$$T(v_k) = a_{1,k}w_1 + \cdots + a_{m,k}w_m$$
and place the entries $a_{1,k}, \ldots, a_{m,k}$ into the k-th column.

Theorem 2.4.1. *For each $T : V \to W$ and bases β and γ, there exists a unique $m \times n$ matrix $[T]_\gamma^\beta$ such that for all $v \in V$,*
$$[T]_\gamma^\beta [v]_\beta = [T(v)]_\gamma \ .$$

Proof. Let $v \in V$. Then write $v = a_1v_1 + \cdots + a_nv_n$.
$$\begin{aligned} T(v) &= a_1T(v_1) + \cdots + a_nT(v_n) \\ &= a_1(a_{1,1}w_1 + \cdots + a_{m,1}w_m) + \cdots + a_n(a_{1,n}w_1 + \cdots + a_{m,n}w_m) \ . \end{aligned}$$
Collecting terms,
$$T(v) = (a_1a_{1,1} + \cdots + a_na_{1,n})w_1 + \cdots + (a_1a_{m,1} + \cdots + a_na_{m,n})w_m \ .$$
This gives the coordinates of $T(v)$ in terms of γ:
$$[T(v)]_\gamma = \begin{bmatrix} a_1a_{1,1} + \cdots + a_na_{1,n} \\ \cdots \\ a_1a_{m,1} + \cdots + a_na_{m,n} \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \cdots & \cdots & \cdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} a_1 \\ \cdots \\ a_n \end{bmatrix} = [T]_\gamma^\beta [v]_\beta \ .$$
Suppose that A and B are two matrices such that for all $v \in V$,
$$A[v]_\beta = [T(v)]_\gamma = B[v]_\beta \ .$$
Take $v = v_k$. Then $[v]_\beta = \vec{e}_k$ and $A[v]_\beta$ is the k-th column of A (and similarly for B). Therefore A and B have all the same columns. This means $A = B$.

□

Theorem 2.4.2. *Given bases β and γ of V and W, the map $T \mapsto [T]_\gamma^\beta$ is an isomorphism.*

Proof. It is one-to-one. To show linear let $c \in F$ and $T, U \in \mathcal{L}(V, W)$. For each $k = 1, \ldots, n$,
$$\begin{aligned} [cT + U]_\gamma^\beta [v_k]_\beta &= [cT(v) + U(v)]_\beta = c[T(v_k)]_\beta + [U(v_k)]_\beta = c[T]_\gamma^\beta [v_k]_\beta + [U]_\gamma^\beta [v_k]_\beta \\ &= (c[T]_\gamma^\beta + [U]_\gamma^\beta)[v_k]_\beta \ . \end{aligned}$$
However the left side is the k-th column of $[cT + U]_\gamma^\beta$ and the right side is the k-th column of $c[T]_\gamma^\beta + [U]_\gamma^\beta$.

Both spaces have the same dimension, so the map is also onto and thus an isomorphism.

□

Examples:

1. **Change of coordinates:** Let V be finite dimensional and β, β' two bases of V. How do we express coordinates in β in terms of those in β'? For each $v \in V$,

$$[v]_{\beta'} = [Iv]_{\beta'} = [I]^{\beta}_{\beta'}[v]_{\beta} \ .$$

   We multiply by the matrix $[I]^{\beta}_{\beta'}$.

2. Suppose that V, W and Z are vector spaces over the same field. Let $T : V \to W$ and $U : W \to Z$ be linear with β, γ and δ bases for V, W and Z.

   (a) $UT : V \to Z$ is linear. If $v_1, v_2 \in V$ and $c \in F$ then

$$(UT)(cv_1 + v_2) = U(T(cv_1 + v_2)) = U(cT(v_1) + T(v_2)) = c(UT)(v_1) + (UT)(v_2) \ .$$

   (b) For each $v \in V$,

$$[(UT)v]_{\delta} = [U(T(v))]_{\delta} = [U]^{\gamma}_{\delta}[T(v)]_{\gamma} = [U]^{\gamma}_{\delta}[T]^{\beta}_{\gamma}[v]_{\beta} \ .$$

   Therefore

$$[UT]^{\beta}_{\delta} = [U]^{\gamma}_{\delta}[T]^{\beta}_{\gamma} \ .$$

   (c) If T is an isomorphism from V to W,

$$Id = [I]^{\beta}_{\beta} = [T]^{\beta}_{\gamma}[T^{-1}]^{\gamma}_{\beta} \ .$$

   Similarly,

$$Id = [I]^{\gamma}_{\gamma} = [T^{-1}]^{\gamma}_{\beta}[T]^{\beta}_{\gamma}$$

   This implies that $[T^{-1}]^{\gamma}_{\beta} = \left([T]^{\beta}_{\gamma}\right)^{-1}$.

3. To change coordinates back,

$$[I]^{\beta'}_{\beta} = \left([I^{-1}]^{\beta}_{\beta'}\right)^{-1} = \left([I]^{\beta}_{\beta'}]\right)^{-1} \ .$$

Definition 2.4.3. *An $n \times n$ matrix A is invertible is there exists an $n \times n$ matrix B such that*

$$I = AB = BA \ .$$

Remark. *If β, β' are bases of V then*

$$I = [I]^{\beta}_{\beta} = [I]^{\beta'}_{\beta}[I]^{\beta}_{\beta'} \ .$$

Therefore each change of basis matrix is invertible.

Now how do we relate the matrix of T with respect to different bases?

Theorem 2.4.4. *Let V and W be finite-dimensional vector spaces over F with $\beta, \widetilde{\beta}$ bases for V and $\gamma, \widetilde{\gamma}$ bases for W.*

1. *If $T : V \to W$ is linear then there exist invertible matrices P and Q such that*
$$[T]_{\gamma}^{\beta} = P[T]_{\gamma'}^{\beta'}Q \ .$$
2. *If $T : V \to W$ is linear then there exists an invertible matrix P such that*
$$[T]_{\beta}^{\beta} = P^{-1}[T]_{\beta'}^{\beta'}P \ .$$

Proof.
$$[T]_{\gamma}^{\beta} = [I]_{\gamma}^{\gamma'}[T]_{\gamma'}^{\beta'}[I]_{\beta'}^{\beta} \ .$$
Also
$$[T]_{\beta}^{\beta} = [I]_{\beta}^{\beta'}[T]_{\beta'}^{\beta'}[I]_{\beta'}^{\beta} = \left([I]_{\beta'}^{\beta}\right)^{-1}[T]_{\beta'}^{\beta'}[I]_{\beta'}^{\beta} \ .$$
□

Definition 2.4.5. *Two $n \times n$ matrices A and B are similar if there exists an $n \times n$ invertible matrix P such that*
$$A = P^{-1}BP \ .$$

Theorem 2.4.6. *Let A and B be $n \times n$ matrices with entries from F. If A and B are similar then there exists an n-dimensional vector space V, a linear transformation $T : V \to V$, and bases β, β' such that*
$$A = [T]_{\beta}^{\beta} \text{ and } B = [T]_{\beta'}^{\beta'} \ .$$

Proof. Suppose that $A = P^{-1}BP$. Define the linear transformation $L_A : F^n \to F^n$ by
$$L_A(\vec{v}) = A \cdot \vec{v} \ .$$
If we choose β to be the standard basis then
$$[L_A]_{\beta}^{\beta} = A \ .$$

Next we will show that if $\beta' = \{\vec{p}_1, \ldots, \vec{p}_n\}$ are the columns of P then β' is a basis and $P = [I]_{\beta}^{\beta'}$, where $I : F^n \to F^n$ is the identity map. If we prove this, then $P^{-1} = [I]_{\beta'}^{\beta}$ and so
$$B = P^{-1}AP = [I]_{\beta'}^{\beta}[L_A]_{\beta}^{\beta}[I]_{\beta}^{\beta'} = [L_A]_{\beta'}^{\beta'} \ .$$
Why is β' a basis? Note that
$$\vec{p}_k = P \cdot \vec{e}_k \ ,$$
so that if L_P is invertible then β' will be the image of a basis and thus a basis. But for all $\vec{v} \in F^n$,
$$L_{P^{-1}}L_P\vec{v} = P^{-1} \cdot P \cdot \vec{v} = \vec{v}$$
and
$$L_P L_{P^{-1}}\vec{v} = \vec{v} \ .$$
So $(L_P)^{-1} = L_{P^{-1}}$. This completes the proof. □

The moral: Similar matrices represent the same transformation but with respect to two different bases. Any property of matrices that is invariant under conjugation can be viewed as a property of the underlying transformation.

Example: Trace. Given an $n \times n$ matrix A with entries from F, define

$$Tr\ A = \sum_{i=1}^{n} a_{i,i}\ .$$

Note that if P is another matrix (not nec. invertible),

$$Tr\ (AP) = \sum_{i=1}^{n} (AP)_{i,i} = \sum_{i=1}^{n} \left[\sum_{l=1}^{n} a_{i,l} p_{l,i} \right] = \sum_{l=1}^{n} \left[\sum_{i=1}^{n} p_{l,i} a_{i,l} \right] = \sum_{l=1}^{n} (PA)_{l,l} = Tr\ (PA)\ .$$

Therefore if P is invertible,

$$Tr\ (P^{-1}AP) = Tr\ (APP^{-1}) = Tr\ A\ .$$

This means that trace is invariant under conjugation. Thus if $T : V \to V$ is linear (and V is finite dimensional) then $Tr\ T$ can be defined as

$$Tr\ [T]_{\beta}^{\beta}$$

for any basis β.

2.5 Exercises

Notation:

1. A *group* is a pair $(G, \cdot)$, where G is a set and $\cdot : G \times G \to G$ is a function (usually called product) such that

   (a) there is an identity element e; that is, an element with the property

   $$e \cdot g = g \cdot e = g \text{ for all } g \in G\ ,$$

   (b) for all $g \in G$ there is an inverse element in G called g^{-1} such that

   $$g^{-1} \cdot g = g \cdot g^{-1} = e\ ,$$

   (c) and the operation is associative: for all $g, h, k \in G$,

   $$(g \cdot h) \cdot k = g \cdot (h \cdot k)\ .$$

   If the operation is commutative; that is, for all $g, h \in G$, we have $g \cdot h = h \cdot g$ then we call G *abelian*.

2. If $(G, \cdot)$ is a group and H is a subset of G then we call H a subgroup of G if $(H, \cdot|_{H\times H})$ is a group. Equivalently (and analogously to vector spaces and subspaces), $H \subseteq G$ is a subgroup of G if and only if

   (a) for all $h_1, h_2 \in H$ we have $h_1 \cdot h_2 \in H$ and

   (b) for all $h \in H$, we have $h^{-1} \in H$.

3. If G and H are groups then a function $\Phi : G \to H$ is called a *group homomorphism* if

$$\Phi(g_1 \cdot g_2) = \Phi(g_1) \cdot \Phi(g_2) \text{ for all } g_1 \text{ and } g_2 \in G \ .$$

   Note that the product on the left is in G whereas the product on the right is in H. We define the *kernel* of Φ to be

$$Ker(\Phi) = \{g \in G : \Phi(g) = e_H\} \ .$$

   Here, e_H refers to the identity element of H.

4. A group homomorphism $\Phi : G \to H$ is called

   - a *monomorphism* (or injective, or one-to-one), if $\Phi(g_1) = \Phi(g_2) \Rightarrow g_1 = g_2$;
   - an *epimorphism* (or surjective, or onto), if $\Phi(G) = H$;
   - an *isomorphism* (or bijective), if it is both injective and surjective.

   A group homomorphism $\Phi : G \to G$ is called an *endomorphism* of G. An endomorphism which is also an isomorphism is called an *automorphism*. The set of automorphisms of a group G is denoted by $\mathrm{Aut}(G)$.

5. Recall that, if F is a field (with operations $+$ and $\cdot$) then $(F, +)$ and $(F \setminus \{0\}, \cdot)$ are abelian groups, with identity elements 0 and 1 respectively. If F and G are fields and $\Phi : F \to G$ is a function then we call Φ a *field homomorphism* if

   (a) for all $a, b \in F$ we have
$$\Phi(a + b) = \Phi(a) + \Phi(b) \ ,$$

   (b) for all $a, b \in F$ we have
$$\Phi(ab) = \Phi(a)\Phi(b) \ ,$$

   (c) and $\Phi(1_F) = 1_G$. Here 1_F and 1_G are the multiplicative identities of F and G respectively.

6. If V and W are vector spaces over the same field F then we define their *product* to be the set
$$V \times W = \{(v, w) : v \in V \text{ and } w \in W\} \ .$$
   which becomes a vector space under the operations

$$(v, w)+(v', w') = (v+v', w+w'), \qquad c(v, w) = (cv, cw) \qquad v, v' \in V, w, w' \in W, c \in F.$$

If you are familiar with the notion of an external direct sum, notice that the product of two vector spaces is the same as their external direct sum. The two notions cease being equivalent when one considers infinitely many factors/summands.

If Z is another vector space over F then we call a function $f : V \times W \to Z$ *bilinear* if

(a) for each fixed $v \in V$, the function $f_v : W \to Z$ defined by

$$f_v(w) = f((v, w))$$

is a linear transformation as a function of w and

(b) for each fixed $w \in W$, the function $f_w : V \to Z$ defined by

$$f_w(v) = f((v, w))$$

is a linear transformation as a function of v.

Exercises:

1. Suppose that G and H are groups and $\Phi : G \to H$ is a homomorphism.

   (a) Prove that if H' is a subgroup of H then the *inverse image*

   $$\Phi^{-1}(H') = \{g \in G : \Phi(g) \in H'\}$$

   is a subgroup of G. Deduce that $Ker(\Phi)$ is a subgroup of G.

   (b) Prove that if G' is a subgroup of G, then the *image* of G' under Φ,

   $$\Phi(G') = \{\Phi(g) | g \in G'\}$$

   is a subgroup of H.

   (c) Prove that Φ is one-to-one if and only if $Ker(\Phi) = \{e_G\}$. (Here, e_G is the identity element of G.)

2. Prove that every field homomorphism is one-to-one.

3. Let V and W be finite dimensional vector spaces with $dim\ V = n$ and $dim\ W = m$. Suppose that $T : V \to W$ is a linear transformation.

   (a) Prove that if $n > m$ then T cannot be one-to-one.

   (b) Prove that if $n < m$ then T cannot be onto.

   (c) Prove that if $n = m$ then T is one-to-one if and only if T is onto.

4. Let F be a field, V, W be finite-dimensional F-vector spaces, and Z be any F-vector space. Choose a basis $\{v_1, \ldots, v_n\}$ of V and a basis $\{w_1, \ldots, w_m\}$ of W. Let

$$\{z_{i,j} : 1 \leq i \leq n, 1 \leq j \leq m\}$$

be any set of mn vectors from Z. Show that there is precisely one bilinear transformation $f : V \times W \to Z$ such that

$$f(v_i, w_j) = z_{i,j} \text{ for all } i, j \ .$$

5. Let V be a vector space and $T : V \to V$ a linear transformation. Show that the following two statements are equivalent.

   (A) $V = R(T) \oplus N(T)$, where $R(T)$ is the range of T and $N(T)$ is the nullspace of T.

   (B) $N(T) = N(T^2)$, where T^2 is T composed with itself.

6. Let V, W and Z be finite-dimensional vector spaces over a field F. If $T : V \to W$ and $U : W \to Z$ are linear transformations, prove that

$$rank(UT) \leq \min\{rank(U), rank(T)\} \ .$$

   Prove also that if either of U or T is invertible, then the rank of UT is equal to the rank of the other one. Deduce that if $P : V \to V$ and $Q : W \to W$ are isomorphisms then the rank of QTP equals the rank of T.

7. Let V and W be finite-dimensional vector spaces over a field F and $T : V \to W$ be a linear transformation. Show that there exist ordered bases β of V and γ of W such that

$$\left([T]_\gamma^\beta\right)_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 0 \text{ or } 1 & \text{if } i = j \end{cases} \ .$$

8. The purpose of this question is to show that the row rank of a matrix is equal to its column rank. Note that this is obviously true for a matrix of the form described in the previous exercise. Our goal will be to put an arbitrary matrix in this form without changing either its row rank or its column rank. Let A be an $m \times n$ matrix with entries from a field F.

   (a) Show that the column rank of A is equal to the rank of the linear transformation $L_A : F^n \to F^m$ defined by $L_A(\vec{v}) = A \cdot \vec{v}$, viewing $\vec{v}$ as a column vector.

   (b) Use question 1 to show that if P and Q are invertible $n \times n$ and $m \times m$ matrices respectively then the column rank of QAP equals the column rank of A.

   (c) Show that the row rank of A is equal to the rank of the linear transformation $R_A : F^m \to F^n$ defined by $R_A(\vec{v}) = \vec{v} \cdot A$, viewing $\vec{v}$ as a row vector.

   (d) Use question 1 to show that if P and Q are invertible $n \times n$ and $m \times m$ matrices respectively then the row rank of QAP equals the row rank of A.

(e) Show that there exist $n \times n$ and $m \times m$ matrices P and Q respectively such that QAP has the form described in question 2. Deduce that the row rank of A equals the column rank of A.

9. Given an angle $\theta \in [0, 2\pi)$ let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the function which rotates a vector clockwise about the origin by an angle θ. Find the matrix of T_θ relative to the standard basis. You do not need to prove that T_θ is linear.

10. Given $m \in \mathbb{R}$, define the line

$$L_m = \{(x, y) \in \mathbb{R}^2 : y = mx\} .$$

(a) Let T_m be the function which maps a point in $\mathbb{R}^2$ to its closest point in L_m. Find the matrix of T_m relative to the standard basis. You do not need to prove that T_m is linear.

(b) Let R_m be the function which maps a point in $\mathbb{R}^2$ to the reflection of this point about the line L_m. Find the matrix of T_m relative to the standard basis. You do not need to prove that R_m is linear.

Hint for (a) and (b): first find the matrix relative to a carefully chosen basis and then perform a change of basis.

11. **The quotient space V/W.** Let V be an $\mathbb{F}$-vector space, and $W \subset V$ a subspace. A subset $S \subset V$ is called a *W-affine subspace of V*, if the following holds:

$$\forall s, s' \in S : s - s' \in W \qquad \text{and} \qquad \forall s \in S, w \in W : s + w \in S.$$

(a) Let S and T be W-affine subspaces of V and $c \in \mathbb{F}$. Put

$$S + T := \{s + t : s \in S, t \in T\}, \qquad cT := \begin{cases} \{ct : t \in T\} & c \neq 0 \\ W & c = 0 \end{cases} .$$

Show that $S + T$ and cT are again W-affine subspaces of V.

(b) Show that the above operations define an $\mathbb{F}$-vector space structure on the set of all W-affine subspaces of V.

We will write V/W for the set of W-affine subspaces of V. We now know that it is a vector space. Note that the *elements* of V/W are *subsets* of V.

(c) Show that if $v \in V$, then

$$v + W := \{v + w : w \in W\}$$

is a W-affine subspace of V. Show moreover that for any W-affine subspace $S \subset V$ there exists a $v \in V$ such that $S = v + W$.

(d) Show that the map $p : V \to V/W$ defined by $p(v) = v + W$ is linear and surjective.

(e) Compute the nullspace of p and, if the dimension of V is finite, the dimension of V/W.

A helpful way to think about the quotient space V/W is to think of it as "being" the vector space V, but with a new notion of "equality" of vectors. Namely, two vectors $v_1, v_2 \in V$ are now seen as equal if $v_1 - v_2 \in W$. Use this point of view to find a solution for the following exercise. When you find it, use the formal definition given above to write your solution rigorously.

12. Let V and X be $\mathbb{F}$-vector spaces, and $f \in L(V, X)$. Let W be a subspace of V contained in $N(f)$. Consider the quotient space V/W and the map $p : V \to V/W$ from the previous exercise.

    (a) Show that there exists a unique $\tilde{f} \in L(V/W, X)$ such that $f = \tilde{f} \circ p$.

    (b) Show that $\tilde{f}$ is injective if and only if $W = N(f)$.

3 Dual spaces

3.1 Definitions

Consider the space F^n and write each vector as a column vector. When can we say that a vector is zero? When all coordinates are zero. Further, we say that two vectors are the same if all of their coordinates are the same. This motivates the definition of the coordinate maps

$$e_i^* : F^n \to F \text{ by } e_i^*(\vec{v}) = i\text{-th coordinate of } \vec{v} \ .$$

Notice that each e_i^* is a linear function from F^n to F. Furthermore they are linearly independent, so since the dimension of $\mathcal{L}(F^n, F)$ is n, they form a basis.

Last it is clear that a vector $\vec{v}$ is zero if and only if $e_i^*(\vec{v}) = 0$ for all i. This is true if and only if $f(\vec{v}) = 0$ for all f which are linear functions from $F^n \to F$. This motivates the following definition.

Definition 3.1.1. *If V is a vector space over F we define the dual space V^* as the space of linear functionals*

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

We can view this as the space $\mathcal{L}(V, F)$, where F is considered as a one-dimensional vector space over itself.

Suppose that V is finite dimensional and $f \in V^*$. Then by the rank-nullity theorem, either $f \equiv 0$ or $N(f)$ is a $dim\ V - 1$ dimensional subspace of V. Conversely, you will show in the homework that any $dim\ V - 1$ dimensional subspace W (that is, a *hyperspace*) is the nullspace of some linear functional.

Definition 3.1.2. *If $\beta = \{v_1, \ldots, v_n\}$ is a basis for V then we define the dual basis $\beta^* = \{f_{v_1}, \ldots, f_{v_n}\}$ as the unique functionals satisfying*

$$f_{v_i}(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \ .$$

From our proof of the dimension of $\mathcal{L}(V, W)$ we know that β^* is a basis of V^*.

Proposition 3.1.3. *Given a basis $\beta = \{v_1, \ldots, v_n\}$ of V and dual basis β^* of V^* we can write*

$$f = f(v_1)f_{v_1} + \cdots + f(v_n)f_{v_n} \ .$$

In other words, the coefficients for f in the dual basis are just $f(v_1), \ldots, f(v_n)$.

Proof. Given $f \in V^*$, we can write

$$f = a_1 f_{v_1} + \cdots + a_n f_{v_n} \ .$$

To find the coefficients, we evaluate both sides at v_k. The left is just $f(v_k)$. The right is

$$a_k f_{v_k}(v_k) = a_k \ .$$

Therefore $a_k = f(v_k)$ are we are done. □

3.2 Annihilators

We now study annihilators.

Definition 3.2.1. *If $S \subseteq V$ we define the annihilator of S as*

$$S^\perp = \{f \in V^* : f(v) = 0 \text{ for all } v \in S\} \ .$$

Theorem 3.2.2. *Let V be a vector space and $S \subseteq V$.*

1. *$S^\perp$ is a subspace of V^* (although S does not have to be a subspace of V).*
2. *$S^\perp = (Span\ S)^\perp$.*
3. *If $\dim V < \infty$ and U is a subspace of V then whenever $\{v_1, \ldots, v_k\}$ is a basis for U and $\{v_1, \ldots, v_n\}$ is a basis for V,*

$$\{f_{v_{k+1}}, \ldots, f_{v_n}\} \text{ is a basis for } U^\perp \ .$$

Proof. First we show that $S^\perp$ is a subspace of V^*. Note that the zero functional obviously sends every vector in S to zero, so $0 \in S^\perp$. If $c \in F$ and $f_1, f_2 \in S^\perp$, then for each $v \in S$,

$$(cf_1 + f_2)(v) = cf_1(v) + f_2(v) = 0 \ .$$

So $cf_1 + f_2 \in S^\perp$ and $S^\perp$ is a subspace of V^*.

Next we show that $S^\perp = (Span\ S)^\perp$. To prove the forward inclusion, take $f \in S^\perp$. Then if $v \in Span\ S$ we can write

$$v = a_1 v_1 + \cdots + a_m v_m$$

for scalars $a_i \in F$ and $v_i \in S$. Thus

$$f(v) = a_1 f(v_1) + \cdots + a_m f(v_m) = 0 \ ,$$

so $f \in (Span\ S)^\perp$. On the other hand if $f \in (Span\ S)^\perp$ then clearly $f(v) = 0$ for all $v \in S$ (since $S \subseteq Span\ S$). This completes the proof of item 2.

For the third item, we know that the functionals $f_{v_{k+1}}, \ldots, f_{v_n}$ are linearly independent. Therefore we just need to show that they span $U^\perp$. To do this, take $f \in U^\perp$. We can write f in terms of the dual basis $f_{v_1}, \ldots, f_{v_n}$:

$$f = a_1 f_{v_1} + \cdots + a_k f_{v_k} + a_{k+1} f_{v_{k+1}} + \cdots + a_n f_{v_n} \ .$$

Using the formula we have for the coefficients, we get $a_j = f(v_j)$, which is zero for $j \leq k$. Therefore

$$f = a_{k+1} f_{v_{k+1}} + \cdots + a_n f_{v_n}$$

and we are done.

□

Corollary 3.2.3. *If V is finite dimensional and W is a subspace then*

$$dim\ V = dim\ W + dim\ W^{\perp}\ .$$

Definition 3.2.4. *For $S' \subseteq V^*$ we define*

$$^{\perp}(S') = \{v \in V : f(v) = 0 \text{ for all } f \in S'\}\ .$$

In the homework you will prove similar properties for $^{\perp}(S')$.

Fact: $v \in V$ is zero if and only if $f(v) = 0$ for all $f \in V^*$. One implication is easy. To prove the other, suppose that $v \neq \vec{0}$ and and extend $\{v\}$ to a basis for V. Then the dual basis has the property that $f_v(v) \neq 0$.

Proposition 3.2.5. *If $W \subseteq V$ is a subspace and V is finite-dimensional then $^{\perp}(W^{\perp}) = W$.*

Proof. If $w \in W$ then for all $f \in W^{\perp}$, we have $f(w) = 0$, so $w \in^{\perp} (W^{\perp})$. Suppose conversely that $w \in V$ has $f(w) = 0$ for all $f \in W^{\perp}$. If $w \notin W$ then build a basis $\{v_1, \ldots, v_n\}$ of V such that $\{v_1, \ldots, v_k\}$ is a basis for W and $v_{k+1} = w$. Then by the previous proposition, $\{f_{v_{k+1}}, \ldots, f_{v_n}\}$ is a basis for $W^{\perp}$. However $f_w(w) = 1 \neq 0$, which is a contradiction, since $f_w \in W^{\perp}$. □

3.3 Double dual

Lemma 3.3.1. *If $v \in V$ is nonzero and dim $V < \infty$, there exists a linear functional $f_v \in V^*$ such that $f_v(v) = 1$. Therefore $v = \vec{0}$ if and only if $f(v) = 0$ for all $f \in V^*$.*

Proof. Extend $\{v\}$ to a basis of V and consider the dual basis. f_v is in this basis and $f_v(v) = 1$. □

For each $v \in V$ we can define the evaluation map $\widetilde{v} : V^* \to F$ by

$$\widetilde{v}(f) = f(v)\ .$$

Theorem 3.3.2. *Suppose that V is finite-dimensional. Then the map $\Phi : V \to V^{**}$ given by*

$$\Phi(v) = \widetilde{v}$$

is an isomorphism.

Proof. First we show that if $v \in V$ then $\Phi(v) \in V^{**}$. Clearly $\widetilde{v}$ maps V^* to F, but we just need to show that $\widetilde{v}$ is linear. If $f_1, f_2 \in V^*$ and $c \in F$ then

$$\widetilde{v}(cf_1 + f_2) = (cf_1 + f_2)(v) = cf_1(v) + f_2(v) = c\widetilde{v}(f_1) + \widetilde{v}(f_2)\ .$$

Therefore $\Phi(v) \in V^{**}$. We now must show that Φ is linear and either one-to-one or onto (since the dimension of V is equal to the dimension of V^{**}). First if $v_1, v_2 \in V$, $c \in F$ then we want to show that

$$\Phi(cv_1 + v_2) = c\Phi(v_1) + \Phi(v_2)\ .$$

Both sides are elements of V^{**} so we need to show they act the same on elements of V^*. Let $f \in V^*$. Then

$$\Phi(cv_1+v_2)(f) = f(cv_1+v_2) = cf(v_1)+f(v_2) = c\Phi(v_1)(f)+\Phi(v_2)(f) = (c\Phi(v_1)+\Phi(v_2))(f) \ .$$

Finally to show one-to-one, we show that $N(\Phi) = \{0\}$. If $\Phi(v) = \vec{0}$ then for all $f \in V^*$,

$$0 = \Phi(v)(f) = f(v) \ .$$

This implies $v = \vec{0}$. □

Theorem 3.3.3. *Let V be finite dimensional.*

1. *If $\beta = \{v_1, \ldots, v_n\}$ is a basis for V then $\Phi(\beta) = \{\Phi(v_1), \ldots, \Phi(v_n)\}$ is the double dual of this basis.*
2. *If W is a subspace of V then $\Phi(W)$ is equal to $(W^\perp)^\perp$.*

Proof. Recall that the dual basis of β is $\beta^* = \{f_{v_1}, \ldots, f_{v_n}\}$, where

$$f_{v_i}(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \ .$$

Since Φ is an isomorphism, $\Phi(\beta)$ is a basis of V^{**}. Now

$$\Phi(v_i)(f_{v_k}) = f_{v_k}(v_i)$$

which is 1 if $i = k$ and 0 otherwise. This means $\Phi(\beta) = \beta^{**}$.

Next if W is a subspace, let $w \in W$. Letting $f \in W^\perp$,

$$\Phi(w)(f) = f(w) = 0 \ .$$

So $w \in (W^\perp)^\perp$. However, since Φ is an isomorphism, $\Phi(W)$ is a subspace of $(W^\perp)^\perp$. But they have the same dimension, so they are equal.

□

3.4 Dual maps

Definition 3.4.1. *Let $T : V \to W$ be linear. We define the dual map $T^* : W^* \to V^*$ by*

$$T^*(g)(v) = g(T(v)) \ .$$

Theorem 3.4.2. *Let V and W be finite dimensional and let β and γ be bases for V and W. If $T : V \to W$ is linear, so is T^*. If β^* and γ^* are the dual bases, then*

$$[T^*]_{\beta^*}^{\gamma^*} = \left([T]_\gamma^\beta\right)^t \ .$$

Proof. First we show that T^* is linear. If $g_1, g_2 \in W^*$ and $c \in F$ then for each $v \in V$,

$$\begin{aligned} T^*(cg_1 + g_2)(v) &= (cg_1 + g_2)(T(v)) = cg_1(T(v)) + g_2(T(v)) \\ &= cT^*(g_1)(v) + T^*(g_2)(v) = (cT^*(g_1) + T^*(g_2))(v) \ . \end{aligned}$$

Next let $\beta = \{v_1, \ldots, v_n\}$, $\gamma = \{w_1, \ldots, w_m\}$, $\beta^* = \{f_{v_1}, \ldots, f_{v_n}\}$ and $\gamma^* = \{g_{w_1}, \ldots, g_{w_m}\}$ and write $[T]_\gamma^\beta = (a_{i,j})$. Then recall from the lemma that for any $f \in V^*$ we have

$$f = f(v_1)f_{v_1} + \cdots + f(v_n)f_{v_n} \ ,$$

Therefore the coefficient of f_{v_i} for $T^*(g_{w_k})$ is

$$T^*(g_{w_k})(v_i) = g_{w_k}(T(v_i)) = g_{w_k}(a_{1,i}w_1 + \cdots + a_{m,i}w_m) = a_{k,i} \ .$$

So

$$T^*(g_{w_k}) = a_{k,1}f_{v_1} + \cdots + a_{k,n}f_{v_n} \ .$$

This is the k-th column of $[T^*]_{\beta^*}^{\gamma^*}$.

□

Theorem 3.4.3. *If V and W are finite dimensional and $T : V \to W$ is linear then $R(T^*) = (N(T))^\perp$ and $(R(T))^\perp = N(T^*)$.*

Proof. If $g \in N(T^*)$ then $T^*(g)(v) = 0$ for all $v \in V$. If $w \in R(T)$ then $w = T(v)$ for some $v \in V$. Then $g(w) = g(T(v)) = T^*(g)(v) = 0$. Thus $g \in (R(T))^\perp$. If, conversely, $g \in (R(T))^\perp$ then we would like to show that $T^*(g)(v) = 0$ for all $v \in V$. We have

$$T^*(g)(v) = g(T(v)) = 0 \ .$$

Let $f \in R(T^*)$. Then $f = T^*(g)$ for some $g \in W^*$. If $T(v) = 0$, we have $f(v) = T^*(g)(v) = g(T(v)) = 0$. Therefore $f \in (N(T))^\perp$. This gives $R(T^*) \subseteq (N(T))^\perp$. To show the other direction,

$$dim\ R(T^*) = m - dim\ N(T^*)$$

and

$$dim\ (N(T))^\perp = n - dim\ N(T) \ .$$

However, by part 1, $dim\ N(T^*) = m - dim\ R(T)$, so

$$dim\ R(T^*) = dim\ R(T) = n - dim\ N(T) = dim\ (N(T))^\perp \ .$$

This gives the other inclusion.

□

3.5 Exercises

Notation:

1. Recall the definition of a bilinear function. Let F be a field, and V, W and Z be F-vector spaces. A function $f : V \times W \to Z$ is called *bilinear* if

   (a) for each $v \in V$ the function $f_v : W \to Z$ given by $f_v(w) = f(v, w)$ is linear as a function of w and

   (b) for each $w \in W$ the function $f_w : V \to Z$ given by $f_w(v) = f(v, w)$ is linear as a function of v.

   When Z is the F-vector space F, one calls f a *bilinear form*.

2. Given a bilinear function $f : V \times W \to Z$, we define its *left kernel* and its *right kernel* as

$$LN(f) = \{v \in V : f(v, w) = 0 \ \forall w \in W\},$$
$$RN(f) = \{w \in W : f(v, w) = 0 \ \forall v \in V\}.$$

   More generally, for subspaces $U \subset V$ and $X \subset W$ we define their orthogonal complements

$$U^{\perp_f} = \{w \in W : f(u, w) = 0 \ \forall u \in U\},$$
$${}^{\perp_f}X = \{v \in V : f(v, x) = 0 \ \forall x \in X\}.$$

   Notice that $LN(f) = {}^{\perp_f}W$ and $RN(f) = V^{\perp_f}$.

Exercises:

1. Let V and W be vector spaces over a field F and let $f : V \times W \to F$ be a bilinear form. For each $v \in V$, we denote by f_v the linear functional $W \to F$ given by $f_v(w) = f(v, w)$. For each $w \in W$, we denote by f_w the linear functional $V \to F$ given by $f_w(v) = f(v, w)$.

   (a) Show that the map
$$\Phi : V \to W^*, \qquad \Phi(v) = f_v$$
   is linear and its kernel is $LN(f)$.

   (b) Analogously, show that the map
$$\Psi : W \to V^*, \qquad \Psi(w) = f_w$$
   is linear and its kernel is $RN(f)$.

   (c) Assume now that V and W are finite-dimensional. Show that the map $W^{**} \to V^*$ given by composing Ψ with the inverse of the canonical isomorphism $W \to W^{**}$ is equal to Φ^*, the map dual to Φ.

(d) Assuming further $\dim(V) = \dim(W)$, conclude that the following statements are equivalent:

i. $LN(f) = \{0\}$,

ii. $RN(f) = \{0\}$,

iii. Φ is an isomorphism,

iv. Ψ is an isomorphism.

2. Let V, W be finite-dimensional F-vector spaces. Denote by ξ_V and ξ_W the canonical isomorphisms $V \to V^{**}$ and $W \to W^{**}$. Show that if $T : V \to W$ is linear then $\xi_W^{-1} \circ T^{**} \circ \xi_V = T$.

3. Let V be a finite-dimensional F-vector space. Show that any basis $(\lambda_1, \ldots, \lambda_n)$ of V^* is the dual basis to some basis $(v_1, \ldots, v_n)$ of V.

4. Let V be an F-vector space, and $S' \subset V^*$ a subset. Recall the definition

$$^{\perp}S' = \{v \in V : \lambda(v) = 0 \ \forall \lambda \in S'\}.$$

Imitating a proof given in class, show the following:

(a) $^{\perp}S' = {}^{\perp}\mathrm{span}(S')$.

(b) Assume that V is finite-dimensional, let $U' \subset V^*$ be a subspace, and let $(\lambda_1, \ldots, \lambda_n)$ be a basis for V^* such that $(\lambda_1, \ldots, \lambda_k)$ is a basis for U'. If $(v_1, \ldots, v_n)$ is the basis of V from the previous exercise, then $(v_{k+1}, \ldots, v_n)$ is a basis for $^{\perp}U'$. In particular, $\dim(U') + \dim(^{\perp}U') = \dim(V^*)$.

5. Let V be a finite-dimensional F-vector space, and $U \subset V$ a hyperplane (that is, a subspace of V of dimension $dim\ V - 1$).

(a) Show that there exists $\lambda \in V^*$ with $N(\lambda) = U$.

(b) Show that if $\mu \in V^*$ is another functional with $N(\mu) = U$, then there exists $c \in F^{\times}$ with $\mu = c\lambda$.

6. (From Hoffman-Kunze)

(a) Let A and B be $n \times n$ matrices with entries from a field F. Show that $Tr\ (AB) = Tr\ (BA)$.

(b) Let $T : V \to V$ be a linear transformation on a finite-dimensional vector space. Define the trace of T as the trace of the matrix of T, represented in some basis. Prove that the definition of trace does not depend on the basis thus chosen.

(c) Prove that on the space of $n \times n$ matrices with entries from a field F, the trace function Tr is a linear functional. Show also that, conversely, if some linear functional g on this space satisfies $g(AB) = g(BA)$ then g is a scalar multiple of the trace function.

4 Determinants

4.1 Permutations

Now we move to permutations. These will be used when we talk about the determinant.

Definition 4.1.1. *A permutation on n letters is a function $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ which is a bijection.*

The set of all permutations forms a group under composition. There are $n!$ elements. There are two main ways to write a permutation.

1. **Row notation:**
$$\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 2 & 3 & 5 & 1 \end{matrix}$$
Here we write the elements of $\{1, \ldots, n\}$ in the first row, in order. In the second row we write the elements they are mapped to, in order.

2. **Cycle decomposition:**
$$(1\ 6)(2\ 4\ 3)(5)$$
All cycles are disjoint. It is easier to compose permutations this way. Suppose π is the permutation given above and π' is the permutation
$$\pi' = (1\ 2\ 3\ 4\ 5)(6)\ .$$
Then the product of $\pi\pi'$ is (here we will apply π' first)
$$\big[(1\ 6)(2\ 3\ 4)(5)\big]\big[(1\ 2\ 3\ 4\ 5)(6)\big] = (1\ 3\ 2\ 4\ 5\ 6)\ .$$
It is a simple fact that each permutation has a cycle decomposition with disjoint cycles.

Definition 4.1.2. *A transposition is a permutation that swaps two letters and fixes the others. Removing the fixed letters, it looks like $(i\ j)$ for $i \neq j$. An adjacent transposition is one that swaps neighboring letters.*

Lemma 4.1.3. *Every permutation π can be written as a product of transpositions. (Not necessarily disjoint.)*

Proof. All we need to do is write a cycle as a product of transpositions. Note
$$(a_1\ a_2\ \cdots\ a_k) = (a_1\ a_k)(a_1\ a_{k-1}) \cdots (a_1\ a_3)(a_1\ a_2)\ .$$
□

Definition 4.1.4. *A pair of numbers (i, j) is an inversion pair for π if $i < j$ but $\pi(i) > \pi(j)$. Write $N_{inv}(\pi)$ for the number of inversion pairs of π.*

For example in the permutation (13)(245), also written as

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 1 & 5 & 2 \end{matrix},$$

we have inversion pairs $(1,3),(1,5),(2,3),(2,5),(4,5)$.

Lemma 4.1.5. *Let π be a permutation and $\sigma = (k\ k+1)$ be an adjacent transposition. Then $N_{inv}(\sigma\pi) = N_{inv}(\pi) \pm 1$. If $\sigma_1, \ldots, \sigma_m$ are adjacent transpositions then*

$$N_{inv}(\pi\sigma_1 \cdots \sigma_m) - N_{inv}(\pi) = \begin{cases} even & m\ even \\ odd & m\ odd \end{cases}.$$

Proof. Let $a < b \in \{1, \ldots, n\}$. If $\pi(a), \pi(b) \in \{k, k+1\}$ then $\sigma\pi(a) - \sigma\pi(b) = -(\pi(a) - \pi(b))$ so (a,b) is an inversion pair for π if and only if it is not one for $\sigma\pi$. We claim that in all other cases, the sign of $\sigma\pi(a) - \sigma\pi(b)$ is the same as the sign of $\pi(a) - \pi(b)$. If neither of $\pi(a)$ and $\pi(b)$ is in $\{k, k+1\}$ then $\sigma\pi(a) - \sigma\pi(b) = \pi(a) - \pi(b)$. The other cases are somewhat similar: if $\pi(a) = k$ but $\pi(b) > k+1$ then $\sigma\pi(a) - \sigma\pi(b) = k + 1 - \pi(b) < 0$ and $\pi(a) - \pi(b) = k - \pi(b) < 0$. Keep going.

Therefore $\sigma\pi$ has exactly the same inversion pairs as π except for $(\pi^{-1}(a), \pi^{-1}(b))$, which switches status. This proves the lemma. □

Definition 4.1.6. *Given a permutation π on n letters, we say that π is even if it can be written as a product of an even number of transpositions and odd otherwise. This is called the signature (or sign) of a permutation:*

$$sgn(\pi) = \begin{cases} +1 & \text{if } \pi \text{ is even} \\ -1 & \text{if } \pi \text{ is odd} \end{cases}.$$

Theorem 4.1.7. *If π can be written as a product of an even number of transpositions, it cannot be written as a product of an odd number of transpositions. In other words, signature is well-defined.*

Proof. Suppose that $\pi = s_1 \cdots s_k$ and $\pi = t_1 \cdots t_j$. We want to show that $k - j$ is even. In other words, if k is odd, so is j and if k is even, so is j.

Now note that each transposition can be written as a product of an odd number of adjacent transpositions:

$$(5\ 1) = (5\ 4)(4\ 3)(3\ 2)(1\ 2)(2\ 3)(3\ 4)(4\ 5)$$

so write $s_1 \cdots s_k = \widetilde{s}_1 \cdots \widetilde{s}_{k'}$ and $t_1 \cdots t_j = \widetilde{t}_1 \cdots \widetilde{t}_{j'}$ where $k - k'$ is even and $j - j'$ is even.

We have $0 = N_{inv}(id) = N_{inv}(\widetilde{t}_{j'}^{-1} \cdots \widetilde{t}_1^{-1}\pi)$, which is $N_{inv}(\pi)$ plus an even number if j' is even or and odd number if j' is odd. This means that $N_{inv}(\pi) - j'$ is even. The same argument works for k, so $N_{inv}(\pi) - k'$ is even. Now

$$j - k = j - j' + j' - N_{inv}(\pi) + N_{inv}(\pi) - k' + k' - k$$

is even.

□

Corollary 4.1.8. *For any two permutations π and π',*

$$sgn(\pi\pi') = sgn(\pi)sgn(\pi') \ .$$

4.2 Determinants: existence and uniqueness

Given n vectors $\vec{v}_1, \ldots, \vec{v}_n$ in $\mathbb{R}^n$ we want to define something like the volume of the parallelepiped spanned by these vectors. What properties would we expect of a volume?

1. $vol(e_1, \ldots, e_n) = 1$.
2. If two of the vectors $\vec{v}_i$ are equal the volume should be zero.
3. For each $c > 0$, $vol(c\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n) = c\ vol(\vec{v}_1, \ldots, \vec{v}_n)$. Same in other arguments.
4. For each $\vec{v}_1'$, $vol(\vec{v}_1 + \vec{v}_1', \vec{v}_2, \ldots, \vec{v}_n) = vol(\vec{v}_1, \ldots, \vec{v}_n) + vol(\vec{v}_1', \vec{v}_2, \ldots, \vec{v}_n)$. Same in other arguments.

Using the motivating example of the volume, we define a multilinear function as follows.

Definition 4.2.1. *If V is an n-dimensional vector space over F then define*

$$V^n = \{(v_1, \ldots, v_n) : v_i \in V \text{ for all } i = 1, \ldots, n\} \ .$$

A function $f : V^n \to F$ is called multilinear if for each i and vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n \in V$, the function $f_i : V \to F$ is linear, where

$$f_i(v) = f(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n) \ .$$

A multilinear function f is called alternating if $f(v_1, \ldots, v_n) = 0$ whenever $v_i = v_j$ for some $i \neq j$.

Proposition 4.2.2. *Let $f : V^n \to F$ be a multilinear function. If F does not have characteristic two then f is alternating if and only if for all $v_1, \ldots, v_n$ and $i < j$,*

$$f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) \ .$$

Proof. Suppose that f is alternating. Then

$$\begin{aligned} 0 &= f(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) \\ &= f(v_1, \ldots, v_i, \ldots, v_i + v_j, \ldots, v_n) + f(v_1, \ldots, v_j, \ldots, v_i + v_j, \ldots, v_n) \\ &= f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) \ . \end{aligned}$$

Conversely suppose that f has the property above. Then if $v_i = v_j$,

$$\begin{aligned} f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) &= -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) \\ &= -f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) \ . \end{aligned}$$

Since F does not have characteristic two, this means this is zero. □

Corollary 4.2.3. *Let $f : V^n \to F$ be an n-linear alternating function. Then for each $\pi \in S_n$,*

$$f(v_{\pi(1)}, \ldots, v_{\pi(n)}) = sgn(\pi)\ f(v_1, \ldots, v_n)\ .$$

Proof. Write $\pi = \sigma_1 \cdots \sigma_k$ where the σ_i's are transpositions and $(-1)^k = sgn(\pi)$. Then

$$f(v_{\pi(1)}, \ldots, v_{\pi(n)}) = -f(v_{\sigma_1 \cdots \sigma_{k-1}(1)}, \ldots, v_{\sigma_1 \cdots \sigma_{k-1}(n)})\ .$$

Applying this $k-1$ more times gives the corollary. □

Theorem 4.2.4. *Let $\{v_1, \ldots, v_n\}$ be a basis for V. There is at most one multilinear alternating function $f : V^n \to F$ such that $f(v_1, \ldots, v_n) = 1$.*

Proof. Let $u_1, \ldots, u_n \in V$ and write

$$u_k = a_{1,k}v_1 + \cdots + a_{n,k}v_n\ .$$

Then

$$\begin{aligned} f(u_1, \ldots, u_n) &= \sum_{i_1=1}^{n} a_{i_1,1} f(v_{i_1}, u_2, \ldots, u_n) \\ &= \sum_{i_1, \ldots, i_n = 1}^{n} a_{i_1,1} \cdots a_{i_n,n} f(v_{i_1}, \ldots, v_{i_n})\ . \end{aligned}$$

However whenever two i_j's are equal, we get zero, so we can restrict the sum to all distinct i_j's. So this is

$$\sum_{i_1, \ldots, i_n = 1\ distinct}^{n} a_{i_1,1} \cdots a_{i_n,n} f(v_{i_1}, \ldots, v_{i_n})\ .$$

This can now be written as

$$\begin{aligned} & \sum_{\pi \in S_n} a_{\pi(1),1} \cdots a_{\pi(n),n} f(v_{\pi(1)}, \ldots, v_{\pi(n)}) \\ = & \sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n} f(v_1, \ldots, v_n) \\ = & \sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n}\ . \end{aligned}$$

□

We now find that if $dim\ V = n$ and $f : V^n \to F$ is an n-linear alternating function with $f(v_1, \ldots, v_n) = 1$ for some fixed basis $\{v_1, \ldots, v_n\}$ then we have a specific form for f. Writing vectors $u_1, \ldots, u_n$ as

$$u_k = a_{1,k}v_1 + \cdots + a_{n,k}v_n\ ,$$

then we have

$$f(u_1, \ldots, u_n) = \sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n}\ .$$

Now we would like to show that the formula above indeed does define an n-linear alternating function with the required property.

1. **Alternating.** Suppose that $u_i = u_j$ for some $i < j$. We will then split the set of permutations into two classes. Let $A = \{\pi \in S_n : \pi(i) < \pi(j)\}$. Letting $\sigma_{i,j} = (ij)$, write $\hat{\pi}$ for $\pi \in A$ for the permutation $\pi\sigma_{i,j}$.

$$\begin{aligned} f(u_1, \ldots, u_n) &= \sum_{\pi \in A} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n} + \sum_{\tau \in S_n \setminus A} sgn(\tau)\ a_{\tau(1),1} \cdots a_{\tau(n),n} \\ &= \sum_{\pi \in A} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n} + \sum_{pi \in A} sgn(\pi\sigma_{i,j})\ a_{\pi\sigma_{i,j}(1),1} \cdots a_{\pi\sigma_{i,j}(n),n}\ . \end{aligned}$$

   However, $\pi\sigma_{i,j}(k) = \pi(k)$ when $k \neq i, j$ and $u_i = u_j$ so this equals

$$\sum_{\pi \in A} [sgn(\pi) + sgn(\pi\sigma_{i,j})] a_{\pi(1),1} \cdots a_{\pi(n),n} = 0\ .$$

2. **1 at the basis.** Note that for $u_i = v_i$ for all i we have

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

   Therefore the value of f is

$$\sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n} = sgn(id) a_{1,1} \cdots a_{n,n} = 1\ .$$

3. **Multilinear.** Write

$$u_k = a_{1,k} v_1 + \cdots + a_{n,k} v_n \text{ for } k = 1, \ldots, n$$

   and write $u = b_1 v_1 + \cdots + b_n v_n$. Now for $c \in F$,

$$cu + u_1 = (cb_1 + a_{1,1}) v_1 + \cdots + (cb_n + a_{n,1}) v_n\ .$$

$$\begin{aligned} & f(cu + u_1, u_2, \ldots, u_n) \\ = & \sum_{\pi \in S_n} sgn(\pi)\ [cb_{\pi(1)} + a_{\pi(1),1}] a_{\pi(2),2} \cdots a_{\pi(n),n} \\ = & c \sum_{\pi \in S_n} sgn(\pi)\ b_{\pi(1)} a_{\pi(2),2} \cdots a_{\pi(n),n} + \sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} a_{\pi(2),2} \cdots a_{\pi(n),n} \\ = & cf(u, u_2, \ldots, u_n) + f(u_1, \ldots, u_n)\ . \end{aligned}$$

4.3 Properties of determinants

Theorem 4.3.1. *Let $f : V^n \to F$ be a multilinear alternating function and let $\{v_1, \ldots, v_n\}$ be a basis with $f(v_1, \ldots, v_n) \neq 0$. Then $\{u_1, \ldots, u_n\}$ is linearly dependent if and only if $f(u_1, \ldots, u_n) = 0$.*

Proof. One direction is on the homework: suppose that $f(u_1, \ldots, u_n) = 0$ but that $\{u_1, \ldots, u_n\}$ is linearly independent. Then write

$$v_k = a_{1,k}u_1 + \cdots + a_{n,k}u_n \ .$$

By the same computation as above,

$$f(v_1, \ldots, v_n) = f(u_1, \ldots, u_n) \sum_{\pi \in S_n} sgn(\pi)\ a_{\pi(1),1} \cdots a_{\pi(n),n} = 0 \ ,$$

which is a contradiction. Therefore $\{u_1, \ldots, u_n\}$ is linearly dependent.

Conversely, if $\{u_1, \ldots, u_n\}$ is linearly dependent then for $n \geq 2$ we can find some u_j and scalars a_i for $i \neq j$ such that

$$u_j = \sum_{i \neq j} a_i u_i \ .$$

Now we have

$$f(u_1, \ldots, u_j, \ldots, u_n) = \sum_{i \neq j} a_i f(u_1, \ldots, u_i, \ldots, u_n) = 0 \ .$$

□

Definition 4.3.2. *On the space F^n we define $det : F^n \to F$ as the unique alternating n-linear function that gives $det(e_1, \ldots, e_n) = 1$. If A is an $n \times n$ matrix then we define*

$$det\ A = det(\vec{a}_1, \ldots, \vec{a}_n) \ ,$$

where $\vec{a}_k$ is the k-th column of A.

Corollary 4.3.3. *An $n \times n$ matrix A over F is invertible if and only if $det\ A \neq 0$.*

Proof. We have $det\ A \neq 0$ if and only if the columns of A are linearly independent. This is true if and only if A is invertible. □

We start with the multiplicative property of determinants.

Theorem 4.3.4. *Let A and B be $n \times n$ matrices over a field F. Then*

$$det\ (AB) = det\ A \cdot det\ B \ .$$

Proof. If $det\ B = 0$ then B is not invertible, so it cannot have full column rank. Therefore neither can AB (by a homework problem). This means $det\ (AB) = 0$ and we are done.

Otherwise $det\ B \neq 0$. Define a function $f : M_{n\times n}(F) \to F$ by

$$f(A) = \frac{det\ (AB)}{det\ B}\ .$$

We claim that f is n-linear, alternating and assigns the value 1 to the standard basis (that is, the identity matrix).

1. f **is alternating.** If A has two equal columns then its column rank is not full. Therefore neither can be the column rank of AB and we have $det\ (AB) = 0$. This implies $f(A) = 0$.

2. $f(I) = 1$**.** This is clear since $IB = B$.

3. f **is** n**-linear.** This follows because *det* is.

But there is exactly one function satisfying the above. We find $f(A) = det\ A$ and we are done. □

For the rest of the lecture we will give further properties of determinants.

- $\det A = \det A^t$. This is on homework.

- det is alternating and n-linear as a function of rows.

- If A is (a block matrix) of the form

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

  then $\det A = \det B \det D$. This is also on homework.

- The determinant is unchanged if we add a multiple of one column (or row) to another. To show this, write a matrix A as a collection of columns $(\vec{a}_1, \ldots, \vec{a}_n)$. For example if we add a multiple of column 1 to column 2 we get

$$\begin{aligned} \det(\vec{a}_1, c\vec{a}_1 + \vec{a}_2, \vec{a}_3, \ldots, \vec{a}_n) &= \det c(\vec{a}_1, \vec{a}_1, \vec{a}_3, \ldots, \vec{a}_n) + \det(\vec{a}_1, \vec{a}_2, \vec{a}_3, \ldots, \vec{a}_n) \\ &= \det(\vec{a}_1, \vec{a}_2, \vec{a}_3, \ldots, \vec{a}_n) \end{aligned}$$

- $\det cA = c^n \det A$.

We will now discuss cofactor expansion.

Definition 4.3.5. *Let $A \in M_{n\times n}(F)$. For $i, j \in [1, n]$ define the (i, j)-minor of A (written $A(i|j)$) to be the $(n-1)\times(n-1)$ matrix obtained from A by removing the i-th row and the j-th column.*

Theorem 4.3.6 (Laplace expansion)**.** *Let $A \in M_{n\times n}(F)$ for $n \geq 2$ and fix $j \in [1, n]$. We have*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det A(i|j) \ .$$

Proof. Let us begin by taking $j = 1$. Now write the column $\vec{a}_1$ as

$$\vec{a}_1 = A_{1,1}e_1 + a_{2,1}e_2 + \cdots + A_{n,1}e_n \ .$$

Then we get

$$\det A = \det(\vec{a}_1, \ldots, \vec{a}_n) = \sum_{i=1}^{n} A_{i,1} \det(e_i, \vec{a}_2, \ldots, \vec{a}_n) \ . \tag{2}$$

We now consider the term $\det(e_i, \vec{a}_2, \ldots, \vec{a}_n)$. This is the determinant of the following matrix:

$$\begin{pmatrix} 0 & A_{1,2} & \cdots & A_{1,n} \\ & & \cdots & \\ 1 & A_{i,2} & \cdots & A_{i,n} \\ & & \cdots & \\ 0 & A_{n,2} & \cdots & A_{n,n} \end{pmatrix} \ .$$

Here, the first column is 0 except for a 1 in the i-th spot. We can now swap the i-th row to the top using $i-1$ adjacent transpositions $(12)\cdots(i-1\ i)$. We are left with the determinant of the matrix

$$(-1)^{i-1} \begin{pmatrix} 1 & A_{i,2} & \cdots & A_{i,n} \\ 0 & A_{1,2} & \cdots & A_{1,n} \\ & & \cdots & \\ 0 & A_{i-1,2} & \cdots & A_{i-1,n} \\ 0 & A_{i+1,2} & \cdots & A_{i+1,n} \\ & & \cdots & \\ 0 & A_{n,2} & \cdots & A_{n,n} \end{pmatrix} \ .$$

This is a block matrix of the form

$$\begin{pmatrix} 1 & B \\ 0 & A(i|1) \end{pmatrix} \ .$$

By the remarks earlier, the determinant is equal to $\det A(i|j) \times 1$. Plugging this into formula (5), we get

$$\det A = \sum_{i=1}^{n} (-1)^{i-1} A_{i,1} \det A(i|1) \ ,$$

which equals $\sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det A(i|j)$.

If $j \neq 1$ then we perform $j-1$ adjacent column switches to get the j-th column to the first. This gives us a new matrix $\widetilde{A}$. For this matrix, the formula holds. Compensating for

the switches,

$$\begin{aligned}\det A = (-1)^{j-1}\det \widetilde{A} &= (-1)^{j-1}\sum_{i=1}^{n}(-1)^{i-1}\widetilde{A}_{i,1}\det \widetilde{A}(i|1) \\ &= \sum_{i=1}^{n}(-1)^{i+j}A_{i,j}\det A(i|j)\ .\end{aligned}$$

Check the last equality. This completes the proof.

□

We have discussed determinants of matrices (and of n vectors in F^n). We will now define for transformations.

Definition 4.3.7. *Let V be a finite dimensional vector space over a field F. If $T : V \to V$ is linear, we define* $\det T$ *as* $\det[T]_\beta^\beta$ *for any basis β of V.*

- Note that $\det T$ does not depend on the choice of basis. Indeed, if β' is another basis,

$$\det[T]_{\beta'}^{\beta'} = \det\left([I]_\beta^{\beta'}[T]_\beta^\beta[I]_{\beta'}^\beta\right) = \det[T]_\beta^\beta\ .$$

- If T and U are linear transformations from V to V then $\det TU = \det T \det U$.
- $\det cT = c^{dim\ V}\det T$.
- $\det T = 0$ if and only if T is non-invertible.

4.4 Exercises

Notation:

1. $\underline{n} = \{1, \ldots, n\}$ is the finite set of natural numbers between 1 and n;
2. S_n is the set of all bijective maps $\underline{n} \to \underline{n}$;
3. For a sequence $k_1, \ldots, k_t$ of distinct elements of $\underline{n}$, we denote by $(k_1 k_2 \ldots k_t)$ the element σ of S_n which is defined by

$$\sigma(i) = \begin{cases} k_s, & i = k_{s-1}, 1 < s < t+1 \\ k_1, & i = k_t \\ i, & i \notin \{k_1, \ldots, k_t\} \end{cases}$$

   Elements of this form are called cycles (or t-cycles). Two cycles $(k_1 \ldots k_t)$ and $(l_1 \ldots l_s)$ are called disjoint if the sets $\{k_1, \ldots, k_t\}$ and $\{l_1, \ldots, l_s\}$ are disjoint.

4. Let $\sigma \in S_n$. A subset $\{k_1, \ldots, k_t\} \subset \underline{n}$ is called an orbit of σ if the following conditions hold

- For any $j \in \mathbb{N}$ there exists an $1 \leq i \leq t$ such that $\sigma^j(k_1) = k_i$.
- For any $1 \leq i \leq t$ there exists a $j \in \mathbb{N}$ such that $k_i = \sigma^j(k_1)$.

Here σ^j is the product of j-copies of σ.

5. Let V and W be two vector spaces over an arbitrary field F, and $k \in \mathbb{N}$. Recall that a k-linear map $f : V^k \to W$ is called
   - alternating, if $f(v_1, \ldots, v_k) = 0$ whenever the vectors $(v_1, \ldots, v_k)$ are not distinct;
   - skew-symmetric, if $f(v_1, \ldots, v_k) = -f(v_{\tau(1)}, \ldots, v_{\tau(k)})$ for any transposition $\tau \in S_k$;
   - symmetric, if $f(v_1, \ldots, v_k) = f(v_{\tau(1)}, \ldots, v_{\tau(k)})$ for any transposition $\tau \in S_k$.
6. If k and n are positive integers such that $k \leq n$ the binomial coefficient $\binom{n}{k}$ is defined as
$$\binom{n}{k} = \frac{n!}{k!(n-k)!} .$$
   Note that this number is equal to the number of distinct subsets of size k of a finite set of size n.
7. Given an $\mathbb{F}$-vector space V, denote by $\mathrm{Alt}^k(V)$ the set of alternating k-linear forms (functions) on V; that is, the set of alternating k-linear map $V^k \to \mathbb{F}$.

Exercises:

1. Prove that composition of maps defines a group law on S_n. Show that this group is abelian only if $n \leq 2$.
2. Let $f : S_n \to \mathbb{Z}$ be a function which is multiplicative, i.e. $f(\sigma\tau) = f(\sigma)f(\tau)$. Show that f must be one of the following three functions: $f(\sigma) = 0$, $f(\sigma) = 1$, $f(\sigma) = \mathrm{sgn}(\sigma)$.
3. (From Dummit-Foote) List explicitly the 24 permutations of degree 4 and state which are odd and which are even.
4. Let $k_1, \ldots, k_t$ be a sequence of distinct elements of $\underline{n}$. Show that $\mathrm{sgn}((k_1 \ldots k_t)) = (-1)^{t-1}$.
5. Let $\pi \in S_n$ be the element $(k_1 \ldots k_t)$ from the previous exercise, and let $\sigma \in S_n$ be any element. Find a formula for the element $\sigma\pi\sigma^{-1}$.
6. Let $\sigma = (k_1 \ldots k_t)$ and $\tau = (l_1 \ldots l_s)$ be disjoint cycles. Show that then $\sigma\tau = \tau\sigma$. One says that σ and τ commute.
7. Let $\sigma \in S_n$. Show that σ can be written as a product of disjoint (and hence, by the previous exercise, commuting) cycles.
   Hint: Consider the orbits of σ.

8. If G is a group and $S \subseteq G$ is a subset, define $\langle S \rangle$ to be the intersection of all subgroups of G that contain S. (This is the subgroup of G generated by S.)

   (a) Show that if S is a subset of G then $\langle S \rangle$ is a subgroup of G.

   (b) Let S be a subset of G and define

   $$\overline{S} = S \cup \{s^{-1} : s \in S\} .$$

   Show that

   $$\langle S \rangle = \{a_1 \cdots a_k : k \geq 1 \text{ and } a_i \in \overline{S} \text{ for all i}\} .$$

9. Prove that $S_n = \langle (12), (12 \cdots n) \rangle$.
   Hint: Use exercise 5.

10. Let V be an finite-dimensional vector space over some field F, W an arbitrary vector space over F, and $k > \dim(V)$. Show that every alternating k-linear function $V^k \to W$ is identically zero. Give an example (choose F, V, W, and k as you wish, as long as $k > \dim(V)$) of a skew-symmetric k-linear function $V^k \to W$ which is not identically zero.

11. (From Hoffman-Kunze) Let $\mathbb{F}$ be a field and $f : \mathbb{F}^2 \to \mathbb{F}$ be a 2-linear alternating function. Show that

    $$f(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix}) = (ad - bc) f(e_1, e_2).$$

    Find an analogous formula for $\mathbb{F}^3$. Deduce from this the formula for the determinant of a 2×2 and a 3×3 matrix.

12. Let V be an n-dimensional vector space over a field $\mathbb{F}$. Suppose that $f : V^n \to \mathbb{F}$ is an n-linear alternating function such that $f(v_1, \ldots, v_n) \neq 0$ for some basis $\{v_1, \ldots, v_n\}$ of V. Show that $f(u_1, \ldots, u_n) = 0$ implies that $\{u_1, \ldots, u_n\}$ is linearly dependent.

13. Let V and W be vector spaces over a field F, and $f : V \to W$ a linear map.

    (a) For $\eta \in \mathrm{Alt}^k(W)$ let $f^*\eta$ be the function on V^k defined by

    $$[f^*\eta](v_1, \ldots, v_k) = \eta(f(v_1), \ldots, f(v_k)).$$

    Show that $f^*\eta \in \mathrm{Alt}^k(V)$.

    (b) Show that in this way we obtain a linear map $f^* : \mathrm{Alt}^k(W) \to \mathrm{Alt}^k(V)$.

    (c) Show that, given a third vector space X over F and a linear map $g : W \to X$, one has $(g \circ f)^* = f^* \circ g^*$.

    (d) Show that if f is an isomorphism, then so is f^*.

14. For $n \geq 2$, we call $M \in M_{n\times n}(\mathbb{F})$ a *block upper-triangular* matrix if there exists k with $1 \leq k \leq n-1$ and matrices $A \in M_{k\times k}(\mathbb{F})$, $B \in M_{k\times(n-k)}(\mathbb{F})$ and $C \in M_{(n-k)\times(n-k)}(\mathbb{F})$ such that M has the form

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} .$$

That is, the elements of M are given by

$$M_{i,j} = \begin{cases} A_{i,j} & 1 \leq i \leq k,\ 1 \leq j \leq k \\ B_{i,j-k} & 1 \leq i \leq k,\ k < j \leq n \\ 0 & k < i \leq n,\ 1 \leq j \leq k \\ C_{i-k,j-k} & k < i \leq n,\ k < j \leq n \end{cases} .$$

We will show in this exercise that

$$\det M = \det A \cdot \det C \ . \tag{3}$$

(a) Show that if $\det C = 0$ then formula (3) holds.

(b) Suppose that $\det C \neq 0$ and define a function $\hat{A} \mapsto \phi\left(\hat{A}\right)$ for $\hat{A} \in M_{k\times k}(\mathbb{F})$ by

$$\phi\left(\hat{A}\right) = [\det C]^{-1} \det \begin{pmatrix} \hat{A} & B \\ 0 & C \end{pmatrix} .$$

That is, $\phi\left(\hat{A}\right)$ is a scalar multiple of the determinant of the block upper-triangular matrix we get when we replace A by $\hat{A}$ and keep B and C fixed.

i. Show that ϕ is k-linear as a function of the columns of $\hat{A}$.

ii. Show that ϕ is alternating and satisfies $\phi(I_k) = 1$, where I_k is the $k \times k$ identity matrix.

iii. Conclude that formula (3) holds when $\det C \neq 0$.

15. Suppose that $A \in M_{n\times n}(\mathbb{F})$ is *upper-triangular*; that is, $a_{i,j} = 0$ when $1 \leq j < i \leq n$. Show that $\det A = a_{1,1}a_{2,2}\cdots a_{n,n}$.

16. Let $A \in M_{n\times n}(\mathbb{F})$ such that $A^k = 0$ for some $k \geq 0$. Show that $\det A = 0$.

17. Let $a_0, a_1, \ldots, a_n$ be distinct complex numbers. Write $M_n(a_0, \ldots, a_n)$ for the matrix

$$\begin{pmatrix} 1 & a_0 & a_0^2 & \cdots & a_0^n \\ 1 & a_1 & a_1^2 & \cdots & a_1^n \\ & & & \cdots & \\ 1 & a_n & a_n^2 & \cdots & a_n^n \end{pmatrix} .$$

The goal of this exercise is to show that

$$\det M_n(a_0, \ldots, a_n) = \prod_{0\leq i<j\leq n} (a_j - a_i) \ . \tag{4}$$

We will argue by induction on n.

(a) Show that if $n = 2$ then formula (5) holds.

(b) Now suppose that $k \geq 3$ and that formula (5) holds for all $2 \leq n \leq k$. Show that it holds for $n = k + 1$ by completing the following outline.

i. Define the function $f : \mathbb{C} \to \mathbb{C}$ by $f(z) = \det M_n(z, a_1, \ldots, a_n)$. Show that f is a polynomial of degree at most n.

ii. Find all the zeros of f.

iii. Show that the coefficient of z^n is $(-1)^n \det M_{n-1}(a_1, \ldots, a_n)$.

iv. Show that formula (5) holds for $n = k + 1$, completing the proof.

18. Show that if $A \in M_{n\times n}(\mathbb{F})$ then $\det A = \det A^t$, where A^t is the transpose of A.

19. Let V be an n-dimensional $\mathbb{F}$-vector space and $k \leq n$. The purpose of this problem is to show that

$$\dim(\mathrm{Alt}^k(V)) = \binom{n}{k},$$

by completing the following steps:

(a) Let W be a subspace of V and let $B = (v_1, \ldots, v_n)$ be a basis for V such that $(v_1, \ldots, v_k)$ is a basis for W. Show that

$$p_{W,B} : V \to W, \qquad v_i \mapsto \begin{cases} v_i, & i \leq k \\ 0, & i > k \end{cases}$$

specifies a linear map with the property that $p_{W,B} \circ p_{W,B} = p_{W,B}$. Such a map (that is, a T such that $T \circ T = T$) is called a *projection*.

(b) With W and B as in the previous part, let d_W be a non-zero element of $\mathrm{Alt}^k(W)$. Show that $[p_{W,B}]^* d_W$ is a non-zero element of $\mathrm{Alt}^k(V)$. (Recall this notation from exercise 3.)

(c) Let $B = (v_1, \ldots, v_n)$ be a basis of V. Let $S_1, \ldots, S_t$ be subsets of $\underline{n} = \{1, \ldots, n\}$. Assume that each S_i has exactly k elements and no two S_i's are the same. Let

$$W_i = \mathrm{Span}(\{v_j : j \in S_i\}).$$

For $i = 1, \ldots, t$, let $d_{W_i} \in \mathrm{Alt}^k(W_i)$ be non-zero. Show that the collection $\{[p_{W_i,B}]^* d_{W_i} : 1 \leq i \leq t\}$ of elements of $\mathrm{Alt}^k(V)$ is linearly independent.

(d) Show that the above collection is also generating, by taking an arbitrary $\eta \in \mathrm{Alt}^k(V)$, an arbitrary collection $u_1, \ldots, u_k$ of vectors in V, expressing each u_i as a linear combination of $(v_1, \ldots, v_k)$ and plugging those linear combinations into η.

In doing this, it may be helpful (although certainly not necessary) to assume that d_{W_i} is the unique element of $\mathrm{Alt}^k(W_i)$ with $d_{W_i}(w_1, \ldots, w_k) = 1$, where $S_i = (w_1, \ldots, w_k)$.

20. Let $A \in M_{n\times n}(F)$ for some field F. Recall that if $1 \leq i, j \leq n$ then the (i,j)-th minor of A, written $A(i|j)$, is the $(n-1)\times(n-1)$ matrix obtained by removing the i-th row and j-th column from A. Define the *cofactor*

$$C_{i,j} = (-1)^{i+j} \det A(i|j) \ .$$

Note that the Laplace expansion for the determinant can be written

$$\det A = \sum_{i=1}^{n} A_{i,j} C_{i,j} \ .$$

(a) Show that if $1 \leq i, j, k \leq n$ with $j \neq k$ then

$$\sum_{i=1}^{n} A_{i,k} C_{i,j} = 0 \ .$$

(b) Define the *classical adjoint* of A, written adj A, by

$$(\text{adj } A)_{i,j} = C_{j,i} \ .$$

Show that $(\text{adj } A)A = (\det A)I$.

(c) Show that $A(\text{adj } A) = (\det A)I$ and deduce that if A is invertible then

$$A^{-1} = (\det A)^{-1} \text{adj } A \ .$$

Hint: begin by applying the result of the previous part to A^t.

(d) Use the formula in the last part to find the inverses of the following matrices:

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 6 & 0 & 1 & 1 \end{pmatrix} \ .$$

21. Consider a system of equations in n variables with coefficients from a field F. We can write this as $AX = Y$ for an $n \times n$ matrix A, an $n \times 1$ matrix X (with entries $x_1, \ldots, x_n$) and an $n \times 1$ matrix Y (with entries $y_1, \ldots, y_n$). Given the matrices A and Y we would like to solve for X.

(a) Show that

$$(\det A)x_j = \sum_{i=1}^{n} (-1)^{i+j} y_i \det A(i|j) \ .$$

(b) Show that if $\det A \neq 0$ then we have

$$x_j = (\det A)^{-1} \det B_j \ ,$$

where B_j is an $n \times n$ matrix obtained from A by replacing the j-th column of A by Y. This is known as *Cramer's rule*.

(c) Solve the following systems of equations using Cramer's rule.

$$\begin{cases} 2x - y + z & = 3 \\ 2y - z & = 1 \\ y - x & = 1 \end{cases} \qquad \begin{cases} 2x - y + z - 2t & = -5 \\ 2x + 2y - 3z + t & = -1 \\ -x + y - z & = -1 \\ 4x - 3y + 2z - 3t & = -8 \end{cases}$$

22. Find the determinants of the following matrices. In the first example, the entries are from $\mathbb{R}$ and in the second they are from $\mathbb{Z}_3$.

$$\begin{pmatrix} 1 & 4 & 5 & 7 \\ 0 & 0 & 2 & 3 \\ 1 & 4 & -1 & 7 \\ 2 & 8 & 10 & 14 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

5 Eigenvalues

5.1 Definitions and the characteristic polynomial

The simplest matrix is λI for some $\lambda \in F$. These act just like the field F. What is the second simplest? A diagonal matrix; that is, a matrix D that satisfied $D_{ij} = 0$ if $i \neq j$.

Definition 5.1.1. *Let V be a finite dimensional vector space over F. An linear transformation T is called diagonalizable if there exists a basis β such that $[T]_\beta^\beta$ is diagonal.*

Proposition 5.1.2. *T is diagonalizable if and only if there exists a basis $\{v_1, \ldots, v_n\}$ of V and scalars $\lambda_1, \ldots, \lambda_n$ such that*

$$T(v_i) = \lambda_i v_i \text{ for all } i \ .$$

Proof. Suppose that T is diagonalizable. Then there is a basis $\{v_1, \ldots, v_n\}$ such that $[T]_\beta^\beta$ is diagonal. Then, writing $D = [T]_\beta^\beta$,

$$[Tv_k]_\beta = D[v_k]_\beta = D \cdot e_k = D_{k,k} \ .$$

Now we can choose $\lambda_k = D_{k,k}$.

If the second condition holds then we see that $[T]_\beta^\beta$ is diagonal with entries $D_{i,j} = 0$ if $i \neq j$ and $D_{i,i} = \lambda_i$. □

This motivates the following definition.

Definition 5.1.3. *If $T : V \to V$ is linear then we call a nonzero vector v an eigenvector of T if there exists $\lambda \in F$ such that $T(v) = \lambda v$. In this case we call λ the eigenvalue for v.*

Theorem 5.1.4. *If $\dim V < \infty$ and $T : V \to V$ is linear then the following are equivalent.*

1. *λ is an eigenvalue of T (for some eigenvector).*
2. *$T - \lambda I$ is not invertible.*
3. *$\det(T - \lambda I) = 0$.*

Proof. If (1) holds then the eigenvector v is a non-zero vector in the nullspace of $T - \lambda I$. Thus $T - \lambda I$ is not invertible. We already know that (2) and (3) are equivalent. If $T - \lambda I$ is not invertible then there is a non-zero vector in its nullspace. This vector is an eigenvector. □

Definition 5.1.5. *If $T : V \to V$ is linear and $\dim V = n$ then we define the characteristic polynomial $c : F \to F$ by*

$$c(\lambda) = \det(T - \lambda I) \ .$$

- Note that $c(\lambda)$ does not depend on the choice of basis.

- We can write in terms of the matrix.

$$c(\lambda) = \det(T - \lambda I) = \det([T - \lambda I]_\beta^\beta) = \det([T]_\beta^\beta - \lambda Id) \ .$$

- Eigenvalues are exactly the roots of $c(\lambda)$.

Facts about $c(x)$.

1. c is a polynomial of degree n. We can see this by analyzing each term in the definition of the determinant: set $B = A - xI$ and see

$$sgn(\pi) \ B_{1,\pi(1)} \cdots B_{n,\pi(n)} \ .$$

   Each term $B_{i,\pi(i)}$ is a polynomial of degree 0 or 1 in x. So the product has degree at most n. A sum of such polynomials is a polynomial of degree at most n.

   In fact, the only term of degree n is

$$sgn(id) \ B_{1,1} \cdots B_{n,n} = (A_{1,1} - x) \cdots (A_{n,n} - x) \ .$$

   So the coefficient of x^n is $(-1)^n$.

2. In the above description of $c(x)$, all terms corresponding to non-identity permutations π have degree at most $n - 2$. Therefore the degree $n - 1$ term comes from $(A_{1,1} - x) \cdots (A_{n,n} - x)$ as well. It is

$$(-1)^{n-1} x^{n-1} \left[A_{1,1} + \cdots + A_{n,n}\right] = (-1)^{n-1} x^{n-1} Tr \ A \ .$$

3. Because $c(0) = det \ A$,

$$c(x) = (-1)^n \left[x^n - Tr \ A \ x^{n-1} + \cdots + det \ A\right] \ .$$

   For $F = \mathbb{C}$ (or any field so that $c(x)$ splits), we can always write $c(x) = (-1)^n (x - \lambda_1) \cdots (x - \lambda_n)$. Thus the constant term in the polynomial is $(-1)^n \prod \lambda_i$. Therefore

$$c(x) = (-1)^n \left[x^n - Tr \ A \ x^{n-1} + \cdots + \prod_{i=1}^{n} \lambda_i\right] \ .$$

   We find $det \ A = \prod \lambda_i$ in $\mathbb{C}$.

Theorem 5.1.6. *If dim $V = n$ and $c(\lambda)$ has n distinct roots then T is diagonalizable. The converse is not true.*

Proof. Write the eigenvalues as $\lambda_1, \ldots, \lambda_n$. For each λ_i we have an eigenvector v_i. We claim that the v_i's are linearly independent. This follows from the lemma:

Lemma 5.1.7. *If $\lambda_1, \ldots, \lambda_k$ are k-distinct eigenvalues associated to eigenvectors $v_1, \ldots, v_k$ then $\{v_1, \ldots, v_k\}$ is linearly independent.*

Proof. Suppose that

$$a_1 v_1 + \cdots + a_k v_k = \vec{0} \ .$$

Take T of both sides

$$a_1 \lambda_1 v_1 + \cdots + a_k \lambda_k v_k = \vec{0} \ .$$

Keep doing this $k-1$ times so we get the system of equations

$$\begin{aligned} a_1 v_1 + \cdots + a_k v_k &= \vec{0} \\ a_1 \lambda_1 v_1 + \cdots + a_k \lambda_k v_k &= \vec{0} \\ &\cdots \\ a_1 \lambda_1^{k-1} v_1 + \cdots + a_k \lambda_k^{k-1} v_k &= \vec{0} \end{aligned}$$

Write each v_i as $[v_i]_\beta$ for some basis β. This is then equivalent to the matrix equation

$$\begin{pmatrix} a_1(v_1)_\beta & a_2(v_2)_\beta & \cdots & a_k(v_k)_\beta \end{pmatrix} \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{k-1} \\ & & \cdots & \\ 1 & \lambda_k & \cdots & \lambda_k^{k-1} \end{pmatrix} = \begin{pmatrix} \vec{0} & \vec{0} & \cdots & \vec{0} \end{pmatrix} \ .$$

Here the left matrix is $n \times k$ and has j-th column equal to the column vector $a_j[v_j]_\beta$. But the middle matrix has nonzero determinant when the λ_i's are distinct: its determinant is $\prod_{1 \le i < j \le k} (\lambda_j - \lambda_i)$. Therefore it is invertible. Multiplying both sides by its inverse, we find $a_i v_i = \vec{0}$ for all i. Since $v_i \neq \vec{0}$, it follows that $a_i = 0$ for all i.

□

□

5.2 Eigenspaces and the main diagonalizability theorem

Definition 5.2.1. *If $\lambda \in F$ we define the eigenspace*

$$E_\lambda = N(T - \lambda I) = \{v \in V : T(v) = \lambda v\} \ .$$

Note that E_λ is a subspace even if λ is not an eigenvalue. Furthermore,

$$E_\lambda \neq \{0\} \text{ if and only if } \lambda \text{ is an eigenvalue of } T$$

and

$$E_\lambda \text{ is } T\text{-invariant for all } \lambda \in F \ .$$

What this means is that if $v \in E_\lambda$ then so is $T(v)$:

$$(T - \lambda I)(T(v)) = (T - \lambda I)(\lambda v) = \lambda (T - \lambda I) v = \vec{0} \ .$$

Definition 5.2.2. *If $W_1, \ldots, W_k$ are subspaces of a vector space V then we write*

$$W_1 \oplus \cdots \oplus W_k$$

for the sum space $W_1 + \cdots + W_k$ and say the sum is direct if

$$W_j \cap [W_1 + \cdots + W_{j-1}] = \{0\} \text{ for all } j = 2, \ldots, k \ .$$

We also say the subspaces are independent.

Theorem 5.2.3. *If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T then*

$$E_{\lambda_1} + \cdots + E_{\lambda_k} = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k} \ .$$

Furthermore

$$dim \ \left(\sum_{i=1}^{k} E_{\lambda_i} \right) = \sum_{i=1}^{k} dim \ E_{\lambda_i} \ .$$

Proof. The theorem will follows directly from the following lemma. The proof is in homework.

Lemma 5.2.4. *Let $W_1, \ldots, W_k$ be subspaces of V. The following are equivalent.*

1. $$W_1 + \cdots + W_k = W_1 \oplus \cdots \oplus W_k \ .$$
2. *Whenever $w_1 + \cdots + w_k = \vec{0}$ for $w_i \in W_i$ for all i, we have $w_i = \vec{0}$ for all i.*
3. *Whenever β_i is a basis for W_i for all i, the β_i's are disjoint and $\beta := \cup_{i=1}^{k} \beta_i$ is a basis for $\sum_{i=1}^{k} W_i$.*

So take $w_1 + \cdots + w_k = \vec{0}$ for $w_i \in E_{\lambda_i}$ for all i. Note that each nonzero w_i is an eigenvector for the eigenvalue λ_i. Remove all the zero ones. If we are left with any nonzero ones, by the previous theorem, they must be linearly independent. This would be a contradiction. So they are all zero.

For the second claim take bases β_i of E_{λ_i}. By the lemma, $\cup_{i=1}^{k} \beta_i$ is a basis for $\sum_{i=1}^{k} E_{\lambda_i}$. This implies the claim.

□

Theorem 5.2.5 (Main diagonalizability theorem)**.** *Let $T : V \to V$ be linear and dim $V < \infty$. The following are equivalent.*

1. *T is diagonalizable.*
2. *$c(x)$ can be written as $(-1)^n (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}$, where $n_i = dim \ E_{\lambda_i}$ for all i.*
3. *$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$, where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T.*

Proof. Suppose first that T is diagonalizable. Then there exists a basis β of eigenvectors for T; that is, for which $[T]_\beta^\beta$ is diagonal. Clearly each diagonal element is an eigenvalue. For each i, call n_i the number of entries on the diagonal that are equal to λ_i. Then $[T - \lambda_i I]_\beta^\beta$ has n_i number of zeros on the diagonal. All other diagonal entries must be non-zero, so the nullspace has dimension n_i. In other words, $n_i = dim\ E_{\lambda_i}$.

Suppose that condition 2 holds. Since c is a polynomial of degree n we must have

$$dim\ E_{\lambda_1} + \cdots + dim`E_{\lambda_k} = dim\ V\ .$$

However since the λ_i's are distinct the previous theorem gives that

$$dim\ \sum_{i=1}^{k} E_{\lambda_i} = dim\ V\ .$$

In other words, $V = \sum_{i=1}^k E_{\lambda_i}$. The previous theorem implies that the sum is direct and the claim follows.

Suppose that condition 3 holds. Then take β_i a basis for E_{λ_i} for all i. Then $\beta = \cup_{i=1}^k \beta_i$ is a basis for V. We claim that $[T]_\beta^\beta$ is diagonal. This is because each vector in β is an eigenvector. This proves 1 and completes the proof. □

5.3 Exercises

1. Let V be an $\mathbb{F}$-vector space and let $W_1, \ldots, W_k$ be subspaces of V. Recall the definition of the sum $\sum_{i=1}^k W_i$. It is the subspace of V given by

   $$\{w_1 + \cdots + w_k : w_i \in W_i\}.$$

   Recall further that this sum is called *direct*, and written as $\bigoplus_{i=1}^k$ if and only if for all $1 < i \leq n$ we have

   $$W_i \cap (\sum_{j=1}^{i-1} W_j) = \{0\}.$$

   Show that the following statements are equivalent:

   (a) The sum $\sum_{i=1}^k W_i$ is direct.

   (b) For any collection $\{w_1, \ldots, w_k\}$ with $w_i \in W_i$ for all i, we have

   $$\sum_{i=1}^{k} w_i = 0 \Rightarrow \forall i : w_i = 0.$$

   (c) If, for each i, β_i is a basis of W_i, then the β_i's are disjoint and their union $\beta = \sqcup_{i=1}^k \beta_i$ is a basis for the subspace $\sum_{i=1}^k W_i$.

   (d) For any $v \in \sum_{i=1}^k W_i$ there exist unique vectors $w_1, \ldots, w_k$ such that $w_i \in W_i$ for all i and $v = \sum_{i=1}^k w_i$.

2. Let V be an $\mathbb{F}$-vector space. Recall that a linear map $p \in L(V,V)$ is called a projection if $p \circ p = p$.

   (a) Show that if p is a projection, then so is $q = \mathrm{id}_V - p$, and we have $p \circ q = q \circ p = 0$.

   (b) Let $W_1, \ldots, W_k$ be subspaces of V and assume that $V = \bigoplus_{i=1}^k W_i$. For $1 \leq t \leq k$, show that there is a unique element $p_t \in \mathcal{L}(V, W_t)$ such that for any choice of vectors $w_1, \ldots, w_k$ such that $w_j \in W_j$ for all j,

$$p_t(w_j) = \begin{cases} w_j & j = t \\ 0 & j \neq t \end{cases}.$$

   (c) Show that each p_t defined in the previous part is a projection. Show furthermore that $\sum_{i=1}^k p_t = \mathrm{id}_V$ and that for $t \neq s$ we have $p_t \circ p_s = 0$.

   (d) Show conversely that if $p_1, \ldots, p_t \in L(V,V)$ are projections with the properties (a) $\sum_{i=1}^k p_i = \mathrm{id}_V$ and (b) $p_i \circ p_j = 0$ for all $i \neq j$, and if we put $W_i = R(p_i)$, then $V = \bigoplus_{i=1}^k W_i$.

3. Let V be an $\mathbb{F}$-vector space.

   (a) If $U \subset V$ is a subspace, W is another $\mathbb{F}$-vector space, and $f \in L(V,W)$, define $f|_U : U \to W$ by

$$f|_U(u) = f(u) \qquad \forall u \in U.$$

   Show that the map $f \mapsto f|_U$ is a linear map $L(V,W) \to L(U,W)$. It is called the *restriction map*.

   (b) Let $f \in L(V,V)$ and let $U \subset V$ be an f-invariant subspace (that is, a subspace U with the property that $f(u) \in U$ whenever $u \in U$). Observe that $f|_U \in L(U,U)$. If $W \subset V$ is another f-invariant subspace and $V = U \oplus W$, show that

$$N(f) = N(f|_U) \oplus N(f|_W), \quad R(f) = R(f|_U) \oplus R(f|_W), \quad \det(f) = \det(f|_U)\det(f|_W).$$

   (c) Let $f, g \in L(V,V)$ be two *commuting* endomorphisms, i.e. we have $f \circ g = g \circ f$. Show that $N(g)$ and $R(g)$ are f-invariant subspaces of V.

4. Consider the matrix

$$A := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

   Show that there does not exist an invertible matrix $P \in M_{2\times 2}(\mathbb{C})$ such that PAP^{-1} is diagonal.

5. Let V be a finite-dimensional $\mathbb{F}$-vector space and $f \in L(V,V)$. Observe that for each natural number k we have $N(f^k) \subset N(f^{k+1})$.

   (a) Show that there exists a natural number k so that $N(f^k) = N(f^{k+1})$.

(b) Show further that for all $l \geq k$ one has $N(f^l) = N(f^k)$.

6. Let V be a finite-dimensional $\mathbb{F}$-vector space and $f \in L(V,V)$.

   (a) Let $U \subset V$ be an f-invariant subspace and $\beta = (v_1, \ldots, v_n)$ a basis of V such that $\beta' = (v_1, \ldots, v_k)$ is a basis for U. Show that

   $$[f]^{\beta}_{\beta} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

   with $A = [f|_U]^{\beta'}_{\beta'} \in M_{k\times k}(\mathbb{F})$, $B \in M_{k\times(n-k)}(\mathbb{F})$, and $C \in M_{(n-k)\times(n-k)}(\mathbb{F})$.

   (b) Let $U, W \subset V$ be f-invariant subspaces with $V = U \oplus W$. Let β' be a basis for U, β'' a basis for W, and $\beta = \beta' \sqcup \beta''$. Show that

   $$[f]^{\beta}_{\beta} = \begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$$

   with $A = [f|_U]^{\beta'}_{\beta'}$ and $C = [f|_W]^{\beta''}_{\beta''}$.

7. Let V be a finite-dimensional $\mathbb{F}$-vector space. Recall that an element $p \in L(V,V)$ with $p^2 = p$ is called a *projection*. On the other hand, an element $i \in L(V,V)$ with $i^2 = \mathrm{id}_V$ is called an *involution*.

   (a) Assume that $\mathrm{char}(\mathbb{F}) \neq 2$. Show that the maps

   $$\begin{array}{rcl} \text{Involutions on V} & \rightleftarrows & \text{Projections in V} \\ i & \mapsto & \frac{1}{2}(\mathrm{id}_V + i) \\ 2p - \mathrm{id}_V & \mapsfrom & p \end{array}$$

   are mutually inverse bijections.

   (b) Show that if $p \in L(V,V)$ is a projection, then the only eigenvalues of p are 0 and 1. Furthermore, $V = E_0(p) \oplus E_1(p)$ (the eigenspaces for p). That is, p is diagonalizable.

   (c) Show that if $i \in L(V,V)$ is an involution, then the only eigenvalues of i are $+1$ and -1. Furthermore, $V = E_{+1}(i) \oplus E_{-1}(i)$. That is, i is diagonalizable.

   Observe that projections and involutions are examples of diagonalizable endomorphisms which do not have $\dim(V)$-many distinct eigenvalues.

8. In this problem we will show that every endomorphism of a vector space over an algebraically closed field can be represented as an upper triangular matrix. This is a simpler result than (and is implied by) the Jordan Canonical form, which we will cover in class soon.

We will argue by (strong) induction on the dimension of V. Clearly the result holds for $dim\ V = 1$. So suppose that for some $k \geq 1$ whenever $dim\ W \leq k$ and $U : W \to W$ is linear, we can find a basis of W with respect to which the matrix of U is upper-triangular. Further, let V be a vector space of dimension $k+1$ over $\mathbb{F}$ and $T : V \to V$ be linear.

(a) Let λ be an eigenvalue of T. Show that the dimension of $R := R(T - \lambda I)$ is strictly less than $dim\ V$ and that R is T-invariant.

(b) Apply the inductive hypothesis to $T|_R$ to find a basis of R with respect to which $T|_R$ is upper-triangular. Extend this to a basis for V and complete the argument.

9. Let $A \in M_{n\times n}(\mathbb{F})$ be upper-triangular. Show that the eigenvalues of A are the diagonal entries of A.

10. Let A be the matrix
$$A = \begin{pmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{pmatrix} .$$

(a) Is A diagonalizable over $\mathbb{R}$? If so, find a basis for $\mathbb{R}^3$ of eigenvectors of A.

(b) Is A diagonalizable over $\mathbb{C}$? If so, find a basis for $\mathbb{C}^3$ of eigenvectors of A.

11. For which values of $a, b, c \in \mathbb{R}$ is the following matrix diagonalizable over $\mathbb{R}$?
$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{pmatrix}$$

12. Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $T : V \to V$ be linear. Suppose that every subspace of V is T-invariant. What can you say about T?

13. Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $T, U : V \to V$ be linear transformations.

(a) Prove that if $I - TU$ is invertible then $I - UT$ is invertible and
$$(I - UT)^{-1} = I + U(I - TU)^{-1}T .$$

(b) Use the previous part to show that TU and UT have the same eigenvalues.

14. Let A be the matrix
$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} .$$

Find A^n for all $n \geq 1$.

Hint: first diagonalize A.

6 Jordan form

6.1 Generalized eigenspaces

It is of course not always true that T is diagonalizable. There can be a couple of reasons for that. First it may be that the roots of the characteristic polynomial do not lie in the field. For instance

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

has characteristic polynomial $x^2 + 1$. Even still it may be that the eigenvalues are in the field, but we still cannot diagonalize. On the homework you will see that the matrix

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

is not diagonalizable over $\mathbb{C}$ (although its eigenvalues are certainly in $\mathbb{C}$). So we resort to looking for a block diagonal matrix.

Suppose that we can show that

$$V = W_1 \oplus \cdots \oplus W_k$$

for some subspaces W_i. Then we can choose a basis for V made up of bases for the W_i's. If the W_i's are T-invariant then the matrix will be in block form.

Definition 6.1.1. *Let $T : V \to V$ be linear. A subspace W of V is T-invariant if $T(w) \in W$ whenever $w \in W$.*

- Each eigenspace is T-invariant. If $w \in E_\lambda$ then

$$(T - \lambda I)T(w) = \lambda(T - \lambda I)w = \vec{0} \ .$$

- Therefore the eigenspace decomposition is a T-invariant direct sum.

To find a general T-invariant direct sum we define generalized eigenspaces.

Definition 6.1.2. *Let $T : V \to V$ be linear. If $\lambda \in F$ then the set*

$$\hat{E}_\lambda = \{v \in V : (T - \lambda I)^k v = \vec{0} \text{ for some } k\}$$

is called the generalized eigenspace for λ.

- This is a subspace. If $v, w \in \hat{E}_\lambda$ and $c \in F$ then there exists k_v and k_w such that

$$(T - \lambda I)^{k_v} v = \vec{0} = (T - \lambda I)^{k_w} w \ .$$

Choosing $k = \max\{k_v, k_w\}$ we find

$$(T - \lambda I)^k (cv + w) = \vec{0} \ .$$

- Each generalized eigenspace is T-invariant. To see this, suppose that $(T-\lambda I)^k v = \vec{0}$. Then because T commutes with $(T-\lambda I)^k$ we have

$$(T-\lambda I)^k Tv = T(T-\lambda I)^k v = \vec{0} \ .$$

To make sure the characteristic polynomial has roots we will take F to be an algebraically closed field. That is, each polynomial with coefficients in F has a root in F.

6.2 Primary decomposition theorem

Theorem 6.2.1 (Primary decomposition theorem)**.** *Let F be algebraically closed and V a finite-dimensional vector space over F. If $T : V \to V$ is linear and $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T then*

$$V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_k} \ .$$

Proof. We follow several steps.

Step 1. $c(x)$ has a root. Therefore T has an eigenvalue. Let λ be this value.

Step 2. Consider the generalized eigenspace $\hat{E}_\lambda$. We first show that there exists k such that

$$\hat{E}_\lambda = N(T-\lambda I)^k \ .$$

Let $v_1, \ldots, v_m$ be a basis for $\hat{E}_\lambda$. Then for each i there is a k_i such that $(T-\lambda I)^{k_i} v_i = \vec{0}$. Choose $k = \max\{k_1, \ldots, k_m\}$. Then $(T-\lambda I)^k$ kills all the basis vectors and thus kills everything in $\hat{E}_\lambda$. Therefore

$$\hat{E}_\lambda \subseteq N(T-\lambda I)^k \ .$$

The other direction is obvious.

Step 3. We now claim that

$$V = R(T-\lambda I)^k \oplus N(T-\lambda I)^k = R(T-\lambda I)^k \oplus \hat{E}_\lambda \ .$$

First we show that the intersection is only the zero vector. Suppose that v is in the intersection. Then $(T-\lambda I)^k v = \vec{0}$ and there exists w such that $(T-\lambda I)^k w = v$. Then $(T-\lambda_1 I)^{2k_1} w = \vec{0}$ so $w \in \hat{E}_{\lambda_1}$. Therefore

$$v = (T-\lambda_1 I)^{k_1} w = \vec{0} \ .$$

By the rank-nullity theorem,

$$dim\ R(T-\lambda_1 I)^{k_1} + dim\ N(T-\lambda_1 I)^{k_1} = dim\ V \ .$$

By the 2-subspace (dimension) theorem,

$$V = R(T-\lambda_1 I)^{k_1} + N(T-\lambda_1 I)^{k_1} \ ,$$

and so it is a direct sum.

Step 4. Write $W_1 = R(T - \lambda_1 I)^{k_1}$ so that

$$V = \hat{E}_{\lambda_1} \oplus W_1 \ .$$

These spaces are T-invariant. To show that note that we know $\hat{E}_{\lambda_1}$ is already. For W_1, suppose that $w \in W_1$. Then there exists u such that

$$w = (T - \lambda_1 I)^{k_1} u \ .$$

So

$$(T - \lambda_1 I)^{k_1}(T - \lambda_1 I)u = (T - \lambda_1 I)w \ .$$

Therefore $(T - \lambda_1 I)w \in W_1$ and thus W_1 is $(T - \lambda_1 I)$-invariant. If $w \in W_1$ then

$$Tw = (T - \lambda_1 I)w + \lambda_1 I w \in W_1 \ ,$$

so W_1 is T-invariant.

Step 5. We now argue by induction and do the base case. Let $e(T)$ be the number of distinct eigenvalues of T. Note $e(T) \geq 1$.

We first assume $e(T) = 1$. In this case we write λ_1 for the eigenvalue and see

$$V = \hat{E}_{\lambda_1} \oplus R(T - \lambda_1 I)^{k_1} = \hat{E}_{\lambda_1} \oplus W_1 \ .$$

We claim that the second space is only the zero vector. Otherwise we restrict T to it to get an operator T_{W_1}. Then T_{W_1} has an eigenvalue λ. So there is a nonzero vector $w \in W_1$ such that

$$Tw = T_{W_1} w = \lambda w \ ,$$

so w is an eigenvector for T. But T has only one eigenvalue so $\lambda = \lambda_1$. This means that $w \in \hat{E}_{\lambda_1}$ and thus

$$w \in \hat{E}_{\lambda_1} \cap W_1 = \{\vec{0}\} \ .$$

This is a contradiction, so

$$V = \hat{E}_{\lambda_1} \oplus \{\vec{0}\} = \hat{E}_{\lambda_1} \ ,$$

and we are done.

Step 6. Suppose the theorem is true for any transformation U with $e(U) = k$ ($k \geq 1$). Then suppose that $e(T) = k + 1$. Let $\lambda_1, \ldots, \lambda_{k+1}$ be the distinct eigenvalues of T and decompose as before:

$$V = \hat{E}_{\lambda_1} \oplus R(T - \lambda_1 I)^{k_1} = \hat{E}_{\lambda_1} \oplus W_1 \ .$$

Now restrict T to W_1 and call it T_{W_1}.

Claim 6.2.2. *T_{W_1} has eigenvalues $\lambda_2, \ldots, \lambda_{k+1}$ with the generalized eigenspaces from T: they are $\hat{E}_{\lambda_2}, \ldots, \hat{E}_{\lambda_{k+1}}$.*

Once we show this we will be done: we will have $e(T_{W_1}) = k$ and so we can apply the theorem:

$$W_1 = \hat{E}_{\lambda_2} \oplus \cdots \oplus \hat{E}_{\lambda_{k+1}} \ ,$$

so

$$V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_{k+1}} \ .$$

Proof. We first show that each of $\hat{E}_{\lambda_2}, \ldots, \hat{E}_{\lambda_{k+1}}$ is in W_1. For this we want a lemma and a definition:

Definition 6.2.3. *If $p(x)$ is a polynomial with coefficients in F and $T : V \to V$ is linear, where V is a vector space over F, we define the transformation*

$$P(T) = a_n T^n + \cdots + a_1 T + a_0 I \ ,$$

where $p(x) = a_n x^n + \cdots + a_1 x + a_0$.

Lemma 6.2.4. *Suppose that $p(x)$ and $q(x)$ are two polynomials with coefficients in F. If they have no common root then there exist polynomials $a(x)$ and $b(x)$ such that*

$$a(x)p(x) + b(x)q(x) = 1 \ .$$

Proof. Homework □

Now choose $v \in \hat{E}_{\lambda_j}$ for some $j = 2, \ldots, k+1$. By the decomposition we can write $v = u + w$ where $u \in \hat{E}_{\lambda_1}$ and $w \in W_1$. We can now write

$$\hat{E}_{\lambda_1} = N(T - \lambda_1 I)^{k_1} \text{ and } \hat{E}_{\lambda_j} = N(T - \lambda_j I)^{k_j}$$

and see

$$\vec{0} = (T - \lambda_j I)^{k_j} v = (T - \lambda_j I)^{k_j} u + (T - \lambda_j I)^{k_j} w \ .$$

However $\hat{E}_{\lambda_1}$ and W_1 are T-invariant so they are $(T - \lambda_j I)^{k_j}$-invariant. This is a sum of vectors equal to zero, where on is in $\hat{E}_{\lambda_1}$, the other is in W_1. Because these spaces direct sum to V we know both vectors are zero. Therefore

$$u \text{ satisfies } (T - \lambda_j I)^{k_j} u = \vec{0} = (T - \lambda_1 I)^{k_1} u \ .$$

In other words, $p(T)u = q(T)u = \vec{0}$, where $p(x) = (x - \lambda_j)^{k_j}$ and $q(x) = (x - \lambda_1)^{k_1}$. Since these polynomials have no root in common we can find $a(x)$ and $b(x)$ as in the lemma. Finally,

$$u = (a(T)p(T) + b(T)q(T))u = \vec{0} \ .$$

This implies that $v = w \in W_1$ and therefore all of $\hat{E}_{\lambda_2}, \ldots, \hat{E}_{\lambda_{k+1}}$ are in W_1.

Because of the above statement, we now know that all of $\lambda_2, \ldots, \lambda_{k+1}$ are eigenvalues of T_{W_1}. Furthermore if λ_{W_1} is an eigenvalue of T_{W_1} then it is an eigenvalue of T. It cannot be λ_1 because then any eigenvector for T_{W_1} with eigenvalue λ_{W_1} would have to be in $\hat{E}_{\lambda_1}$

but also in W_1 so it would be zero, a contradiction. Therefore the eigenvalues of T_{W_1} are precisely $\lambda_2, \ldots, \lambda_{k+1}$.

Let $\hat{E}_{\lambda_j}^{W_1}$ be the generalized eigenspace for T_{W_1} corresponding to λ_j. We want

$$\hat{E}_{\lambda_j}^{W_1} = \hat{E}_{\lambda_j}, \ j = 2, \ldots, k+1 \ .$$

If $w \in \hat{E}_{\lambda_j}^{W_1}$ then there exists k such that $(T_{W_1} - \lambda_j I)^k w = \vec{0}$. But now on W_1, $(T_{W_1} - \lambda_j I)^k$ is the sam as $(T - \lambda_j I)^k$, so

$$(T - \lambda_j I)^k w = (T_{W_1} - \lambda_j I)^k w = \vec{0} \ ,$$

so that $\hat{E}_{\lambda_j}^{W_1} = \hat{E}_{\lambda_j}$. To show the other inclusion, take $w \in \hat{E}_{\lambda_j}$, Since $E_{\lambda_j} \subseteq W_1$, this implies that $w \in W_1$. Now since there exists k such that $(T - \lambda I)^k w = \vec{0}$, we find

$$(T_{W_1} - \lambda_j I)^k w = (T - \lambda_j I)^k w = \vec{0} \ ,$$

and we are done. We find

$$V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_{k+1}} \ .$$

□

□

6.3 Nilpotent operators

Now we look at the operator T on the generalized eigenspaces. We need only restrict T to each eigenspace to determine the action on all of V. So for this purpose we will assume that T has only one generalized eigenspace: there exists $\lambda \in F$ such that

$$V = \hat{E}_{\lambda} \ .$$

In other words, for each $v \in V$ there exists k such that $(T - \lambda I)^k v = \vec{0}$. Recall we can then argue that there exists k^* such that

$$V = N(T - \lambda I)^{k^*} \ ,$$

or, if $U = T - \lambda I$, $U^{k^*} = 0$.

Definition 6.3.1. *Let $U : V \to V$ be linear. We say that U is nilpotent if there exists k such that*

$$U^k = 0 \ .$$

The smallest k for which $U^k = 0$ is called the degree of U.

The point of this section will be to give a structure theorem for nilpotent operators. It can be seen as a special case of Jordan form when all eigenvalues are zero. To prove this structure theorem, we will look at the nullspaces of powers of U. Note that if $k = deg(U)$, then $N(U^k) = V$ but $N(U^{k-1}) \neq V$. We get then an increasing chain of subspaces

$$N_0 \subseteq N_1 \subseteq \cdots \subseteq N_k, \text{ where } N_j = N(U^j) \ .$$

- If $v \in N_j \setminus N_{j-1}$ then $U(v) \in N_{j-1} \setminus N_{j-2}$.

  Proof. v has the property that $U^j v = \vec{0}$ but $U^{j-1} v \neq \vec{0}$. Therefore $U^{j-1}(Uv) = \vec{0}$ but $U^{j-2}(Uv) \neq \vec{0}$. □

Definition 6.3.2. *If $W_1 \subseteq W_2$ are subspaces of V then we say that $v_1, \ldots, v_m \in W_2$ are linearly independent mod W_1 if*

$$a_1 v_1 + \cdots + a_m v_m \in W_1 \text{ implies } a_i = 0 \text{ for all } i \ .$$

Lemma 6.3.3. *Suppose that $dim\ W_2 - dim\ W_1 = l$ and $v_1, \ldots, v_m \in W_2 \setminus W_1$ are linearly independent mod W_1. Then*

1. *$m \leq l$ and*
2. *we can choose $l - m$ vectors $v_{m+1}, \ldots, v_l$ in $W_2 \setminus W_1$ such that $\{v_1, \ldots, v_l\}$ are linearly independent mod W_1.*

Proof. It suffices to show that we can add just one vector. Let $w_1, \ldots, w_t$ be a basis for W_1. Then define

$$X = Span(\{w_1, \ldots, w_t, v_1, \ldots, v_m\}) \ .$$

Then this set is linearly independent. Indeed, if

$$a_1 w_1 + \cdots + a_t w_t + b_1 v_1 + \cdots + b_m v_m = \vec{0} \ ,$$

then $b_1 v_1 + \cdots + b_m v_m \in W_1$, so all b_i's are zero. Then

$$a_1 w_1 + \cdots + a_t w_t = \vec{0} \ ,$$

so all a_i's are zero. Thus

$$t + m = dim\ X \leq dim\ W_2 = t + l$$

or $m \leq l$.

For the second part, if $k = l$, we are done. Otherwise $dim\ X < dim\ W_2$, so there exists $v_{k+1} \in W_2 \setminus X$. To show linear independence mod W_1, suppose that

$$a_1 v_1 + \cdots + a_k v_k + a_{k+1} v_{k+1} = w \in W_1 \ .$$

If $a_{k+1} = 0$ then we are done. Otherwise we can solve for v_{k+1} and see it is in X. This is a contradiction. □

Lemma 6.3.4. *Suppose that for some m, $v_1, \ldots, v_p \in N_m \setminus N_{m-1}$ are linearly independent mod N_{m-1}. Then $U(v_1), \ldots, U(v_p)$ are linearly independent mod N_{m-2}.*

Proof. Suppose that

$$a_1U(v_1) + \cdots + a_pU(v_p) = \vec{n} \in N_{m-2} \ .$$

Then

$$U(a_1v_1 + \cdots + a_pv_p) = \vec{n} \ .$$

Now

$$U^{m-1}(a_1v_1 + \cdots + a_pv_p) = U^{m-1}(\vec{n}) = \vec{0} \ .$$

Therefore $a_1v_1 + \cdots + a_pv_p \in N_{m-1}$. But these are linearly independent mod N_{m-1} so we find that $a_i = 0$ for all i. □

Now we do the following

1. Write $d_m = dim\ N_m - dim\ N_{m-1}$. Starting at the top, choose

   $$\beta_k = v_1^k, \ldots, v_{d_k}^k$$

   which are linearly independent mod N_{k-1}.

2. Move down one level: write $v_i^{k-1} = U(v_i^k)$. Then $\{v_1^{k-1}, \ldots, v_{d_k}^{k-1}\}$ is linearly independent mod N_{k-2}, so $d_k \leq d_{k-1}$. By the lemma we can extend this to

   $$\beta_{k-1} = \{v_1^{k-1}, \ldots, v_{d_k}^{k-1}, v_{d_k+1}^{k-1}, \ldots, v_{d_{k-1}}^{k-1}\} \ ,$$

   a maximal linearly independent set mod N_{k-2} in $N_{k-1} \setminus N_{k-2}$.

3. Repeat.

Note that $d_k + d_{k-1} + \cdots + d_1 = dim\ V$. We claim that if β_i is the set at level i then

$$\beta = \beta_1 \cup \cdots \cup \beta_k$$

is a basis for V. It suffices to show linear independence. For this, we use the following fact.

- If $W_1 \subseteq \cdots \subseteq W_k = V$ is a nested sequence of subspaces with $\beta_i \subseteq W_i \setminus W_{i-1}$ linearly independent mod W_{i-1} then $\beta = \cup_i \beta_i$ is linearly independent. (check).

We have shown the following result.

Definition 6.3.5. *A chain of length l for U is a set $\{v, U(v), U^2(v), \ldots, U^{l-1}(v)\}$ of non-zero vectors such that $U^l(v) = \vec{0}$.*

Theorem 6.3.6. *If $U : V \to V$ is linear and nilpotent (dim $V < \infty$) then there exists a basis of V consisting entirely of chains for U.*

Let $U : V \to V$ be nilpotent. If $C = \{U^{l-1}v, U^{l-2}v, \ldots, U(v), v\}$ is a chain then note that

$$\mathcal{C} = Span(C) \text{ is } U\text{-invariant} .$$

Since V has a basis of chains, say $C_1, \ldots, C_m$, we can write

$$V = \mathcal{C}_1 \oplus \cdots \oplus \mathcal{C}_m .$$

In our situation each C_i is a basis for $\mathcal{C}_i$ so our matrix for U is block diagonal. Each block corresponds to a chain. Then $U|_{\mathcal{C}_i}$ has the following matrix w.r.t. C_i:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & \\ & & & & \cdots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{pmatrix} .$$

Theorem 6.3.7 (Uniqueness of nilpotent form)**.** *Let $U : V \to V$ be linear and nilpotent with $\dim V < \infty$. Write*

$$l_i(\beta) = \#\ \textit{of (maximal) chains of length } i \textit{ in } \beta .$$

Then if β, β' are bases of V consisting of chains for U then

$$l_i(\beta) l_i(\beta') \textit{ for all } i .$$

Proof. Write $K_i(\beta)$ for the set of elements v of β such that $U^i(v) = \vec{0}$ but $U^{i-1}(v) \neq \vec{0}$ (for $i = 1$ we only require $U(v) = \vec{0}$). Let $\tilde{l}_i(\beta)$ be the number of (maximal) chains in β of length at least i.

We first claim that $\#K_i(\beta) = \tilde{l}_i(\beta)$ for all i. To see this note that for each chain C of length at least i there is a unique element $v \in C$ such that $v \in K_i(\beta)$. Conversely, for each $v \in K_i(\beta)$ there is a unique chain of length at least i containing v.

Let n_i be the dimension of $N(U^i)$. We claim that n_i equals the number $m_i(\beta)$ of $v \in \beta$ such that $U^i(v) = \vec{0}$. Indeed, the set of such v's is linearly independent and in $N(U^i)$ so $n_i \geq m_i(\beta)$. However all other v's ($\dim V - m_i(\beta)$ of them) are mapped to distinct elements of β (since β is made up of chains), so $\dim R(U^i) \geq \dim V - m_i(\beta)$, so $n_i \leq m_i(\beta)$.

Because $N(U^i)$ contains $N(U^{i-1})$ for all i (here we take $N(U^0) = \{\vec{0}\}$, we have

$$\tilde{l}_i(\beta) = \#K_i(\beta) = n_i - n_{i-1} .$$

Therefore

$$l_i(\beta) = \tilde{l}_i(\beta) - \tilde{l}_{i+1}(\beta) = (n_i - n_{i-1}) - (n_{i+1} - n_i) .$$

The right side does not depend on β and in fact the same argument shows it is equal to $l_i(\beta')$. This completes the proof.

□

6.4 Existence and uniqueness of Jordan form, Cayley-Hamilton

Definition 6.4.1. *A Jordan block for λ of size l is*

$$J_l^\lambda = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 \\ & & & & \cdots & \\ & & & & & 1 \\ & & & & & \lambda \end{pmatrix} .$$

Theorem 6.4.2 (Jordan canonical form)**.** *If $T : V \to V$ is linear with dim $V < \infty$ and F algebraically closed. Then there is a basis β of V such that $[T]_\beta$ is block diagonal with Jordan blocks.*

Proof. First decompose $V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_k}$. On each $\hat{E}_{\lambda_i}$, the operator $T - \lambda_i I$ is nilpotent. Each chain for $(T-\lambda_i I)|_{\hat{E}_{\lambda_i}}$ gives a block in nilpotent decomposition. Then $T = T - \lambda_i I + \lambda_i I$ gives a Jordan block. □

Draw a picture at this point (of sets of chains). We first decompose

$$V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_k}$$

and then

$$\hat{E}_{\lambda_i} = \mathcal{C}_1^i \oplus \cdots \oplus \mathcal{C}_{k_i}^i \ ,$$

where each $\mathcal{C}_j^i$ is the span of a chain of generalized eigenvectors: $C_j^i = \{v_1, \ldots, v_p\}$, where

$$T(v_1) = \lambda_i v_1, \ T(v_2) = \lambda v_2 + v_1, \ldots, \ T(v_p) = \lambda_i v_p + v_{p-1} \ .$$

Theorem 6.4.3 (Cayley-Hamilton)**.** *Let $T : V \to V$ be linear with dim $V < \infty$ and F algebraically closed. Then*

$$c(T) = 0 \ .$$

Remark. *In fact the theorem holds even if F is not algebraically closed by doing a field extension.*

Lemma 6.4.4. *If $U : V \to V$ is linear and nilpotent with dim $V = n < \infty$ then*

$$U^n = 0 \ .$$

Therefore if $T : V \to V$ is linear and $v \in \hat{E}_\lambda$ then

$$(T - \lambda I)^{dim \ \hat{E}_\lambda} v = \vec{0} \ .$$

Proof. Let β be a basis of chains for U. Then the length of the longest chain is n. □

Lemma 6.4.5. *If $T : V \to V$ is linear with dim $V < \infty$ and β is a basis such that $[T]_\beta^\beta$ is in Jordan form then for each eigenvalue λ, let S_λ be the basis vectors corresponding to blocks for λ. Then*

$$Span(S_\lambda) = \hat{E}_\lambda \text{ for each } \lambda \ .$$

Therefore if

$$c(x) = \prod_{i=1}^{k} (\lambda_i - x)^{n_i} \ ,$$

then $n_i = dim\ \hat{E}_{\lambda_i}$ for each i.

Proof. Write $\lambda_1, \ldots, \lambda_k$ for the distinct eigenvalues of T. Let

$$W_i = Span(S_{\lambda_i}) \ .$$

We may assume that the blocks corresponding to λ_1 appear first, λ_2 appear second, and so on. Since $[T]_\beta^\beta$ is in block form, this means V is a T-invariant direct sum

$$W_1 \oplus \cdots \oplus W_k \ .$$

However for each i, $T - \lambda_i I$ restricted to W_i is in nilpotent form. Thus $(T - \lambda_i I)^{dim\ \hat{E}_{\lambda_i}} v = \vec{0}$ for each $v \in S_{\lambda_i}$. This means

$$W_i \subseteq \hat{E}_{\lambda_i} \text{ for all } i \text{ or } dim\ W_i \leq dim\ \hat{E}_{\lambda_i} \ .$$

But $V = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_k}$, so $\sum_{i=1}^{k} dim\ \hat{E}_{\lambda_i} = dim\ V$. This gives that $dim\ W_i = dim \hat{E}_{\lambda_i}$ for all i, or $W_i = \hat{E}_{\lambda_i}$.

For the second claim, n_i is the number of times that λ_i appears on the diagonal; that is, the dimension of $Span(S_{\lambda_i})$. □

Proof of Cayley-Hamilton. We first factor

$$c(x) = \prod_{i=1}^{k} (\lambda_i - x)^{n_i} \ ,$$

where n_i is called the *algebraic multiplicity* of the eigenvalue λ_i. Let β be a basis such that $[T]_\beta^\beta$ is in Jordan form. If $v \in \beta$ is in a block corresponding to λ_j then $v \in \hat{E}_{\lambda_j}$ and so $(T - \lambda_j I)^{dim\ \hat{E}_{\lambda_j}} v = \vec{0}$ by the previous lemma). Now

$$c(T)v = \left(\prod_{i=1}^{k} (\lambda_i I - T)^{n_i} \right) v = \left(\prod_{i \neq j} (\lambda_i I - T)^{n_i} \right) (\lambda_j I - T)^{n_j} v = \vec{0}$$

since $n_j = dim\ \hat{E}_{\lambda_j}$. □

Finally we have uniqueness of Jordan form.

Theorem 6.4.6. *If β and β' are bases of V for which T is in Jordan form, then the matrices are the same up to permutation of blocks.*

Proof. First we note that the characteristic polynomial can be read off of the matrices and is invariant. This gives that the diagonal entries are the same, and the number of vectors corresponding to each eigenvalue is the same.

We see from the second lemma that if β_i and β_i' are the parts of the bases corresponding to blocks involving λ_i then

$$W_i := Span(\beta_i) = \hat{E}_{\lambda_i} = Span(\beta_i') =: W_i' \ .$$

Restricting T to W_i and to W_i' then gives the blocks for λ_i. But then β_i and β_i' are just bases of $\hat{E}_{\lambda_i}$ of chains for $T - \lambda_i I$. The number of chains of each length is the same, and this is the number of blocks of each size. □

6.5 Exercises

Notation:

1. If $\mathbb{F}$ is a field then we write $\mathbb{F}[X]$ for the set of polynomials with coefficients in $\mathbb{F}$.
2. If $P, Q \in \mathbb{F}[X]$ then we say that P divides Q and write $P|Q$ if there exists $S \in \mathbb{F}[X]$ such that $Q = SP$.
3. If $P \in \mathbb{F}[X]$ then we write the $\deg(P)$ for the largest k such that the coefficient of x^k in P is nonzero. We define the degree of the zero polynomial to be $-\infty$.
4. If $P \in \mathbb{F}[X]$ then we say that P is monic if the coefficient of x^n is 1, where $n =$deg(P).
5. For a complex number z, we denote the complex conjugate by $\bar{z}$, i.e. if $z = a + ib$, with $a, b \in \mathbb{R}$, then $\bar{z} = a - ib$.
6. If V is an F-vector space, recall the definition of $V \times V$: it is the F-vector space whose elements are
$$V \times V = \{(v_1, v_2) : v_1, v_2 \in V\} \ .$$
Vector addition is performed as $(v_1, v_2) + (v_3, v_4) = (v_1 + v_3, v_2 + v_4)$ and for $c \in F$, $c(v_1, v_2)$ is defined as (cv_1, cv_2).

Exercises

1. (a) Show that for $P, Q \in \mathbb{F}[X]$, one has $\deg(PQ) = \deg(P) + \deg(Q)$.
   (b) Show that for $P, D \in \mathbb{F}[X]$ such that D is nonzero, there exist $Q, R \in \mathbb{F}[X]$ such that $P = QD + R$ and $\deg(R) < \deg(D)$.
   Hint: Use induction on $\deg(P)$.

(c) Show that, for any $\lambda \in F$,

$$P(\lambda) = 0 \Leftrightarrow (X - \lambda)|P.$$

(d) Let $P \in \mathbb{F}[X]$ be of the form $p(X) = a(X - \lambda_1)^{n_1} \cdots (X - \lambda_k)^{n_k}$ for some $a, \lambda_1, \ldots, \lambda_k \in \mathbb{F}$ and natural numbers $n_1, \ldots, n_k$. Show that $Q \in \mathbb{F}[X]$ divides P if and only if $Q(X) = b(X - \lambda_1)^{m_1} \cdots (X - \lambda_k)^{m_k}$ for some $b \in \mathbb{F}$ and natural numbers m_i with $m_i \leq n_i$ (we allow $m_i = 0$).

(e) Assume that $\mathbb{F}$ is algebraically closed. Show that every $P \in \mathbb{F}[X]$ is of the form $a(X - \lambda_1)^{n_1} \ldots (X - \lambda_k)^{n_k}$ for some $a, \lambda_1, \ldots, \lambda_k \in \mathbb{F}$ and natural numbers $n_1, \ldots, n_k$ with $n_1 + \cdots + n_k = \deg(P)$. In this case we call the λ_i's the *roots* of P.

2. Let $\mathbb{F}$ be a field and suppose that P, Q are nonzero polynomials in $\mathbb{F}[X]$. Define the subset $\mathcal{S}$ of $\mathbb{F}[X]$ as follows:

$$\mathcal{S} = \{AP + BQ : A, B \in \mathbb{F}[X]\} .$$

(a) Let $D \in \mathcal{S}$ be of minimal degree. Show that D divides both P and Q.
Hint: Use part 2 of the previous problem.

(b) Show that if $S \in F[X]$ divides both P and Q then S divides D.

(c) Conclude that there exists a unique monic polynomial $D \in \mathbb{F}[X]$ satisfying the following conditions

i. D divides P and Q.

ii. If $T \in \mathbb{F}[X]$ divides both P and Q then T divides D.

(Such a polynomial is called the *greatest common divisor* (GCD) of P and Q.)

(d) Show that if $\mathbb{F}$ is algebraically closed and P and Q are polynomials in $\mathbb{F}[X]$ with no common root then there exist A and B in $\mathbb{F}[X]$ such that

$$AP + BQ = 1 .$$

3. Let F be any field, V be an F-vector space, $f \in L(V, V)$, and $W \subset V$ an f-invariant subspace.

(a) Let $p : V \to V/W$ denote the natural map defined in Homework 8, problem 5. Show that there exists an element of $L(V/W, V/W)$, which we will denote by $f|_{V/W}$, such that $p \circ f = f|_{V/W} \circ p$. It is customary to expresses this fact using the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{f} & V \\ \downarrow{\scriptstyle p} & & \downarrow{\scriptstyle p} \\ V/W & \xrightarrow{f|_{V/W}} & V/W \end{array}$$

(b) Let β' be a basis for W and β be a basis for V which contains β'. Show that the image of $\beta'' := \beta \setminus \beta'$ under p is a basis for V/W.

(c) Let β'' be a subset of V with the property that the restriction of p to β'' (which is a map of sets $\beta'' \to V/W$) is injective and its image is a basis for V/W. Show that β'' is a linearly-independent set. Show moreover that if β' is a basis for W, then $\beta := \beta' \sqcup \beta''$ is a basis for V.

(d) Let β, β', and β'' be as above. Show that

$$[f]^{\beta}_{\beta} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

with $A = [f|W]^{\beta'}_{\beta'}$ and $C = [f|_{V/W}]^{p(\beta'')}_{p(\beta'')}$.

4. **The minimal polynomial.** Let F be any field, V an F-vector space, and $f \in L(V, V)$.

(a) Consider the subset $S \subset F[X]$ defined by

$$S = \{P \in F[X] | P(f) = 0\}.$$

Show that S contains a nonzero element.

(b) Let $M_f \in S$ be a monic non-zero element of minimal degree. Show that M_f divides any other element of S. Conclude that M_f as defined is unique. It is called the *minimal polynomial* of f.

(c) Show that the roots of M_f are precisely the eigenvalues of f by completing the following steps.

i. Suppose that $r \in \mathbb{F}$ is such that $M_f(r) = 0$. Show that

$$M_f(X) = Q(X)(X - r)^k$$

for some positive integer k and $Q \in \mathbb{F}[X]$ such that $Q(r) \neq 0$. Prove also that $Q(f) \neq 0$.

ii. Show that if $r \in \mathbb{F}$ satisfies $M_f(r) = 0$ then $f - rI$ is not invertible and thus r is an eigenvalue of f.

iii. Conversely, let λ be an eigenvalue of f with eigenvector v. Show that if $P \in \mathbb{F}[X]$ then

$$P(f)v = P(\lambda)v \ .$$

Conclude that λ is a root of M_f.

(d) Assume that $\mathbb{F}$ is algebraically closed. For each eigenvalue λ of f, express $\mathrm{mult}_\lambda(M_f)$ in terms of the Jordan form of f.

(e) Assume that $\mathbb{F}$ is algebraically closed. Show that f is diagonalizable if and only if $\mathrm{mult}_\lambda(M_f) = 1$ for all eigenvalues λ of f.

(f) Assume that $\mathbb{F}$ is algebraically closed. Under which circumstances the does M_f equal the characteristic polynomial of f?

5. If $T : V \to V$ is linear and V is a finite-dimensional $\mathbb{F}$-vector space with $\mathbb{F}$ algebraically closed, we define the *algebraic multiplicity* of an eigenvalue λ to be $a(\lambda)$, the dimension of the generalized eigenspace $\hat{E}_\lambda$. The *geometric multiplicity* of λ is $g(\lambda)$, the dimension of the eigenspace E_λ. Finally, the index of λ is $i(\lambda)$, the length of the longest chain of generalized eigenvectors in $\hat{E}_\lambda$.

   Suppose that λ is an eigenvalue of T and $g = g(\lambda)$ and $i = i(\lambda)$ are given integers.

   (a) What is the minimal possible value for $a = a(\lambda)$?

   (b) What is the maximal possible for a?

   (c) Show that a can take any value between the answers for the above two questions.

   (d) What is the smallest dimension n of V for which there exist two linear transformations T and U from V to V with all of the following properties? (i) There exists $\lambda \in \mathbb{F}$ which is the only eigenvalue of either T or U, (ii) T and U are not similar transformations and (iii) the geometric multiplicity of λ for T equals that of U and similarly for the index.

6. Find the Jordan form for each of the following matrices over $\mathbb{C}$. Write the minimal polynomial and characteristic polynomial for each. To do this, first find the eigenvalues. Then, for each eigenvalue λ, find the dimensions of the nullspaces of $(A - \lambda I)^k$ for pertinent values of k (where A is the matrix in question). Use this information to deduce the block forms.

$$(a)\ \begin{pmatrix} -1 & 0 & 0 \\ 1 & 4 & -1 \\ -1 & 4 & 0 \end{pmatrix} \qquad (b)\ \begin{pmatrix} 2 & 3 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 2 \end{pmatrix} \qquad (c)\ \begin{pmatrix} 5 & 1 & 3 \\ 0 & 2 & 0 \\ -6 & -1 & -4 \end{pmatrix}$$

$$(d)\ \begin{pmatrix} 3 & 0 & 0 \\ 4 & 2 & 0 \\ 5 & 0 & 2 \end{pmatrix} \qquad (e)\ \begin{pmatrix} 2 & 3 & 2 \\ -1 & 4 & 2 \\ 0 & 1 & 3 \end{pmatrix} \qquad (f)\ \begin{pmatrix} 4 & 1 & 0 \\ -1 & 2 & 0 \\ 1 & 1 & 3 \end{pmatrix}$$

7. (a) The characteristic polynomial of the matrix

$$A = \begin{pmatrix} 7 & 1 & 2 & 2 \\ 1 & 4 & -1 & -1 \\ -2 & 1 & 5 & -1 \\ 1 & 1 & 2 & 8 \end{pmatrix}$$

   is $c(x) = (x - 6)^4$. Find an invertible matrix S such that $S^{-1}AS$ is in Jordan form.

   (b) Find all complex matrices in Jordan form with characteristic polynomial

$$c(x) = (i - x)^3(2 - x)^2 \ .$$

8. **Complexification of finite-dimensional real vector spaces.** Let V be an $\mathbb{R}$-vector space. Just as we can view $\mathbb{R}$ as a subset of $\mathbb{C}$ we will be able to view V as a subset of a $\mathbb{C}$-vector space. This will be useful because $\mathbb{C}$ is algebraically closed so we can, for example, use the theory of Jordan form in $V_{\mathbb{C}}$ and bring it back (in a certain form) to V. We will give two constructions of the complexification; the first is more elementary and the second is the standard construction you will see in algebra.

   We put $V_{\mathbb{C}} = V \times V$.

   (a) Right now, $V_{\mathbb{C}}$ is only an $\mathbb{R}$-vector space. We must define what it means to multiply vectors by complex scalars. For $z \in \mathbb{C}$, $z = a + ib$ with $a, b \in \mathbb{R}$, and $v = (v_r, v_i) \in V_{\mathbb{C}}$, we define the element $zv \in V_{\mathbb{C}}$ to be

   $$(av_r - bv_i, av_i + bv_r).$$

   Show that in this way, $V_{\mathbb{C}}$ becomes a $\mathbb{C}$-vector space. This is the *complexification* of V.

   (b) We now show how to view V as a "subset" of $V_{\mathbb{C}}$. Show that the map $\iota : V \to V_{\mathbb{C}}$ which maps v to $(v, 0)$ is injective and $\mathbb{R}$-linear. (Thus the set $\iota(V)$ is a copy of V sitting in $V_{\mathbb{C}}$.)

   (c) Show that $\dim_{\mathbb{C}}(V_{\mathbb{C}}) = \dim_{\mathbb{R}}(V)$. Conclude that $V_{\mathbb{C}}$ is equal to $\mathrm{span}_{\mathbb{C}}(\iota(V))$. Conclude further that if $v_1, \ldots, v_n$ is an $\mathbb{R}$-basis for V, then $\iota(v_1), \ldots, \iota(v_n)$ is a $\mathbb{C}$-basis for $V_{\mathbb{C}}$.

   (d) **Complex conjugation:** We define the *complex conjugation map* $c : V_{\mathbb{C}} \to V_{\mathbb{C}}$ to be the map $(v_r, v_i) \mapsto (v_r, -v_i)$. Just as $\mathbb{R}$ (sitting inside of $\mathbb{C}$) is invariant under complex conjugation, so will our copy of V (and its subspaces) be inside of $V_{\mathbb{C}}$.

      i. Prove that $c^2 = 1$ and $i(V) = \{v \in V_{\mathbb{C}} | c(v) = v\}$.

      ii. Show that for all $z \in \mathbb{C}$ and $v \in V_{\mathbb{C}}$, we have $c(zv) = \bar{z}c(v)$. Maps with this property are called *anti-linear*.

      iii. In the next two parts, we classify those subspaces of $V_{\mathbb{C}}$ that are invariant under c. Let W be a subspace of V. Show that the $\mathbb{C}$-subspace of $V_{\mathbb{C}}$ spanned by $\iota(W)$ equals

      $$\{(w_1, w_2) \in V_{\mathbb{C}} : w_1, w_2 \in W\}$$

      and is invariant under c.

      iv. Show conversely that if a subspace $\tilde{W}$ of the $\mathbb{C}$-vector space $V_{\mathbb{C}}$ is invariant under c, then there exists a subspace $W \subset V$ such that $\tilde{W} = \mathrm{span}_{\mathbb{C}}(\iota(W))$.

      v. Last, notice that the previous two parts told us the following: The subspaces of $V_{\mathbb{C}}$ which are invariant under conjugation are precisely those which are equal to $W_{\mathbb{C}}$ for subspaces W of the $\mathbb{R}$-vector space V. Show that moreover, in that situation, the restriction of the complex conjugation map $c : V_{\mathbb{C}} \to V_{\mathbb{C}}$ to $W_{\mathbb{C}}$ is equal to the complex conjugation map defined for $W_{\mathbb{C}}$ (the latter map is defined intrinsically for $W_{\mathbb{C}}$, i.e. without viewing it as a subspace of $V_{\mathbb{C}}$).

9. Let V be a finite dimensional $\mathbb{R}$-vector space. For this exercise we use the notation of the previous one.

   (a) Let W be another finite dimensional $\mathbb{R}$-vector space, and let $f \in L(V, W)$. Show that
   $$f_{\mathbb{C}}((v, w)) := (f(v), f(w))$$
   defines an element $f_{\mathbb{C}} \in L(V_{\mathbb{C}}, W_{\mathbb{C}})$.

   (b) Show that for $\underline{v} \in V_{\mathbb{C}}$, we have $f_{\mathbb{C}}(c(\underline{v})) = c(f_{\mathbb{C}}(\underline{v}))$. Show conversely that if $\tilde{f} \in L(V_{\mathbb{C}}, W_{\mathbb{C}})$ has the property that $\tilde{f}(c(\underline{v})) = c(\tilde{f}(\underline{v}))$, the $\tilde{f} = f_{\mathbb{C}}$ for some $f \in L(V, W)$.

10. In this problem we will establish the real Jordan form. Let V be a vector space over $\mathbb{R}$ of dimension $n < \infty$. Let $T : V \to V$ be linear and $T_{\mathbb{C}}$ its complexification.

   (a) If $\lambda \in \mathbb{C}$ is an eigenvalue of $T_{\mathbb{C}}$, and $\hat{E}_\lambda$ is the corresponding generalized eigenspace, show that
   $$c(\hat{E}_\lambda) = \hat{E}_{\bar{\lambda}}.$$

   (b) Show that the non-real eigenvalues of $T_{\mathbb{C}}$ come in pairs. In other words, show that we can list the distinct eigenvalues of $T_{\mathbb{C}}$ as
   $$\lambda_1, \ldots, \lambda_r, \sigma_1, \sigma_2, \ldots, \sigma_{2m} \ ,$$
   where for each $j = 1, \ldots, r$, $\overline{\lambda_j} = \lambda_j$ and for each $i = 1, \ldots, m$, $\sigma_{2i-1} = \overline{\sigma_{2i}}$.

   (c) Because $\mathbb{C}$ is algebraically closed, the proof of Jordan form shows that
   $$V_{\mathbb{C}} = \hat{E}_{\lambda_1} \oplus \cdots \oplus \hat{E}_{\lambda_r} \oplus \hat{E}_{\sigma_1} \oplus \cdots \oplus \hat{E}_{\sigma_{2m}} \ .$$
   Using the previous two points, show that for $j = 1, \ldots, r$ and $i = 1, \ldots, m$, the subspaces of $V_{\mathbb{C}}$
   $$\hat{E}_{\lambda_j} \qquad \text{and} \qquad \hat{E}_{\sigma_{2i-1}} \oplus \hat{E}_{\sigma_{2i}}$$
   are c-invariant.

   (d) Deduce from the results of problem 6, homework 10 that there exist subspaces $X_1, \ldots, X_r$ and $Y_1, \ldots, Y_m$ of V such that for each $j = 1, \ldots, r$ and $i = 1, \ldots, m$,
   $$\hat{E}_{\lambda_j} = Span_{\mathbb{C}}(\iota(X_j)) \text{ and } \hat{E}_{\sigma_{2i-1}} \oplus \hat{E}_{\sigma_{2i}} = Span_{\mathbb{C}}(\iota(Y_i)) \ .$$
   Show that
   $$V = X_1 \oplus \cdots \oplus X_r \oplus Y_1 \oplus \cdots \oplus Y_m \ .$$

   (e) Prove that for each $j = 1, \ldots, r$, the transformation $T - \lambda_j I$ restricted to X_j is nilpotent and thus we can find a basis β_j for X_j consisting entirely of chains for $T - \lambda_j I$.

(f) For each $i = 1, \ldots, m$, let

$$\beta_i = \{(v_1^i, w_1^i), \ldots, (v_{n_i}^i, w_{n_i}^i)\}$$

be a basis of $\hat{E}_{\sigma_{2i-1}}$ consisting of chains for $T_{\mathbb{C}} - \sigma_{2i-1} I$. Prove that

$$\hat{\beta}_i = \{v_1^i, w_1^i, \ldots, v_{n_i}^i, w_{n_i}^i\}$$

is a basis for Y_i. Describe the form of the matrix representation of T restricted to Y_i relative to $\hat{\beta}_i$.

(g) Gathering the previous parts, state and prove a version of Jordan form for linear transformations over finite-dimensional real vector spaces. Your version should be of the form "If $T : V \to V$ is linear then there exists a basis β of V such that $[T]_\beta^\beta$ has the form ..."

7 Bilinear forms

7.1 Definitions

We now switch gears from Jordan form.

Definition 7.1.1. *If V is a vector space over F, a function $f : V \times V \to F$ is called a bilinear form if for fixed $v \in V$, $f(v, w)$ is linear in w and for fixed $w \in V$, $f(v, w)$ is linear in v.*

Bilinear forms have matrix representations similar to those for linear transformations. Choose a basis $\beta = \{v_1, \ldots, v_n\}$ for V and write

$$v = a_1 v_1 + \cdots + a_n v_n, \ w = b_1 v_1 + \cdots + b_n v_n \ .$$

Now

$$f(v, w) = \sum_{i=1}^{n} a_i f(v_i, w) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j f(v_i, v_j) \ .$$

Define an $n \times n$ matrix A by $A_{i,j} = f(v_i, v_j)$. Then this is

$$\sum_{i=1}^{n} a_i \left(\sum_{j=1}^{n} b_j A_{i,j} \right) = \sum_{i=1}^{n} a_i \left(A \cdot \vec{b} \right)_i = (\vec{a})^t \, A \cdot \vec{b} = [v]_\beta^t [f]_\beta^\beta [w]_\beta \ .$$

We have proved:

Theorem 7.1.2. *If $\dim V < \infty$ and $f : V \times V \to F$ is a bilinear form there exists a unique matrix $[f]_\beta^\beta$ such that for all $v, w \in V$,*

$$f(v, w) = [v]_\beta^t [f]_\beta^\beta [w]_\beta \ .$$

Furthermore the map $f \mapsto [f]_\beta^\beta$ is an isomorphism from $Bil(V, F)$ to $M_{n \times n}(F)$.

Proof. We showed existence. To prove uniqueness, suppose that A is another such matrix. Then

$$A_{i,j} = e_i^t A e_j = [v_i]_\beta^t A [v_j]_\beta = f(v_i, v_j) \ .$$

□

If β' is another basis then

$$([v]_{\beta'})^t \left(\left([I]_\beta^{\beta'} \right)^t [f]_\beta^\beta [I]_\beta^{\beta'} \right) [w]_{\beta'} = \left([I]_\beta^{\beta'} [v]_{\beta'} \right)^t [f]_\beta^\beta [I]_\beta^{\beta'} [w]_{\beta'} = ([v]_\beta)^t \, [f]_\beta^\beta [w]_\beta = f(v, w) \ .$$

Therefore

$$[f]_{\beta'}^{\beta'} = \left([I]_\beta^{\beta'} \right)^t [f]_\beta^\beta [I]_\beta^{\beta'} \ .$$

Note that for fixed $v \in V$ the map $L_f(v) : V \to F$ given by $L_f(v)(w) = f(v, w)$ is a linear functional. So f gives a map in $\mathcal{L}(V, V^*)$.

Theorem 7.1.3. *Denote by $Bil(V,F)$ the set of bilinear forms on V. The map $\Phi_L : Bil(V,F) \to \mathcal{L}(V,V^*)$ given by*

$$\Phi_L(f) = L_f$$

is an isomorphism.

Proof. If $f, g \in Bil(V,F)$ and $c \in F$ then

$$\begin{aligned}(\Phi_L(cf+g)(v))(w) &= (L_{cf+g}(v))(w) = (cf+g)(v,w) = cf(v,w) + g(v,w) \\ &= cL_f(v)(w) + L_g(v)(w) = (cL_f(v) + L_g(v))(w) \\ &= (c\Phi_L(f)(v) + \Phi_L(g)(v))(w) \ .\end{aligned}$$

Thus $\Phi_L(cf+g)(v) = c\Phi_L(f)(v) + \Phi_L(g)(v)$. This is the same as $(c\Phi_L(f) + \Phi_L(g))(v)$. Therefore $\Phi_L(cf+g) = c\Phi_L(f) + \Phi_L(g)$. Thus Φ_L is linear.

Now $Bil(V,F)$ has dimension n^2. This is because the map from last theorem is an isomorphism. So does $\mathcal{L}(V,V^*)$. Therefore we only need to show one-to-one or onto. To show one-to-one, suppose that $\Phi_L(f) = 0$. Then for all v, $L_f(v) = 0$. In other words, for all v and $w \in V$, $f(v,w) = 0$. This means $f = 0$. □

Remark. *We can also define $R_f(w)$ by $R_f(w)(v) = f(v,w)$. Then the corresponding map Φ_R is an isomorphism.*

You will prove the following fact in homework. If β is a basis for V and β^* is the dual basis, then for each $f \in Bil(V,F)$,

$$[R_f]^\beta_{\beta^*} = [f]^\beta_\beta \ .$$

Then we have

$$[L_f]^\beta_{\beta^*} = \left([f]^\beta_\beta\right)^t \ .$$

To see this, set $g \in Bil(V,F)$ to be $g(v,w) = f(w,v)$. Then for each $v, w \in V$,

$$([w]_\beta)^t [f]^\beta_\beta [v]_\beta = f(w,v) = g(v,w)$$

Taking transpose on both sides,

$$([v]_\beta)^t \left([f]^\beta_\beta\right)^t [w]_\beta = g(v,w)$$

so $[g]^\beta_\beta = \left([f]^\beta_\beta\right)^t$. But $L_f = R_g$, so

$$\left([f]^\beta_\beta\right)^t = [R_g]^\beta_{\beta^*} = [L_f]^\beta_{\beta^*} \ .$$

Definition 7.1.4. *If $f \in Bil(V,F)$ then we define the rank of f to be the rank of R_f.*

- By the above remark, the rank equals the rank of either of the matrices

$$[f]^\beta_\beta \text{ or } [L_f]^\beta_{\beta^*} \ .$$

Therefore the rank of $[f]^\beta_\beta$ does not depend on the choice of basis.

- For $f \in Bil(V,F)$, define

$$N(f) = \{v \in V : f(v,w) = 0 \text{ for all } w \in V\} .$$

This is just

$$\{v \in V : L_f(v) = 0\} = N(L_f) .$$

But L_f is a map from V to V^*, so we have

$$rank(f) = rank(L_f) = dim\ V - dim\ N(f) .$$

7.2 Symmetric bilinear forms

Definition 7.2.1. *A bilinear form $f \in Bil(V,F)$ is called symmetric if $f(v,w) = f(w,v)$ for all $v, w \in V$. f is called skew-symmetric if $f(v,v) = 0$ for all $v \in V$.*

- The matrix for a symmetric bilinear form is symmetric and the matrix for a skew-symmetric bilinear form is anti-symmetric.
- Furthermore, each symmetric matrix A gives a symmetric bilinear form:

$$f(v,w) = ([v]_\beta)^t\, A \cdot [w]_\beta .$$

Similarly for skew-symmetric matrices.

Lemma 7.2.2. *If $f \in Bil(V,F)$ is symmetric and $char(F) \neq 2$ then $f(v,w) = 0$ for all $v, w \in V$ if and only if $f(v,v) = 0$ for all $v \in V$.*

Proof. One direction is clear. For the other, suppose that $f(v,v) = 0$ for all $v \in V$.

$$f(v+w, v+w) = f(v,v) + 2f(v,w) + f(w,w)$$

$$f(v-w, v-w) = f(v,v) - 2f(v,w) + f(w,w) .$$

Therefore

$$0 = f(v+w, v+w) - f(v-w, v-w) = 4f(v,w) .$$

Here $4f(v,w)$ means $f(v,w)$ added to itself 3 times. If $char(F) \neq 2$ then this implies $f(v,w) = 0$. □

Remark. *If $char(F) = 2$ then the above lemma is false. Take $F = \mathbb{Z}_2$ with $V = F^2$ and f with matrix*

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

Check that this has $f(v,v) = 0$ for all v but $f(v,w)$ is clearly not 0 for all $v, w \in V$.

Definition 7.2.3. *A basis $\beta = \{v_1, \ldots, v_n\}$ of V is orthogonal for $f \in Bil(V,F)$ if $f(v_i, v_j) = 0$ whenever $i \neq j$. It is orthonormal if it is orthogonal and $f(v_i, v_i) = 1$ for all i.*

Theorem 7.2.4 (Diagonalization of symmetric bilinear forms)**.** *Let $f \in Bil(V, F)$ with $char(F) \neq 2$ and $\dim V < \infty$. If f is symmetric then V has an orthogonal basis for f.*

Proof. We argue by induction on $n = \dim V$. If $n = 1$ it is clear. Let us now suppose that the statement holds for all $k < n$ for some $n > 1$ and show that it holds for n. If $f(v, v) = 0$ for all v then f is identically zero and thus we are done. Otherwise we can find some $v \neq 0$ such that $f(v, v) \neq 0$.

Define

$$W = \{w \in V : f(v, w) = 0\} .$$

Since this is the nullspace of $L_f(v)$ and $L_f(v)$ is a nonzero element of V^*, it follows that W is $n-1$ dimensional. Because f restricted to W is still a symmetric bilinear function, we can find a basis β' of W such that $[f]_{\beta'}^{\beta'}$ is diagonal. Write $\beta' = \{v_1, \ldots, v_{n-1}\}$ and $\beta = \beta' \cup \{v\}$. Then we claim β is a basis for V: if

$$a_1 v_1 + \cdots + a_{n-1} v_{n-1} + av = \vec{0} ,$$

then taking $L_f(v)$ of both sides we find $af(v, v) = \vec{0}$, or $a = 0$. Linear independence gives that the other a_i's are zero.

Now it is clear that $[f]_{\beta}^{\beta}$ is diagonal. For $i \neq j$ which are both $< n$ this follows because β' is a basis for which f is diagonal on W. Otherwise one of i, j is n and then the other vector is in W and so $f(v_i, v_j) = 0$.

□

- In the basis β, f has a diagonal matrix. This says that for each symmetric matrix A we can find an invertible matrix S such that

$$S^t AS \text{ is diagonal} .$$

- In fact, if F is a field such that each number has a square root (like $\mathbb{C}$) then we can make a new basis, replacing each element v of β such that $f(v, v) \neq 0$ by $v/\sqrt{f(v, v)}$ and leaving all elements such that $f(v, v) = 0$ to find a basis such that the representation of f is diagonal, with all 1 and 0 on the diagonal. The number of 1's equals the rank of f.

- Therefore if f has full rank and each element of F has a square root, there exists an orthonormal basis of V for f.

Theorem 7.2.5 (Sylvester's law)**.** *Let f be a symmetric bilinear form on $\mathbb{R}^n$. There exists a basis β such that $[f]_{\beta}^{\beta}$ is diagonal, with only 0's, 1's and -1's. Furthermore, the number of each is independent of the choice of basis that puts f into this form.*

Proof. Certainly such a basis exists. Just modify the construction by dividing by $\sqrt{|f(v_i, v_i)|}$ instead. So we show the other claim. Because the number of 0's is independent of the basis, we need only show the statement for 1's.

For a basis β, let $V^+(\beta)$ be the span of the v_i's such that $f(v_i, v_i) > 0$, and similarly for $V^-(\beta)$ and $V^0(\beta)$. Clearly

$$V = V^+(\beta) \oplus V^-(\beta) \oplus V^0(\beta) \ .$$

Note that the number of 1's equals the dimension of $V^+(\beta)$. Furthermore, for each $v \in V^+(\beta)$ we have

$$f(v, v) = \sum_{i=1}^{p} a_i^2 f(v_i, v_i) > 0 \ ,$$

where $v_1, \ldots, v_p$ are the basis vectors for $V^+(\beta)$. A similar argument gives $f(v, v) \leq 0$ for all $v \in V^-(\beta) \oplus V^0(\beta)$.

If β' is another basis we also have

$$V = V^+(\beta') \oplus V^-(\beta') \oplus V^0(\beta') \ .$$

Suppose that $dim\ V^+(\beta') > dim\ V^+(\beta)$. Then

$$dim\ V^+(\beta') + dim\ (V^-(\beta) \oplus V^0(\beta)) > n \ ,$$

so $V^+(\beta')$ intersects $V^-(\beta) \oplus V^0(\beta)$ in at least one non-zero vector, say v. Since $v \in V^+(\beta')$, $f(v, v) > 0$, However, since $v \in V^-(\beta) \oplus V^0(\beta)$, so $f(v, v) \leq 0$, a contradiction. Therefore $dim\ V^+(\beta) = dim\ V^+(\beta')$ and we are done. □

- The subspace $V^0(\beta)$ is unique. We can define

  $$N_L(f) = N(L_f), \ N_R(f) = N(R_f) \ .$$

  In the symmetric case, these are equal and we can define it to be be $N(f)$. We claim that

  $$V^0(\beta) = N(f) \text{ for all } \beta \ .$$

  Indeed, if $v \in V^0(\beta)$ then because the basis β is orthogonal for f, $f(v, \tilde{v}) = 0$ for all $\tilde{v} \in \beta$ and so $v \in N(f)$. On the other hand,

  $$dim\ V^0(\beta) = dim\ V - \left[dim\ V^+(\beta) + dim\ V^-(\beta)\right] = dim\ V - rank\ [f]_\beta^\beta = dim\ N(f) \ .$$

- However the spaces $V^+(\beta)$ and $V^-(\beta)$ are not unique. Let us take $f \in Bil(\mathbb{R}^2, \mathbb{R})$ with matrix (in the standard basis)

  $$[f]_\beta^\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ .$$

  Then $f((a, b), (c, d)) = ac - bd$. Take $v_1 = (2, \sqrt{3})$ and $v_2 = (\sqrt{3}, 2)$. Thus we get

  $$f(v_1, v_1) = (2)(2) - (\sqrt{3})(\sqrt{3}) = 1 \ ,$$

  $$f(v_1, v_2) = (2)(\sqrt{3}) - (\sqrt{3})(2) = 0$$

  $$f(v_2, v_2) = (\sqrt{3})(\sqrt{3}) - (2)(2) = -1 \ .$$

7.3 Sesquilinear and Hermitian forms

One important example of a symmetric bilinear form on $\mathbb{R}^n$ is

$$f(v,w) = v_1w_1 + \cdots + v_nw_n \ .$$

In this case, $\sqrt{f(v,v)}$ actually defines a good notion of length of vectors on $\mathbb{R}^n$ (we will define precisely what this means later). In particular, we have $f(v,v) \geq 0$ for all v. On $\mathbb{C}^n$, however, this is not true. If f is the bilinear form from above, then $f((i,\ldots,i),(i,\ldots,i)) < 0$. But if we define the form

$$f(v,w) = v_1\overline{w_1} + \cdots + v_n\overline{w_n} \ ,$$

then it is true. This is not bilinear, but it is sesquilinear.

Definition 7.3.1. *Let V be a finite dimensional complex vector space. A function $f : V \times V \to \mathbb{C}$ is called sesquilinear if*

1. *for each $w \in V$, the function $v \mapsto f(v,w)$ is linear and*
2. *for each $v \in V$, the function $w \mapsto f(v,w)$ is anti-linear.*

To be anti-linear, it means that $f(v, cw_1 + w_2) = \overline{c}f(v,w_1) + f(v,w_2)$. The sesquilinear form f is called Hermitian if $f(v,w) = \overline{f(w,v)}$.

- Note that if f is hermitian, then $f(v,v) = \overline{f(v,v)}$, so $f(v,v) \in \mathbb{R}$.
  1. If $f(v,v) \geq 0$ (> 0) for all $v \neq 0$ then f is positive semi-definite (positive definite).
  2. If $f(v,v) \leq 0$ (< 0) for all $v \neq 0$ then f is negative semi-definite (negative definite).
- If f is a sesquilinear form and β is a basis then there is a matrix for f: as before, if $v = a_1v_1 + \cdots + a_nv_n$ and $w = b_1v_1 + \cdots + b_nv_n$,

$$f(v,w) = \sum_{i=1}^{n} a_i f(v_i,w) = \sum_{i=1}^{n} \left[\overline{b_j} f(v_i,v_j)\right] = [v]_\beta^t [f]_\beta^\beta \overline{[w]_\beta} \ .$$

- The map $w \mapsto f(\cdot,w)$ is a *conjugate isomorphism* from V to V^*.
- We have the polarization formula

  $$4f(u,v) = f(u+v,u+v) - f(u-v,u-v) + if(u+iv,u+iv) - if(u-iv,u-iv) \ .$$

  From this we deduce that if $f(v,v) = 0$ for all v then $f = 0$.

Theorem 7.3.2 (Sylvester for Hermitian forms)**.** *Let f be a hermitian form on a finite-dimensional complex vector space V. There exists a basis β of V such that $[f]_\beta^\beta$ is diagonal with only 0's, 1's and -1's. Furthermore the number of each does not depend on β so long as the matrix is in diagonal form.*

Proof. Same proof. □

7.4 Exercises

Notation:

1. For all problems below, F is a field of characteristic different from 2, and V is a finite-dimensional F-vector space. We write $\mathrm{Bil}(V, F)$ for the vector space of bilinear forms on V, and $\mathrm{Sym}(V, F)$ for the subspace of symmetric bilinear forms.

2. If B is a bilinear form on V and $W \subset V$ is any subspace, we define the *restriction* of B to W, written $B|_W \in \mathrm{Bil}(W, F)$, by $B|_W(w_1, w_2) = B(w_1, w_2)$.

3. We call $B \in \mathrm{Sym}(V, F)$ *non-degenerate*, if $N(B) = 0$.

Exercises

1. Let $l \in V^*$. Define a symmetric bilinear form B on V by $B(v, w) = l(v)l(w)$. Compute the nullspace of B.

2. Let B be a symmetric bilinear form on V. Suppose that $W \subset V$ is a subspace with the property that $V = W \oplus N(B)$. Show that the $B|_W$ is non-degenerate.

3. Let B be a symmetric bilinear form on V and $char(\mathbb{F}) \neq 2$. Suppose that $W \subset V$ is a subspace such that $B|_W$ is non-degenerate. Show that then $V = W \oplus W^{\perp}$.
   Hint: Use induction on $\dim(W)$.

4. Recall the isomorphism $\Phi : \mathrm{Bil}(V, F) \to L(V, V^*)$ given by

$$\Phi(B)(v)(w) = B(v, w), \qquad v, w \in V.$$

   If β is a basis of V, and β^* is the dual basis of V^*, show that

$$[B]_{\beta}^{\beta} = \left([\Phi(B)]_{\beta^*}^{\beta}\right)^T.$$

5. Let n denote the dimension of V. Let $d \in \mathrm{Alt}^n(V)$, and $B \in \mathrm{Sym}(V, F)$ both be non-zero. We are going to show that there exists a constant $c_{d,B} \in F$ with the property that for any vectors $v_1, \ldots, v_n, w_1, \ldots, w_n \in V$, the following identity holds:

$$\det(B(v_i, w_j)_{i=1\,j=1}^{n\quad n}) = c_{d,B} d(v_1, \ldots, v_n) d(w_1, \ldots, w_n),$$

   by completing the following steps:

   (a) Show that for fixed $(v_1, \ldots, v_n)$, there exists a constant $c_{d,B}(v_1, \ldots, v_n) \in F$ such that

$$\det(B(v_i, w_j)_{i=1\,j=1}^{n\quad n}) = c_{d,B}(v_1, \ldots, v_n) \cdot d(w_1, \ldots, w_n).$$

(b) We now let $(v_1, \ldots, v_n)$ vary. Show that there exists a constant $c_{d,B} \in F$ such that

$$c_{d,B}(v_1, \ldots, v_n) = c_{d,B} \cdot d(v_1, \ldots, v_n).$$

Show further that $c_{d,B} = 0$ precisely when B is degenerate.

6. **The orthogonal group.** Let B be a non-degenerate symmetric bilinear form on V. Consider

$$O(B) = \{f \in L(V,V) | B(f(v), f(w)) = B(v,w)\ \forall v, w \in V\}.$$

(a) Show that if $f \in O(B)$, then $\det(f)$ is either 1 or -1
Hint: Use the previous exercise.

(b) Show that the composition of maps makes $O(B)$ into a group.

(c) Let $V = \mathbb{R}^2$, $B((x_1, x_2), (y_1, y_2)) = x_1y_1 + x_2y_2$. Give a formula for the 2x2-matrices that belong to $O(B)$.

7. **The vector product.** Assume that V is 3-dimensional. Let $B \in Sym(V, F)$ be non-degenerate, and $d \in \mathrm{Alt}^3(V)$ be non-zero.

(a) Show that for any $v, w \in V$ there exists a unique vector $z \in V$ such that for all vectors $x \in V$ the following identity holds: $B(z, x) = d(v, w, x)$.
Hint: Consider the element $d(v, w, \cdot) \in V^*$.

(b) We will denote the unique vector z from part 1 by $v \times w$. Show that $V \times V \to V$, $(v, w) \mapsto v \times w$ is bilinear and skew-symmetric.

(c) For $f \in O(B)$, show that $f(v \times w) = \det(f) \cdot (f(v) \times f(w))$.

(d) Show that $v \times w$ is B-orthogonal to both v and w.

(e) Show that $v \times w = 0$ precisely when v and w are linearly dependent.

8. Let V be a finite dimensional $\mathbb{R}$-vector space. Recall its complexification $V_{\mathbb{C}}$, defined in the exercises of last chapter. It is a $\mathbb{C}$-vector space with $\dim_{\mathbb{C}} V_{\mathbb{C}} = \dim_{\mathbb{R}} V$. As an $\mathbb{R}$-vector space, it equals $V \oplus V$. We have the injection $\iota : V \to V_{\mathbb{C}}$, $v \mapsto (v, 0)$. We also have the complex conjugation map $c(v, w) = (v, -w)$.

(a) Let B be a bilinear form on V. Show that

$$B_{\mathbb{C}}((v,w),(x,y)) := B(v,x) - B(w,y) + iB(v,y) + iB(w,x)$$

defines a bilinear form on $V_{\mathbb{C}}$. Show that $N(B_{\mathbb{C}}) = N(B)_{\mathbb{C}}$. Show that $B_{\mathbb{C}}$ is symmetric if and only if B is.

(b) Show that for $\underline{v}, \underline{w} \in V_{\mathbb{C}}$, we have $B_{\mathbb{C}}(c(\underline{v}), c(\underline{w})) = \overline{B_{\mathbb{C}}(\underline{v}, \underline{w})}$. Show conversely that any bilinear form $\tilde{B}$ on $V_{\mathbb{C}}$ with the property $\tilde{B}(c(\underline{v}), c(\underline{w})) = \overline{\tilde{B}(\underline{v}, \underline{w})}$ is equal to $B_{\mathbb{C}}$ for some bilinear form B on V.

(c) Let B be a symmetric bilinear form on V. Show that

$$B_H((v,w),(x,y)) := B(v,x) + B(w,y) - iB(v,y) + iB(w,x)$$

defines a Hermitian form on $V_{\mathbb{C}}$. Show that $N(B_H) = N(B)_{\mathbb{C}}$.

(d) Show that for $\underline{v}, \underline{w} \in V_{\mathbb{C}}$, we have $B_H(c(\underline{v}), c(\underline{w})) = \overline{B_H(\underline{v}, \underline{w})}$. Show conversely that any Hermitian form $\tilde{B}$ on $V_{\mathbb{C}}$ with the property $\tilde{B}(c(\underline{v}), c(\underline{w})) = \overline{\tilde{B}(\underline{v}, \underline{w})}$ is equal to B_H for some bilinear form B on V.

9. Prove that if V is a finite-dimensional $\mathbb{F}$-vector space with $char(F) \neq 2$ and f is a nonzero skew-symmetric bilinear form (that is, a bilinear form such that $f(v,w) = -f(w,v)$ for all $v, w \in V$) then there is no basis β for V such that $[f]^{\beta}_{\beta}$ is upper triangular.

10. For each of the following real matrices A, find an invertible matrix S such that $S^t AS$ is diagonal.

$$\begin{pmatrix} 2 & 3 & 5 \\ 3 & 7 & 11 \\ 5 & 11 & 13 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}.$$

Also find a complex matrix T such that $T^t AT$ is diagonal with only entries 0 and 1.

8 Inner product spaces

8.1 Definitions

We will be interested in positive definite hermitian forms.

Definition 8.1.1. *Let V be a complex vector space. A hermitian form f is called an inner product (or scalar product) if f is positive definite. In this case we call V a (complex) inner product space.*

An example is the standard dot product:

$$\langle u, v\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n} \ .$$

It is customary to write an inner product $f(u, v)$ as $\langle u, v\rangle$. In addition, we write $\|u\| = \sqrt{\langle u, u\rangle}$. This is the *norm* induced by the inner product $\langle \cdot, \cdot\rangle$. In fact, (V, d) is a metric space, using

$$d(u, v) = \|u - v\| \ .$$

Properties of norm. Let $(V, \langle \cdot, \cdot\rangle)$ be a complex inner product space.

1. For all $c \in \mathbb{C}$, $\|cu\| = |c|\|u\|$.
2. $\|u\| = 0$ if and only if $u = \vec{0}$.
3. (Cauchy-Schwarz inequality) For $u, v \in V$,

$$|\langle u, v\rangle| \leq \|u\|\|v\| \ .$$

Proof. Define If u or v is $\vec{0}$ then we are done. Otherwise, set $w = u - \frac{\langle u,v\rangle}{\|v\|^2}v$.

$$0 \leq \langle w, w\rangle = \langle w, u\rangle - \frac{\overline{\langle u, v\rangle}}{\|v\|^2}\langle w, v\rangle \ .$$

However

$$\langle w, v\rangle = \langle u, v\rangle - \frac{\langle u, v\rangle}{\|v\|^2}\langle v, v\rangle = 0 \ ,$$

so

$$0 \leq \langle w, u\rangle = \langle u, u\rangle - \frac{\langle u, v\rangle\overline{\langle u, v\rangle}}{\|v\|^2} \ ,$$

and

$$0 \leq \langle u, u\rangle - \frac{|\langle u, v\rangle|^2}{\|v\|^2} \Rightarrow |\langle u, v\rangle|^2 \leq \|u\|^2\|v\|^2 \ .$$

□

Everything above is an equality so, we have equality if and only if $w = \vec{0}$, or v and u are linearly dependent.

4. (Triangle inequality) For $u, v \in V$,

$$\|u + v\| \leq \|u\| + \|v\| \ .$$

This is also written $\|u - v\| \leq \|u - w\| + \|w - v\|$.

Proof.

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} + \langle v,v\rangle \\ &= \langle u,u\rangle + 2Re\ \langle u,v\rangle + \langle v,v\rangle \\ &\leq \|u\|^2 + 2\,|\langle u,v\rangle| + \|v\|^2 \\ &\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2 \ . \end{aligned}$$

Taking square roots gives the result. □

8.2 Orthogonality

Definition 8.2.1. *Given a complex inner product space $(V, \langle\cdot,\cdot\rangle)$ we say that vectors $u, v \in V$ are orthogonal if $\langle u, v\rangle = 0$.*

Theorem 8.2.2. *Let $v_1, \ldots, v_k$ be nonzero and pairwise orthogonal in a complex inner product space. Then they are linearly independent.*

Proof. Suppose that

$$a_1 v_1 + \cdots + a_k v_k = \vec{0} \ .$$

Then we take inner product with v_i.

$$0 = \langle \vec{0}, v_i\rangle = \sum_{j=1}^{k} a_j \langle v_j, v_i\rangle = a_i \|v_i\|^2 \ .$$

Therefore $a_i = 0$. □

We begin with a method to transform a linearly independent set into an orthonormal set.

Theorem 8.2.3 (Gram-Schmidt)**.** *Let V be a complex inner product space and $v_1, \ldots, v_k \in V$ which are linearly independent. There exist $u_1, \ldots, u_k$ such that*

1. *$\{u_1, \ldots, u_k\}$ is orthonormal and*
2. *for all $j = 1, \ldots, k$, $Span(\{u_1, \ldots, u_j\}) = Span(\{v_1, \ldots, v_j\})$.*

Proof. We will prove by induction. If $k = 1$, we must have $v_1 \neq \vec{0}$, so set $u_1 = v_1/\|v_1\|$. This gives $\|u_1\| = 1$ so that $\{u_1\}$ is orthonormal and certainly the second condition holds.

If $k \geq 2$ then assume the statement holds for all $j = k - 1$. Find vectors $u_1, \ldots, u_{k-1}$ as in the statement. Now to define u_k we set

$$w_k = v_k - [\langle v_k, u_1\rangle u_1 + \cdots + \langle v_k, u_{k-1}\rangle u_{k-1}] \ .$$

We claim that w_k is orthogonal to all u_j's and is not zero. To check the first, let $1 \leq j \leq k-1$ and compute

$$\begin{aligned} \langle w_k, u_j\rangle &= \langle v_k, u_j\rangle - [\langle v_k, u_1\rangle\langle u_1, u_j\rangle + \cdots + \langle v_k, u_{k-1}\rangle\langle u_{k-1}, u_j\rangle] \\ &= \langle v_k, u_j\rangle - \langle v_k, u_j\rangle\langle u_j, u_j\rangle = 0 \ . \end{aligned}$$

Second, if w_k were zero then we would have

$$v_k \in Span(\{u_1, \ldots, u_{k-1}\}) = Span(\{v_1, \ldots, v_{k-1}\}) \ ,$$

a contradiction to linear independence. Therefore we set $u_k = w_k/\|w_k\|$ and we see that $\{u_1, \ldots, u_k\}$ is orthonormal and therefore linearly independent.

Furthermore note that by induction,

$$Span(\{u_1, \ldots, u_k\}) \subseteq Span(\{u_1, \ldots, u_{k-1}, v_k\}) \subseteq Span(\{v_1, \ldots, v_k\}) \ .$$

Since the spaces on the left and right have the same dimension they are equal. □

Corollary 8.2.4. *If V is a finite-dimensional inner product space then V has an orthonormal basis.*

What do vectors look like represented in an orthonormal basis? Let $\beta = \{v_1, \ldots, v_n\}$ be a basis and let $v \in V$. Then

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

Taking inner product with v_j on both sides gives $a_j = \langle v, v_j\rangle$, so

$$v = \langle v, v_1\rangle v_1 + \cdots + \langle v, v_n\rangle v_n \ .$$

Thus we can view in this case (orthonormal) the number $\langle v, v_i\rangle$ as the projection of v onto v_i. We can then find the norm of v easily:

$$\begin{aligned} \|v\|^2 &= \langle v, v\rangle = \langle v, \sum_{i=1}^{n} \langle v, v_i\rangle v_i\rangle = \sum_{i=1}^{n} \overline{\langle v, v_i\rangle}\langle v, v_i\rangle \\ &= \sum_{i=1}^{n} |\langle v, v_i\rangle|^2 \ . \end{aligned}$$

This is known as **Parseval's identity**.

Definition 8.2.5. *If V is an inner product space and W is a subspace of V we define the orthogonal complement of V as*

$$W^\perp = \{v \in V : \langle v, w\rangle = 0 \text{ for all } w \in W\} \ .$$

- $\{\vec{0}\}^\perp = V$ and $V^\perp = \{\vec{0}\}$.
- If $S \subseteq V$ then $S^\perp$ is always a subspace of V (even if S was not). Furthermore,

$$S^\perp = (Span\ S)^\perp \text{ and } \left(S^\perp\right)^\perp = Span\ S \ .$$

Theorem 8.2.6. *Let V be a finite-dimensional inner product space with W a subspace. Then*

$$V = W \oplus W^\perp \ .$$

Proof. Let $\{w_1, \ldots, w_k\}$ be a basis for W and extend it to a basis $\{w_1, \ldots, w_n\}$ for V. Then perform Gram-Schmidt to get an orthonormal basis $\{v_1, \ldots, v_n\}$ such that

$$Span(\{v_1, \ldots, v_j\}) = Span(\{w_1, \ldots, w_j\}) \text{ for all } j = 1, \ldots, n \ .$$

In particular, $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W. We claim that $\{v_{k+1}, \ldots, v_n\}$ is a basis for $W^\perp$. To see this, define $\widetilde{W}$ to be the span of these vectors. Clearly $\widetilde{W} \subseteq W^\perp$. On the other hand,

$$\begin{aligned} W \cap W^\perp &= \{w \in W : \langle w, w'\rangle = 0 \text{ for all } w' \in W\} \\ &\subseteq \{w \in W : \langle w, w\rangle = 0\} = \{\vec{0}\} \ . \end{aligned}$$

This means that $dim\ W + dim\ W^\perp \leq n$, or $dim\ W^\perp \leq n - k$. Since $dim\ \widetilde{W} = n - k$ we see they are equal. □

This leads us to a definition.

Definition 8.2.7. *Let V be a finite dimensional inner product space. If W is a subspace of V we write $P_W : V \to V$ for the operator*

$$P_W(v) = w_1 \ ,$$

where v is written uniquely as $w_1 + w_2$ for $w_1 \in W$ and $w_2 \in W^\perp$. P_W is called the orthogonal projection onto W.

Properties of orthogonal projection.

1. P_W is linear.
2. $P_W^2 = P_W$.
3. $P_{W^\perp} = I - P_W$.

4. For all $v_1, v_2 \in V$,

$$\begin{aligned}\langle P_W(v_1), v_2\rangle &= \langle P_W(v_1), P_W(v_2)\rangle + \langle P_W(v_1), P_{W^\perp}(v_2)\rangle \\ &= \langle P_W(v_1), P_W(v_2)\rangle + \langle P_{W^\perp}(v_1), P_W(v_2)\rangle = \langle v_1, P_W(v_2)\rangle \ .\end{aligned}$$

Alternatively one may define an orthogonal projection as a linear map with properties 2 and 4. That is, if $T : V \to V$ is linear with $T^2 = I$ and $\langle T(v), w\rangle = \langle v, T(w)\rangle$ for all $v, w \in V$ then (check this)

- $V = R(T) \oplus N(T)$ where $N(T) = (R(T))^\perp$ and
- $T = P_{R(T)}$.

Example. Orthogonal projection onto a 1-d subspace. What we saw in the proof of $V = W \oplus W^\perp$ is the following. If W is a subspace of a finite dimensional inner product space, there exists an orthonormal basis of V of the form $\beta = \beta_1 \cup \beta_2$, where β_1 is an orthonormal basis of W and β_2 is an orthonormal basis of $W^\perp$.

Choose an orthonormal basis $\{v_1, \ldots, v_n\}$ of V so that v_1 spans W and the other vectors span $W^\perp$. For any $v \in V$,

$$v = \langle v, v_1\rangle v_1 + \langle v, v_2\rangle v_2 + \cdots + \langle v, v_n\rangle v_n \ ,$$

which is a representation of v in terms of W and $W^\perp$. Thus $P_W(v) = \langle v, v_1\rangle v_1$.

For any vector v' we can define the orthogonal projection onto v' as P_W, where $W = Span\{v'\}$. Then we choose $w' = v'/\|v'\|$ as our first vector in the orthonormal basis and

$$P_{v'}(v) = P_W(v) = \langle v, w'\rangle w' = \frac{\langle v, v'\rangle}{\|v'\|^2} v' \ .$$

Theorem 8.2.8. *If V is a finite dimensional inner product space with W a subspace of V, for each $v \in V$, $P_W(v)$ is the closest vector in W to v (using the distance coming from $\|\cdot\|$). That is, for all $w \in W$,*

$$\|v - P_W(v)\| \leq \|v - w\| \ .$$

Proof. First we note that if $w \in W$ and $w' \in W^\perp$ then the *Pythagoras theorem* holds:

$$\|w + w'\|^2 = \langle w, w'\rangle = \langle w, w\rangle + \langle w', w'\rangle = \|w\|^2 + \|w'\|^2 \ .$$

We now take $v \in V$ and $w \in W$ and write $v - w = P_W(v) + P_{W^\perp}(v) - w = P_W(v) - w + P_{W^\perp}(v)$. Applying Pythagoras,

$$\|v - w\|^2 = \|P_W(v) - w\|^2 + \|P_{W^\perp}(v)\|^2 \geq \|P_{W^\perp}(v)\|^2 = \|P_W(v) - v\|^2 \ .$$

□

8.3 Adjoints

Theorem 8.3.1. *Let V be a finite-dimensional inner product space. If $T : V \to V$ is linear then there exists a unique linear transformation $T^* : V \to V$ such that for all $v, u \in V$,*

$$\langle T(v), u\rangle = \langle v, T^*(u)\rangle \ . \tag{5}$$

We call T^* the adjoint of T.

Proof. We will use the Riesz representation theorem.

Lemma 8.3.2 (Riesz)**.** *Let V be a finite-dimensional inner product space. For each $f \in V^*$ there exists a unique $z \in V$ such that*

$$f(v) = \langle v, z\rangle \text{ for all } v \in V \ .$$

Given this we will define T^* as follows. For $u \in V$ we define the linear functional

$$f_{u,T} : V \to \mathbb{C} \text{ by } f_{u,T}(v) = \langle T(v), u\rangle \ .$$

You can check this is indeed a linear functional. By Riesz, there exists a unique $z \in V$ such that

$$f_{u,T}(v) = \langle v, z\rangle \text{ for all } v \in V \ .$$

We define this z to be $T^*(u)$. In other words, for a given $u \in V$, $T^*(u)$ is the unique vector in V with the property

$$\langle T(v), u\rangle = f_{T,u}(v) = \langle v, T^*(u)\rangle \text{ for all } v \in V \ .$$

Because of this identity, we see that there exists a *function* $T^* : V \to V$ such that for all $u, v \in V$, (5) holds. In other words, given T, we have a way of mapping a vector $u \in V$ to another vector which we call $T^*(u)$. We need to know that it is unique and linear.

Suppose that $R : V \to V$ is another function such that for all $u, v \in V$,

$$\langle T(v), u\rangle = \langle v, R(u)\rangle \ .$$

Then we see that

$$\langle v, T^*(u) - R(u)\rangle = \langle v, T^*(u)\rangle - \langle v, R(u)\rangle = \langle T(v), u\rangle - \langle v, R(u)\rangle = 0$$

for all $u, v \in V$. Choosing $v = T^*(u) - R(u)$ gives that

$$\|T^*(u) - R(u)\| = 0 \ ,$$

or $T^*(u) = R(u)$ for all u. This means $T^* = R$.

To show linearity, let $c \in \mathbb{C}$ and $u_1, u_2, v \in V$.

$$\begin{aligned} \langle T(v), cu_1 + u_2\rangle &= \overline{c}\langle T(v), u_1\rangle + \langle T(v), u_2\rangle = \overline{c}\langle v, T^*(u_1)\rangle + \langle v, T^*(u_2)\rangle \\ &= \langle v, cT^*(u_1) + T^*(u_2)\rangle \ . \end{aligned}$$

This means that

$$\langle v, T^*(cu_1 + u_2) - cT^*(u_1) - T^*(u_2)\rangle = 0$$

for all $v \in V$. Choosing $v = T^*(cu_1 + u_2) - cT^*(u_1) - T^*(u_2)$ give that

$$T^*(cu_1 + u_2) = cT^*(u_1) + T^*(u_2) \ ,$$

or T^* is linear. □

Properties of adjoint.

1. $T^* : V \to V$ is linear. To see this, if $w_1, w_2 \in V$ and $c \in F$ then for all v,

$$\begin{aligned} \langle T(v), cw_1 + w_2\rangle &= \overline{c}\langle T(v), w_1\rangle + \langle T(v), w_2\rangle \\ &= \overline{c}\langle v, T^*(w_1)\rangle + \langle v, T^*(w_2)\rangle \\ &= \langle v, cT^*(w_1) + T^*(w_2)\rangle \ . \end{aligned}$$

   By uniqueness, $T^*(cw_1 + w_2) = cT^*(w_1) + T^*(w_2)$.

2. If β is an orthonormal basis of V then $[T^*]^\beta_\beta = \overline{\left([T]^\beta_\beta\right)^t}$.

   Proof. If β is an orthonormal basis, remembering that $\langle\cdot,\cdot\rangle$ is a sesquilinear form, the matrix of this in the basis β is simply the identity. Therefore

$$\langle T(v), w\rangle = [T(v)]^t_\beta \overline{[w]_\beta} = \left([T]^\beta_\beta [v]_\beta\right)^t \overline{[w]_\beta} = [v]^t_\beta \left([T]^\beta_\beta\right)^t \overline{[w]_\beta} \ .$$

   On the other hand, for all v, w

$$\langle v, T^*(w)\rangle = [v]^t_\beta \overline{[T^*(w)]_\beta} = [v]^t_\beta \overline{[T^*]^\beta_\beta [w]_\beta} \ .$$

   Therefore

$$[v]^t_\beta \overline{[T^*]^\beta_\beta} \overline{[w]_\beta} = [v]^t_\beta \left([T]^\beta_\beta\right)^t \overline{[w]_\beta} \ .$$

   Choosing $v = v_i$ and $w = v_j$ tells that all the entries of the two matrices are equal. □

3. $(T + S)^* = T^* + S^*$.

   Proof. If $v, w \in V$,

$$\langle (T + S)(v), w\rangle = \langle T(v), w\rangle + \langle S(v), w\rangle = \langle v, T^*(w)\rangle + \langle v, S^*(w)\rangle \ .$$

   This equals $\langle v, (T^* + S^*)(w)\rangle$. □

4. $(cT)^* = \overline{c}T^*$. This is similar.

5. $(TS)^* = S^*T^*$.

   Proof. For all $v, w \in V$,

   $$\langle (TS)(v), w\rangle = \langle T(S(v)), w\rangle = \langle S(v), T^*(w)\rangle = \langle v, S^*(T^*(w))\rangle \ .$$

   This is $\langle v, (S^*T^*)(w)\rangle$. □

6. $(T^*)^* = T$.

   Proof. If $v, w \in V$,

   $$\langle T^*(v), w\rangle = \overline{\langle w, T^*(v)\rangle} = \overline{\langle T(w), v\rangle} = \langle v, T(w)\rangle \ .$$

   □

8.4 Spectral theory of self-adjoint operators

Definition 8.4.1. *If V is an inner product space and $T : V \to V$ is linear we say that T*

1. *self-adjoint if $T = T^*$;*
2. *skew-adjoint if $T = -T^*$;*
3. *unitary if T is invertible and $T^{-1} = T^*$;*
4. *normal if $TT^* = T^*T$.*

Note all the above operators are normal. Also orthogonal projections are self-adjoint. Draw relation to complex numbers (SA is real, skew is purely imaginary, unitary is unit circle).

Theorem 8.4.2. *Let V be an inner product space and $T : V \to V$ linear with λ an eigenvalue of T.*

1. *If T is self-adjoint then λ is real.*
2. *If T is skew-adjoint then λ is purely imaginary.*
3. *If T is unitary then $|\lambda| = 1$ and $|det\ T| = 1$.*

Proof. Let $T : V \to V$ be linear and λ and eigenvalue for eigenvector v.

1. Suppose $T = T^*$. Then

   $$\langle \lambda \|v\|^2 = \langle T(v), v\rangle = \langle v, T(v)\rangle = \overline{\lambda}\|v\|^2 \ .$$

   But $v \neq 0$ so $\lambda = \overline{\lambda}$.

2. Suppose that $T = -T^*$. Define $S = iT$. Then $S^* = \bar{i}T^* = -iT^* = iT = S$, so S is self-adjoint. Now $i\lambda$ is an eigenvalue of S:

$$S(v) = (iT)(v) = i\lambda v \ .$$

This means $i\lambda$ is real, or λ is purely imaginary.

3. Suppose $T^* = T^{-1}$. Then

$$|\lambda|^2\|v\|^2 = \langle T(v), T(v)\rangle = \langle v, T^{-1}Tv\rangle = \|v\|^2 \ .$$

This means $|\lambda| = 1$. Furthermore, *det* T is the product of eigenvalues, so $|det\ T| = 1$.

□

What do these operators look like? If β is an orthonormal basis for V then

1. If $T = T^*$ then $[T]^\beta_\beta = \overline{\left([T]^\beta_\beta\right)^t}$.

2. If $T = -T^*$ then $[T]^\beta_\beta = -\overline{\left([T]^\beta_\beta\right)^t}$.

Lemma 8.4.3. *Let V be a finite-dimensional inner product space. If $T : V \to V$ is linear, the following are equivalent.*

1. *T is unitary.*
2. *For all $v \in V$, $\|T(v)\| = \|v\|$.*
3. *For all $v, w \in V$, $\langle T(v), T(w)\rangle = \langle v, w\rangle$.*

Proof. If T is unitary then $\langle T(v), T(v)\rangle = \langle v, T^{-1}Tv\rangle = \langle v, v\rangle$. This shows 1 implies 2. If T preserves norm then it also preserves inner product by the polarization identity. This proves 2 implies 3. To see that 3 implies 1, we take $v, w \in V$ and see

$$\langle v, w\rangle = \langle T(v), T(w)\rangle = \langle v, T^*T(w)\rangle \ .$$

This implies that $\langle v, w - T^*T(w)\rangle = 0$ for all v. Taking $v = w - T^*T(w)$ gives that $T^*T(w) = w$. Thus T must be invertible and $T^* = T^{-1}$. □

- Furthermore, T is unitary if and only if T maps orthonormal bases to orthonormal bases. In particular, $[T]^\beta_\beta$ has orthonormal columns whenever β is orthonormal.
- For β an orthonormal basis, the unitary operators are exactly those whose matrices relative to β have orthonormal columns.

We begin with a definition.

Definition 8.4.4. *If V is a finite-dimensional inner product space and $T : V \to V$ is linear, we say that T is unitarily diagonalizable if there exists an orthonormal basis β of V such that $[T]_\beta^\beta$ is diagonal.*

Note that T is unitarily diagonalizable if and only if there exists a unitary operator U such that

$$U^{-1}TU \text{ is diagonal .}$$

Theorem 8.4.5 (Spectral theorem)**.** *Let V be a finite-dimensional inner product space. If $T : V \to V$ is self-adjoint then T is unitarily diagonalizable.*

Proof. We will use induction on $dim\ V = n$. If $n = 1$ just choose a vector of norm 1. Otherwise suppose the statement is true for $n < k$ and we will show it for $k \geq 2$. Since T has an eigenvalue λ, it has an eigenvector, v_1. Choose v_1 with norm 1.

Let $U_\lambda = T - \lambda I$. We claim that

$$V = N(U_\lambda) \oplus R(U_\lambda) \ .$$

To show this we need only prove that $R(U_\lambda) = N(U_\lambda)^\perp$. This will follow from a lemma:

Lemma 8.4.6. *Let V be a finite-dimensional inner product space and $U : V \to V$ linear. Then*

$$R(U) = N(U^*)^\perp \ .$$

Proof. If $w \in R(U)$ then let $z \in N(U^*)$. There exists $v \in V$ such that $U(v) = w$. Therefore

$$\langle w, z\rangle = \langle U(v), z\rangle = \langle v, U^*(z)\rangle = 0 \ .$$

Therefore $R(U) \subseteq N(U^*)^\perp$. For the other containment, note that $dim\ R(U) = dim\ R(U^*)$ (since the matrix of U^* in an orthonormal basis is just the conjugate transpose of that of U). Therefore

$$dim\ R(U) = dim\ R(U^*) = dim\ V - dim\ N(U^*) = dim\ N(U^*)^\perp \ .$$

□

Now we apply the lemma. Note that since T is self-adjoint,

$$U_\lambda^* = (T - \lambda I)^* = T^* - \overline{\lambda} I = T - \lambda I = U_\lambda \ ,$$

since $\lambda \in \mathbb{R}$. Thus using the lemma with U_λ,

$$V = N(U_\lambda^*) \oplus N(U_\lambda^*)^\perp = N(U_\lambda) \oplus R(U_\lambda) \ .$$

Note that these are T-invariant spaces and $dim\ R(U_\lambda) < k$ since λ is an eigenvalue of T. Thus there is an orthonormal basis β' of $R(U_\lambda)$ such that T restricted to this space is diagonal. Taking

$$\beta = \beta' \cup \{v_1\}$$

gives now an orthonormal basis such that $[T]_\beta^\beta$ is block diagonal. But the block for $\{v_1\}$ is of size 1, so $[T]_\beta^\beta$ is diagonal. □

- Note that if T is skew-adjoint, iT is self-adjoint, so we can find an orthonormal basis β such that $[iT]_\beta^\beta$ is diagonal. This implies that T itself can be diagonalized by β: its matrix is just $-i[iT]_\beta^\beta$.

8.5 Normal and commuting operators

Lemma 8.5.1. *Let $U, T : V \to V$ be linear and F be algebraically closed. Write $\hat{E}^T_{\lambda_1}, \ldots, \hat{E}^T_{\lambda_k}$ for the generalized eigenspaces for T. If T and U commute then*

$$V = \hat{E}^T_{\lambda_1} \oplus \cdots \oplus \hat{E}^T_{\lambda_k}$$

is both a T-invariant direct sum and a U-invariant direct sum.

Proof. We need only show that the generalized eigenspaces of T are U-invariant. If $v \in N(T - \lambda_i I)^m$ then

$$(T - \lambda_i I)^m (U(v)) = U(T - \lambda_i I)^m v = \vec{0} \, .$$

□

Theorem 8.5.2. *Let $U, T : V \to V$ be linear and F algebraically closed. Suppose that T and U commute. Then*

1. *If T and U are diagonalizable then there exists a basis β such that both $[T]_\beta^\beta$ and $[U]_\beta^\beta$ are diagonal.*
2. *If V is an inner product space and T and U are self-adjoint then we can choose β to be orthonormal.*

Proof. Suppose that T and U are diagonalizable. Then the direct sum of eigenspaces for T is simply

$$E^T_{\lambda_1} \oplus \cdots \oplus E^T_{\lambda_k} \, ,$$

the eigenspaces. For each j, choose a Jordan basis β_j for U on $E^T_{\lambda_j}$. Set $\beta = \cup_{j=1}^k \beta_j$. These are all eigenvectors for T so $[T]_\beta^\beta$ is diagonal. Further, $[U]_\beta^\beta$ is in Jordan form. But since U is diagonalizable, its Jordan form is diagonal. By uniqueness, $[U]_\beta^\beta$ is diagonal.

If T and U are self-adjoint the decomposition

$$V = E^T_{\lambda_1} \oplus \cdots \oplus E^T_{\lambda_k}$$

is orthogonal. For each j, choose an orthonormal basis β_j of $E^T_{\lambda_j}$ of eigenvectors for U (since U is self-adjoint on $E^T_{\lambda_j}$). Now $\beta = \cup_{i=1}^k \beta_i$ is an orthonormal basis of eigenvectors for both T and U. □

Theorem 8.5.3. *Let V be a finite-dimensional inner product space. If $T : V \to V$ is linear then T is normal if and only if T is unitarily diagonalizable.*

Definition 8.5.4. *If V is an inner product space and $T : V \to V$ is linear, we write*

$$T = (1/2)(T + T^*) + (1/2)(T - T^*) = T_1 + T_2$$

and call these operators the self-adjoint part and the skew-adjoint part of T respectively.

Of course each part of a linear transformation can be unitarily diagonalized on its own. We now need to diagonalize them simultaneously.

Proof. If T is unitarily diagonalizable, then, taking U such that $T = U^{-1}DU$ for a diagonal D, we get

$$TT^* = U^{-1}DU\left(U^{-1}DU\right)^* = U^*DUU^*D^*U = U^*DD^*U = \cdots = T^*T \ ,$$

or T is normal.

Suppose that T is normal. Then T_1 and T_2 commute. Note that T_1 is self-adjoint and iT_2 is also. They commute so we can find an orthonormal basis β such that $[T_1]^\beta_\beta$ and $[iT_2]^\beta_\beta$ are diagonal. Now

$$[T]^\beta_\beta = [T_1]^\beta_\beta - i[iT_2]^\beta_\beta$$

is diagonal. □

8.6 Exercises

Notation

1. If V is a vector space over $\mathbb{R}$ and $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a positive-definite symmetric bilinear form, then we call $\langle \cdot, \cdot \rangle$ a (real) inner product. The pair $(V, \langle \cdot, \cdot \rangle)$ is called a (real) inner product space.

2. If $(V, \langle \cdot, \cdot \rangle)$ is a real inner product space and S is a subset of V we say that S is orthogonal if $\langle v, w \rangle = 0$ whenever $v, w \in S$ are distinct. We say S is orthonormal if S is orthogonal and $\langle v, v \rangle = 1$ for all $v \in S$.

3. If f is a symmetric bilinear form on a vector space V the *orthogonal group* is the set

$$O(f) = \{T : V \to V \mid f(T(u), T(v)) = f(u, v) \text{ for all } u, v \in V\} \ .$$

Exercises

1. Let V be a complex inner product space. Let $T \in L(V, V)$ be such that $T^* = -T$. We call such T skew-self-adjoint. Show that the eigenvalues of T are purely imaginary. Show further that V is the orthogonal direct sum of the eigenspaces of T. In other words, V is a direct sum of the eigenspaces and $\langle v, w \rangle = 0$ if v and w are in distinct eigenspaces.
   Hint: Construct from T a suitable self-adjoint operator and apply the known results from the lecture to that operator.

2. Let V be a complex inner product space, and $T \in L(V, V)$.

   (a) Show that T is unitary if and only if it maps orthonormal bases to orthonormal bases.

   (b) Let β be an orthonormal basis of V. Show that T is unitary if and only if the columns of the matrix matrix $[T]_\beta^\beta$ form a set of orthonormal vectors in $\mathbb{C}^n$ with respect to the standard hermitian form (standard dot product).

3. Let $(V, \langle \cdot, \cdot \rangle)$ be a complex inner product space, and η be a Hermitian form on V (in addition to $\langle \cdot, \cdot \rangle$). Show that there exists an orthonormal basis of V such that $[\eta]_\beta^\beta$ is diagonal, by completing the following steps:

   (a) Show that for each $w \in V$, there exists a unique vector, which we call Aw, in V with the property that for all $v \in V$,

   $$\eta(v, w) = \langle v, Aw \rangle.$$

   (b) Show the the map $A : V \to V$ which sends a vector $w \in V$ to the vector Aw just defined, is linear and self-adjoint.

   (c) Use the spectral theorem for self-adjoint operators to complete the problem.

4. Let $(V, \langle \cdot, \cdot \rangle)$ be a real inner product space.

   (a) Define $\| \cdot \| : V \to \mathbb{R}$ by

   $$\|v\| = \sqrt{\langle v, v \rangle} \ .$$

   Show that for all $v, w \in V$,

   $$|\langle v, w \rangle| \leq \|v\| \|w\| \ .$$

   (b) Show that $\| \cdot \|$ is a norm on V.

   (c) Show that there exists an orthonormal basis β of V.

5. Let $(V, \langle \cdot, \cdot \rangle)$ be a real inner product space and $T : V \to V$ be linear.

   (a) Prove that for each $f \in V^*$ there exists a unique $z \in V$ such that for all $v \in V$,

   $$f(v) = \langle v, z \rangle \ .$$

   (b) For each $u \in V$ define $f_{u,T} : V \to V$ by

   $$f_{u,T}(v) = \langle T(v), u \rangle \ .$$

   Prove that $f_{u,T} \in V^*$. Define $T^t(u)$ to be the unique $u \in V$ such that for all $v \in V$,

   $$\langle T(v), u \rangle = \langle v, T^t(u) \rangle$$

   and show that T^t is linear.

(c) Show that if β is an orthonormal basis for V then

$$[T^t]_\beta^\beta = \left([T]_\beta^\beta\right)^t \ .$$

6. Let $(V, \langle\cdot,\cdot\rangle)$ be a real inner product space and define the complexification of $\langle\cdot,\cdot\rangle$ as in homework 11 by

$$\langle (v,w),(x,y)\rangle_{\mathbb{C}} = \langle v,x\rangle + \langle w,y\rangle - i\langle v,y\rangle + i\langle w,x\rangle \ .$$

(a) Show that $\langle\cdot,\cdot\rangle_{\mathbb{C}}$ is an inner product on $V_{\mathbb{C}}$.

(b) Let $T : V \to V$ be linear.

i. Prove that $(T_{\mathbb{C}})^* = (T^t)_{\mathbb{C}}$.

ii. If $T^t = T$ then we say that T is *symmetric*. Show in this case that $T_{\mathbb{C}}$ is Hermitian.

iii. It $T^t = -T$ then we say T is *anti-symmetric*. Show in this case that $T_{\mathbb{C}}$ is skew-adjoint.

iv. If T is invertible and $T^t = T^{-1}$ then we say that T is *orthogonal*. Show in this case that $T_{\mathbb{C}}$ is unitary. Show that this is equivalent to

$$T \in O(\langle\cdot,\cdot\rangle) \ ,$$

where $O(\langle\cdot,\cdot\rangle)$ is the orthogonal group for $\langle\cdot,\cdot\rangle$.

7. Let $(V, \langle\cdot,\cdot\rangle)$ be a real inner product space and $T : V \to V$ be linear.

(a) Suppose that $TT^t = T^tT$. Show then that $T_{\mathbb{C}}$ is normal. In this case, we can find a basis β of $V_{\mathbb{C}}$ such that β is orthonormal (with respect to $\langle\cdot,\cdot\rangle_{\mathbb{C}}$) and $[T_{\mathbb{C}}]_\beta^\beta$ is diagonal. Define the subspaces of V

$$X_1, \ldots, X_r, Y_1, \ldots, Y_{2m}$$

as in problem 1, question 3. Show that these are mutually orthogonal; that is, if v, w are in different subspaces then $\langle v, w\rangle = 0$.

(b) If T is symmetric then show that there exists an orthonormal basis β of V such that $[T]_\beta^\beta$ is diagonal.

(c) If T is skew-symmetric, what is the form of the matrix of T in real Jordan form?

(d) $A \in M_{2\times 2}(\mathbb{R})$ is called a *rotation matrix* if there exists $\theta \in [0, 2\pi)$ such that

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \ .$$

If T is orthogonal, show that there exists a basis β of V such that $[T]_\beta^\beta$ is block diagonal, and the blocks are either 2×2 rotation matrices or 1×1 matrices consisting of 1 or -1.

Hint. Use the real Jordan form.