Linear Algebra
Supplemental Lecture Notes for MTH 414
Fall 2001

Ulrich Meierfrankenfeld

November 10, 2001

1 Introduction

These notes are about the material I covered in class but can not be found in the text book [Axler].

2 Fields

Let S be a set. A binary operation on S is a function

$$* : S \times S \to S, \quad (a, b) \to a * b$$

For example both addition and multiplication are binary operations on $\mathbb{R}$, (the set of real numbers).

You are probably used to various properties of addition and multiplication of real and complex numbers: commutative, associative and distributive. For most theorems in Linear Algebra it is just these properties which are important. For this reason we will (in contrast to the text book) consider vector spaces over arbitrary *fields* and not only over $\mathbb{R}$ and $\mathbb{C}$.

Definition 2.1 *A* field *is a tuple* $(\mathbb{F}, +, \cdot)$*, where both* $+$ *and* $\cdot$ *are binary operation on* $\mathbb{F}$ *such that*

(aa) $a + b = b + a$ *(additive Commutative Law)*

for all $a, b \in \mathbb{F}$

(ab) $(a + b) + c = a + (b + c)$ *(additive Associative Law)*

for all $a, b, c \in \mathbb{F}$.

(ac) *There exists an element* $0_{\mathbb{F}}$ *in* $\mathbb{F}$ *such that*

$0_{\mathbb{F}} + a = a = a + 0_{\mathbb{F}}$ *(additive Identity)*

for all $a \in \mathbb{F}$

(ad) *For all* $a \in \mathbb{F}$ *there exists an element* $-a \in F$ *so that*
$$a + (-a) = 0_{\mathbb{F}} = (-a) + a \qquad \textit{(additive Inverse)}$$

(ba) $$a \cdot b = b \cdot a \qquad \textit{(multiplicative Commutative Law)}$$
for all $a, b \in \mathbb{F}$

(bb) $$(a \cdot b) \cdot c = a \cdot (b \cdot c) \qquad \textit{(multiplicative Associative Law)}$$
for all $a, b, c \in \mathbb{F}$.

(bc) *There exists an element* $1_{\mathbb{F}}$ *in* $\mathbb{F}$ *such that*
$$1_{\mathbb{F}} \cdot a = a = a \cdot 1_{\mathbb{F}} \qquad \textit{(multiplicative Identity)}$$
for all $a \in \mathbb{F}$

(bd) *For all* $a \in \mathbb{F}$ *there exists an element* $\frac{1}{a} \in F$ *so that*
$$a \cdot \tfrac{1}{a} = 0_{\mathbb{F}} = \tfrac{1}{a} \cdot a \qquad \textit{(multiplicative Inverse)}$$

(ca) $$a \cdot (b + c) = a \cdot b + a \cdot c \qquad \textit{(Right Distributive Law)}$$
for all $a, b, c \in \mathbb{F}$.

(cb) $$(a + b) \cdot c = a \cdot c + b \cdot c \qquad \textit{(Left Distributive Law)}$$
for all $a, b, c \in \mathbb{F}$.

(d) $$0_{\mathbb{F}} \neq 1_{\mathbb{F}}$$

If $(\mathbb{F}, +, \cdot)$ is a field the binary operations "+" and "$\cdot$" will be called addition and multiplication, respectively. To simplify notation we will often just write 0 for $0_{\mathbb{F}}$ and 1 for $1_{\mathbb{F}}$. But the reader should be aware that now 0 has two different meanings (depending on the context) namely the natural number 0 and the identity 0 for the addition in a field. Similarly 1 now has a double meaning. For example consider the following equation

$$1 + 1 = 0$$

This is obviously nonsense when 0 and 1 are viewed as natural numbers. But this equation can be true if 0 and 1 are the additive and multiplicative identities in a field. Indeed let as consider the following example. Let $\mathbb{F}_2 = \{0, 1\}$ and define addition and multiplication by the following charts:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Let us verify that $\mathbb{F}_2$ with this addition and multiplication is indeed a field.

It is obvious from the addition table that addition is commutative. To check that addition is associative let $a, b, c \in \mathbb{F}$. If one of a, b, c is 0, then both $(a+b)+c$ and $a+(b+c)$ are equal to the sum of the other two. If $a = b = c = 1$ then both $(a+b)+c$ and $a+(b+c)$

are equal to 1. So addition is associative. From the addition table we see that 0 is an additive identity. Also 0 is an additive inverse of 0 and 1 is an additive inverse of 1.

Again from the multiplication table we see that multiplication is commutative. If one of a, b, c is zero then both $(a \cdot b) \cdot c$ and $a \cdot (b \cdot c)$ are zero. If $a = b = c = 1$ then both $(a \cdot b) \cdot c$ and $a \cdot (b \cdot c)$ are equal to $1 \cdot 1 = 1$. From the multiplication table, 1 is a multiplicative identity. 1 is the only non-zero element and its multiplicative inverse is 1.

Since multiplication is commutative, we only need to verify the right distributive law. If $a = 0$ then both sides are 0. If $a = 1$ then both sides are equal to $b + c$.

Finally $1 \neq 0$ and so all the field axioms hold for $\mathbb{F}_2$.

So we indeed found a field in which $1 + 1 = 0$.

Let $\mathbb{F}$ be field, n a positive integer and $a \in \mathbb{F}$. Then na denotes the element

$$na = \underbrace{a + a + \ldots + a}_{n-\text{times}}.$$

Similarly a^n denotes the element

$$a^n = \underbrace{a \cdot a \cdot \ldots \cdot a}_{n-\text{times}}.$$

a^0 is defined as $1_{\mathbb{F}}$ and $0 \cdot a$ as $0_{\mathbb{F}}$. Further we set $(-n)a = n(-a)$ and if $a \neq 0$, $a^{-n} = (\frac{1}{a})^n$.

Also we will write $n_{\mathbb{F}}$ for $n \cdot 1_{\mathbb{F}}$, and, if no confusion is possible, n for $n_{\mathbb{F}}$. So depending on the context, 3 stands for the integer 3 or for the field element $3_{\mathbb{F}} = 1_{\mathbb{F}} + 1_{\mathbb{F}} + 1_{\mathbb{F}}$.

3 Polynomials and the field of rational function

Let $\mathbb{F}$ be a field. A polynomial over $\mathbb{F}$ is an expression of the form

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

where each $a_i, 0 \leq i \in n$ is an element of $\mathbb{F}$ and n is a non-negative integer.

The degree of a polynomial is the largest i with $a_i \neq 0$. Two polynomials are defined to be equal if the have the same degree and the same coefficients. So

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m$$

if and only if $a_i = b_i$ for all i. Here and in the future we will set $a_i = 0$ for all $i > n$.

The set of all polynomials over $\mathbb{F}$ will be denoted by $\mathcal{P}(\mathbb{F})$. Each polynomial $p = \sum_{i=0}^{n} a_i x^n$ gives rise to a function

$$\mathbb{F} \to \mathbb{F}, \quad a \to p(a) := \sum_{i=0} a_i a^n$$

You are probably used to thinking of polynomials as functions. But in the context of arbitrary fields this is a little dangerous. It can happen that two different polynomials give rise to the same function:

Consider the field $\mathbb{F}_2 = \{0, 1\}$ from above. Also consider the two polynomials x and x^3 (Here x stands for $1_{\mathbb{F}}x$). By definition, x and x^3 are different polynomials. But their corresponding functions are the same. Namely if we plug 0 in into x and x^3 we both times get 0. And if we plug 1 into x and x^3 we both times get 1. Now 0 and 1 are the only field elements available and the two polynomial functions associated to x and x^3 over $\mathbb{F}_2$ are the same.

We now define addition and multiplication of polynomials as follows:

$$\sum_{i=0}^{n} a_i x^i \quad + \quad \sum_{i=0}^{n} b_i x^i \quad = \quad \sum_{i=0}^{n} (a_i + b_i) x^i$$

and

$$\sum_{i=0}^{n} a_i x^n \quad \cdot \quad \sum_{i=0}^{m} b_i x^i \quad = \quad \sum_{i=0}^{n+m} (\sum_{k=0}^{i} a_k b_{i-k}) x^i$$

It is now easy to verify that almost all of the properties of a field hold for $\mathcal{P}(\mathbb{F})$. Indeed only one fails. Namely polynomials of degree 1 and larger do not have a multiplicative inverse. To circumvent this problem we introduce the set $\mathcal{R}(\mathbb{F})$ of rational functions over $\mathbb{F}$. A rational function over $\mathbb{F}$ is an expression of the form $\frac{p}{q}$ where p and q are polynomials over $\mathcal{F}$ with $q \neq 0$. Two such expressions $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ are defined to be equal if $p_1 q_2 = q_2 p_1$. Addition and multiplication of rational functions is defined as follows:

$$\frac{p_1}{q_1} + \frac{p_2}{q_2} = \frac{p_1 q_2 + q_1 p_2}{q_1 q_2}$$

and

$$\frac{p_1}{q_1} \cdot \frac{p_2}{q_2} = \frac{p_1 p_2}{q_1 q_2}.$$

A rational function $\frac{p}{q}$ is different from 0 (that is from $0_{\mathcal{R}(\mathbb{F})} = \frac{0}{1}$) if and only $p \neq 0$. In this case it has a multiplicative inverse namely $\frac{q}{p}$). It is now straightforward to check that $\mathcal{R}(\mathbb{F})$ is a field.

Notice that we obtain a second example of a field with $2_{\mathbb{F}} = 0_{\mathbb{F}}$ namely $\mathbb{F} = \mathcal{R}(\mathbb{F}_2)$.

4 Direct sums and linear independency

Definition 4.1 *Let $(U_1, U_2, \ldots U_n)$ be a list of subspaces of V.*

(a) *Let W be a subspace of V. We say that W is direct sum of the U_i's, and write*

$$W = \bigoplus_{i=1}^{n} U_i,$$

provided that for every $w \in W$ there exist **unique** *$u_i \in U_i$, $1 \leq i \leq n$, with*

$$w = \sum_{i=1}^{n} u_i.$$

(b) *$(U_1, U_2, \ldots, U_n)$ is called linearly independent if*

$$0 = \sum_{i=1}^{n} u_i$$

with $u_i \in U_i$ (for $1 \leq i \leq n$) implies $u_i = 0$ for all $1 \leq i \leq n$.

Let $V = \mathbb{F}^2$, $U_1 = \{(x, x) \mid x \in \mathbb{F}\}$ and $U_2 = \{(y, -y) \mid y \in \mathbb{F}\}$. It is easy to see that U_1 and U_2 are subspaces of V. We would like to compute their sum $U_1 + U_2$. At a first glance one would guess that $U_1 + U_2 = \mathbb{F}^2$. In order to show this we need to verify that each $(a, b) \in \mathbb{F}^2$ is the sum of an element in U_1 and an element from U_2. That is we have to find a solution to the equation

$$(a, b) = (x, x) + (y, -y)$$

That is

$$a = x + y \text{ and } b = x - y$$

Adding and subtracting the two equation we obtain

$$2x = a + b \text{ and } 2y = a - b$$

Now we have to distinguish two cases

Case 1: $2 \neq 0$ in $\mathbb{F}$.

Then we can divide by 2 to obtain the solution $x = \frac{a+b}{2}$ and $y = \frac{a-b}{2}$. So if $2 \neq 0$ we indeed have $V = U_1 + U_2$.

Case 2: $2 = 0$ in $\mathbb{F}$.

Then $2x = 0 = 2y$. So the two equations now read $0 = a + b$ and $0 = a - b$. Hence $a = b$ (which since $2 = 0$ and so $1 = -1$ is the same as $a = -b$.) So for $2 = 0$ we get $U_1 + U_2 = U_1$. Actually looking back we observe that for $2 = 0$, $y = -y$ for all $y \in \mathbb{F}$ and so $U_1 = U_2$.

Proposition 4.2 [**different ways to recognize direct sums**] *Let $U_1, \ldots, U_m$ be subspaces of V and put $W = U_1 + U_2 + \ldots + U_m$. Then the following four statements are equivalent:*

(a) $W = U_1 \oplus U_2 \oplus \ldots \oplus U_m$.

(b) *For all* $1 \leq i \leq m$

$$U_i \cap (U_1 + \ldots U_{i-1} + U_{i+1} + \ldots + U_m) = 0.$$

(c) $(U_1, U_2, \ldots, U_m)$ *is linearly independent.*

(d) *Whenever* $1 \leq i_1 < i_2 < \ldots < i_k \leq m$ *and* $0_V \neq u_{i_l} \in U_{i_l}$ *(for* $1 \leq l \leq k$*) then*

$$(u_{i_1}, \ldots, u_{i_k})$$

is linearly independent.

Proof: We will show that (a) implies (b), that (b) implies (c), that (c) implies (d) and finally that (d) implies (a).

(a)$\Longrightarrow$ (*b*):

Let $1 \leq i \leq m$ and suppose $u \in U_i \cap \sum_{j \neq i} U_j$. We need to show that $u = 0$. Note that $u \in U_i$ and there exists $u_j \in U_j$, $j \neq i$ with

$$u = u_1 + \ldots + u_{i-1} + u_{i+1} + \ldots + u_m$$

Hence

$$u = u_1 + \ldots + u_{i-1} + 0 + u_{i+1} + \ldots + u_m$$

and

$$u = 0 + \ldots + 0 + u + 0 + \ldots + 0$$

are two ways to write u as a sum of elements from $U_1, U_2, \ldots, U_m$. Recall that we are assuming that (a) holds, that is W is the direct sum of the U_i's. But this means there is only one way to write a given element as such a sum. Hence $u_j = 0$ for all $i \neq j$ and $u = 0$. Thus (b) holds.

(b)$\Longrightarrow$ (*c*):

Suppose that $u_i \in U_i$ for $1 \leq i \leq m$ and that

$$0 = u_1 + u_2 + \ldots + u_m$$

Then

$$-u_i = \sum_{j \neq i} u_j$$

The element on the left hand side of the equation is contained in U_i. The element on the right hand side is in $\sum_{j \neq i} U_j$. So $-u_i \in U_i \cap \sum_{j \neq i} U_j = \{0\}$.

Thus $u_i = 0$ for all i and (c) holds.

(c)$\Longrightarrow$ (d):

To simplify notation we assume without loss that $k = m$ and so $i_l = l$.

Suppose that $\sum_{i=1}^{m} a_i u_i = 0$. Since $a_i u_i \in U_i$, the definition of linear independence of subspaces implies $a_i u_i = 0$ for all i. Since $u_i \neq 0$ we conclude, $a_i = 0$ and so $(u_1, \ldots, u_m)$ is linearly independent.

(d)$\Longrightarrow$ (a):

Let $w \in W$ and suppose that $w = \sum_{i=1}^{m} u_i$ and $w = \sum_{i=1} u'_i$ for some $u_i, u'_i \in U_i$. Let $1 \leq i_1 < i_2 < \ldots < i_k \leq m$ be such that $u_i \neq u'_i$ if an only if $i = i_l$ for some $1 \leq l \leq k$. Subtracting the two equation we obtain

$$0 = (u_1 - u'_1) + (u_2 - u'_2) + \ldots + (u_m - u'_m) = 1 \cdot (u_{i_1} - u'_{i_1}) + 1 \cdot (u_{i_1} - u'_{i_2}) + \ldots + 1 \cdot (u_{i_k} - u'_{i_k})$$

Since U_{i_l} is a subspace of V, $0 \neq u_{i_l} - u'_{i_l} \in U_{i_l}$. Thus from (c) we conclude that $k = 0$. Hence $u_i = u'_i$ for all $1 \leq i \leq m$ and (a) holds. □

Remark 4.3 [**Condition d**] *If all the $U'_i s$ in the preceding Proposition 4.2 are non-zero, condition (d) can be replaced by:*

All list $(u_1, u_2, \ldots, u_m)$ of non-zero vectors with $u_i \in U_i$ for all $1 \leq i \leq m$ are linearly independent. □

Here is an examples which shows that linear independency can depend on the characteristic of $\mathbb{F}$. Consider the vectors $(1, -1)$ and $(1, 2)$ in $\mathbb{F}^2$. Are they linearly independent?

Suppose that $a_1(1, -1) + a_2(1, 2) = (0, 0)$.

Then

$$a_1 + a_2 = 0$$

and

$$-a_1 + 2a_2 = 0$$

Adding the two equations we obtain:

$$3a_2 = 0$$

Multiplying the first equation by 2 and subtracting the second gives:

$$3a_1 = 0$$

Case 1: $3 \neq 0$ in $\mathbb{F}$.

Then we can divide by 3 to obtain $a_1 = a_2 = 0$. So the two vectors are linearly independent.

Case 2 $3 = 0$ in $\mathbb{F}$.

We no longer are able to conclude that $a_1 = a_2 = 0$. $a_1 = 1$ and $a_2 = -1$ is a solution of the above equations. We double check:

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

So if $\operatorname{char} \mathbb{F} = 3$, the two vectors are linearly dependent.

Lemma 4.4 [**decomposing direct sums**] *Let U, W, X and Y be subspaces of V such that $V = U \oplus W$ and $W = X \oplus Y$. Then*

$$V = U \oplus X \oplus Y.$$

Proof: Since $V = U + W$ and $W = X + Y$ we have $V = U + (X + Y) = U + X + Y$. Now suppose that $u \in U, x \in X$ and $y \in Y$ with $u + x + y = 0$. Then $u \in U$ and $x + y \in W$. So since V is the direct sum of U and W we get $u = 0$ and $x + y = 0$. Now W is the direct sum of X and Y and so $x = 0$ and $y = 0$. Hence by 4.2

$$V = U \oplus X \oplus Y.$$

□

Proposition 4.5 [**different ways to recognize linear independence**] *Let $(v_1, v_2, \ldots v_n)$ be a list of vectors in V. Then the following statements are equivalent:*

(a) $(v_1, v_2, \ldots, v_n)$ *is linearly independent.*

(b) $v_i \neq 0$ *for all* $1 \leq i \leq n$ *and*

$$\operatorname{Span}(v_1, v_2, \ldots, v_n) = \mathbb{F}v_1 \oplus \mathbb{F}v_2 \oplus \ldots \oplus \mathbb{F}v_n$$

(c) *For each* $w \in \operatorname{Span}(v_1, v_2, \ldots, v_n)$ *there exist* **unique** $a_i \in \mathbb{F}$, $1 \leq i \leq n$, *with*

$$w = a_1v_1 + a_2v_2 + \ldots + a_nv_n$$

Proof: (a)$\Longrightarrow$ (b):

Here we assume that $(v_1, v_2, \ldots v_n)$ is linearly independent and need to show that each $v_i \neq 0_V$ and that the span of the v_i is the direct sum of the subspaces $\mathbb{F}v_i$. So suppose that $v_i = 0$ for some $1 \leq i \leq n$. Then

$$0 = 0 \cdot v_1 + 0 \cdot v_2 + \ldots 0 \cdot v_{i-1} + 1 \cdot v_i + 0 \cdot v_{i+1} + \ldots 0 \cdot v_n$$

But this contradicts the linear independence of the v_i's.

To verify that $\operatorname{Span}(v_i \mid 1 \leq i \leq n) = \bigoplus_{i=1}^{n} \mathbb{F}v_i$ we use condition (c) of Proposition 4.2 So suppose $u_i \in \mathbb{F}v_i$ with

$$0 = u_1 + u_2 + \ldots + u_n$$

Since $u_i \in \mathbb{F}v_i$ there exists $a_i \in \mathbb{F}$ with $u_i = a_i v_i$. So we get

$$0 = a_1u_1 + a_2u_2 + \dots a_nv_n$$

The linear independence of the v_i gives $a_i = 0$ for all i. Hence $u_i = a_iv_i = 0 \cdot v_i = 0$. So by 4.2 (b) holds.

(b)$\Longrightarrow$ (c):

By definition of "Span" any element in Span$(v_i \mid 1 \leq i \leq n)$ is a linear combination of the v_i's. Hence we just need to verify the uniqueness.

So suppose that

$$\sum_{i=1}^{n} a_iv_i = \sum_{i=1}^{n} a_i'v_i$$

for some $a_i, a_i' \in \mathbb{F}$. Put $u_i = a_iv_i$ and $u_i' = a_i'v_i$. Then both u_i and u_i' are in $\mathbb{F}v_i$ and

$$\sum_{i=1}^{n} u_i = \sum_{i=1}^{n} u_i'$$

Thus by the definition of "direct sum" $u_i = u_i'$ for all $1 \leq i \leq n$. Hence $a_iv_i = a_i'v_i$ and $(a_i - a_i')v_i = 0$. By assumption $v_i \neq 0$ and so by a homework problem $a_i - a_i' = 0$. Hence $a_i = a_i'$ and (c) holds.

(c)$\Longrightarrow$ (a):

Suppose that $0 = a_1v_1 + \dots a_2v_2 + \dots + a_nv_n$. Since also $0 = 0 \cdot v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_n$ the uniqueness assertion in (c) implies $a_i = 0$ for all $1 \leq i \leq n$. Thus $(v_1, v_2, \dots, v_n)$ is linearly independent and (a) holds. □

5 Bases and Dimensions

The main goal in this section is to prove that every linear independent list is contained in a basis and every spanning list contains a basis. The theorems and proofs presented here differ from the ones in the book and also from the once in class.

Let $\mathfrak{v} = (v_1, v_2, \dots v_m)$ and $\mathfrak{u} = (u_1, u_2, \dots, u_n)$ be list of vectors in V. We say that $\mathfrak{u}$ is spanning list for if Span$(\mathfrak{v}) = V$.

We say that $\mathfrak{v}$ is a sublist of $\mathfrak{u}$ (or that $\mathfrak{u}$ is contained in $\mathfrak{v}$) if there exists positive integers $1 \leq i_1 < i_2 < \dots < i_n \leq m$ such that $u_1 = v_{i_1}, u_2 = v_{i_2}, \dots u_n = u_{i_n}$. For example (v_2, v_4) is a sublist of $(v_1, v_2, v_3, v_4, v_5)$. If $\mathfrak{u}$ is a sublist of $\mathfrak{v}$ and $\mathfrak{u} \neq \mathfrak{v}$, $\mathfrak{v}$ is called a proper sublist of $\mathfrak{u}$. The set $\{i_1, i_2, \dots, i_k\}$ is denoted by $I(\mathfrak{u})$.

Lemma 5.1 [**linear independence of sublists**] *Let* $\mathfrak{v}$ *be a list of vectors in* V *and* $\mathfrak{u}$ *a sublist of* $\mathfrak{v}$.

(a) Span$(\mathfrak{u}) \leq$ Span$(\mathfrak{v})$

(b) *If* $\mathfrak{v}$ *is linearly independent, so is* $\mathfrak{u}$.

(c) *If* $\mathfrak{u}$ *is a spanning list for* V*, so is* $\mathfrak{v}$.

Proof: Let $v = \sum_{i \in I(\mathfrak{u})} a_i v_i$ be a linear combination of $\mathfrak{u}$. Define $a_i = 0$ for all $i \notin I(\mathfrak{u})$. Then v is also a linear combination of v, namely $v = \sum_{i=1}^{n} a_i v_i$.

In particular, (a) holds.

Suppose now that $\mathfrak{v}$ is linearly independent and $v = 0$. Then $a_i = 0$ for all $1 \leq i \leq n$ and so also $a_i = 0$ for all $i \in I(\mathfrak{u})$). Hence $\mathfrak{u}$ is linearly independent and (b) is proved.

Suppose next that $\mathrm{Span}(\mathfrak{u}) = V$. Then by (a)

$$V = \mathrm{Span}(\mathfrak{u}) \leq \mathrm{Span}(\mathfrak{v}) \leq V$$

and so $V = \mathrm{Span}(\mathfrak{v})$ and (c) holds. □

Lemma 5.2 [**sublists and linear independence**] *Let* $\mathfrak{v} = (v_1, v_2, \ldots, v_n)$ *a list of vectors in* V*. Then* $\mathfrak{v}$ *is linear independent if and only if* $\mathrm{Span}(\mathfrak{u}) \neq \mathrm{Span}(\mathfrak{v})$ *for all proper sublists* $\mathfrak{u}$ *of* $\mathfrak{v}$.

Proof: $\Longrightarrow$:

Suppose that $\mathfrak{v}$ is linearly independent. Let $\mathfrak{u}$ be a proper sublist of $\mathfrak{v}$. Then there exists $1 \leq i \leq n$ with $i \notin I(\mathfrak{u})$. By 4.5 and 4.2 $\mathbb{F}v_i \cap \sum_{j \neq i} \mathbb{F}v_j = 0$. Since $\mathrm{Span}(\mathfrak{u}) \leq \sum_{j \neq i} \mathbb{F}v_j$ we conclude $v_i \notin \mathrm{Span}(\mathfrak{u})$ and so $\mathrm{Span}(\mathfrak{u}) \neq \mathrm{Span}(\mathfrak{v})$.

$\Longleftarrow$:

Suppose that $\mathrm{Span}(\mathfrak{u}) \neq \mathrm{Span}(\mathfrak{v})$ for all proper sublists $\mathfrak{u}$ of $\mathfrak{v}$. We need to show that $\mathfrak{v}$ is linearly independent. If not, there exist $a_i \in \mathbb{F}, 1 \leq i \leq n$, not all zero with $\sum_{i=1}^{n} a_i v_i = 0$. Pick j with $a_j \neq 0$. Then

$$v_j = \sum_{i \neq j} -\frac{a_i}{a_j} v_i \in \mathrm{Span}(\mathfrak{u})$$

where $\mathfrak{u} = (v_1, v_2 \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$. But then $\mathrm{Span}(\mathfrak{u})$ is a subspace of V containing all the elements of $\mathfrak{v}$ and so $\mathrm{Span}(\mathfrak{v}) \leq \mathrm{Span}(\mathfrak{u}) \leq \mathrm{Span}(\mathfrak{v})$ and $\mathrm{Span}(\mathfrak{u}) = \mathrm{Span}(\mathfrak{v})$. This contradiction shows that $\mathfrak{v}$ is linearly independent. □

Given lists of vectors $\mathfrak{u}$ and $\mathfrak{v}$ in V.

We say that $\mathfrak{u}$ is a maximal linearly independent sublist of $\mathfrak{v}$ provided that

(a) $\mathfrak{u}$ is a sublist of $\mathfrak{v}$,
(b) $\mathfrak{u}$ is linearly independent, and
(c) no other sublist of $\mathfrak{v}$ which contains $\mathfrak{u}$ is linearly independent.

For example $((1,0,0),(0,1,0))$ is a maximal linearly independent sublist of

$$((1,0,0),(1,-1,0),(0,1,0)).$$

We say that $\mathfrak{u}$ is a minimal spanning sublist of $\mathfrak{v}$ provided that

(a) $\mathfrak{u}$ is a sublist of $\mathfrak{v}$,
(b) $\mathfrak{u}$ is a spanning list for V, and
(c) no other sublist of $\mathfrak{u}$ is a spanning list for V.

For example $((1,0),(0,1))$ is a minimal spanning sublist of

$$((1,0),(0,1),(1,1)).$$

Proposition 5.3 [**characterizing bases**] *Let $\mathfrak{w} = (w_1, w_2, \dots, w_n)$ be a spanning list for V and $\mathfrak{v} = (v_1, v_2, \dots, v_m)$ be a sublist of $\mathfrak{w}$. Then the following are equivalent:*

(a) *$\mathfrak{v}$ is a basis for V.*

(b) *$\mathfrak{v}$ is a maximal linearly independent sublist of $\mathfrak{w}$.*

(c) *$\mathfrak{v}$ is a minimal spanning subset of w.*

Proof: (a)$\Longrightarrow$ (b):

Suppose that $\mathfrak{v}$ is a basis for V. Then by definition of a basis $\mathfrak{v}$ is linearly independent and a spanning set. Now let $\mathfrak{u}$ be a sublist of $\mathfrak{w}$ containing $\mathfrak{v}$ with $\mathfrak{u} \neq \mathfrak{v}$. Then $V = \text{Span}(\mathfrak{v}) \leq \text{Span}(\mathfrak{u}) \leq V$ and so $\text{Span}(\mathfrak{u}) = \text{Span}(\mathfrak{v})$. Hence 5.2 implies that $\mathfrak{u}$ is linearly dependent. So $\mathfrak{v}$ is a maximal linearly independent sublist of $\mathfrak{w}$. That is (b) holds.

(b)$\Longrightarrow$ (c):

Let $1 \leq i \leq n$. By maximality of $\mathfrak{v}$, $(w_i, v_1, v_2, \dots, v_m)$ is linearly dependent. So there exists $a_j, 0 \leq i \leq n$ in $\mathbb{F}$, not all 0, such that

$$a_0 w_i + a_1 v_1 + \dots a_m v_m = 0$$

Suppose that $a_0 = 0$. Then

$$a_1 v_1 + \dots a_m v_m = 0$$

and the linear independence of $\mathfrak{v}$ forces $a_1 = a_2 \dots a_m = 0$
a contradiction. Thus $a_0 \neq 0$ and
$w_i = -\frac{a_1}{a_0} v_1 - \frac{a_2}{a_0} v_2 - \dots - \frac{a_m}{a_0} v_m$ and so $w_i \in \text{Span}(\mathfrak{v})$.
Since this is true for all $1 \leq i \leq n$ we get $V = \text{Span}(\mathfrak{w}) \leq \text{Span}(\mathfrak{v}) \subseteq V$ and so $\mathfrak{v}$ is a spanning list.

Since $\mathfrak{v}$ is linearly independent, 5.2 implies that $\text{Span}(\mathfrak{u}) \neq \text{Span}(\mathfrak{v})$ for all proper sublists $\mathfrak{u}$ of $\mathfrak{v}$. Hence $\mathfrak{v}$ is a minimal spanning sublists of $\mathfrak{w}$ and (c) holds.

(c)$\Longrightarrow$ (a):

By assumption $\mathfrak{v}$ is a spanning list and no proper sublist of $\mathfrak{v}$ spans $V = \text{Span}(\mathfrak{v})$. Thus by 5.2 $\mathfrak{v}$ is also linearly independent and hence a basis. □

Proposition 5.4 [**Existence of bases**] *Let $\mathfrak{v} = (v_1, \dots, v_m)$ be a linearly independent list of vectors in V and $\mathfrak{w} = (w_1, \dots, w_n)$ a spanning list of V. Then $m \leq n$ and there exists a sublists $\mathfrak{u} = (u_1, \dots, u_k)$ of $\mathfrak{w}$ with $k \leq n - m$ such that*

$$(\mathfrak{v}, \mathfrak{u}) = (v_1, \dots, v_m, u_1, \dots, u_k)$$

is a basis for V.

Proof: Replacing $\mathfrak{w}$ by a minimal spanning sublist, we may assume that no proper sublist of w spans V. Then by 5.3 $\mathfrak{w}$ is a basis for W.

We prove the proposition by induction on m. If $m = 0$ the proposition holds with $\mathfrak{u} = \mathfrak{w}$. So suppose $m \neq 0$ and that proposition holds for $m - 1$. Let $v^- = (v_1, v_2 \dots, v_{m-1})$. Then by the induction assumption there exists a sublist $\mathfrak{u}^+$ of $\mathfrak{w}$ of length at most $n-(m-1)$ such that $(\mathfrak{v}^-, \mathfrak{u}^+)$ is a basis for V. Since $(\mathfrak{v}^-, \mathfrak{u}^+)$ spans V so does $(\mathfrak{v}, \mathfrak{u}^+)$. Let $\mathfrak{u}$ be a sublist of $\mathfrak{u}^+$ such that $(\mathfrak{v}, \mathfrak{u})$ is a maximal linearly independent subset of $(\mathfrak{v}, \mathfrak{u}^+)$. We conclude from 5.3 that $(\mathfrak{v}, \mathfrak{u})$ is basis for V. It remains to show that $\mathfrak{u}$ has length not exceeding $n - m$. Otherwise, since $\mathfrak{u}^+$ has length at most $n - m + 1$ we get $\mathfrak{u} = \mathfrak{u}^+$. But then $(\mathfrak{v}^-, \mathfrak{u}^+)$ is contained in $(\mathfrak{v}, \mathfrak{u})$, a contradiction to 5.2 since $(\mathfrak{v}, \mathfrak{u})$ is linearly independent and both $(\mathfrak{v}, \mathfrak{u})$ and $(\mathfrak{v}^-, \mathfrak{u}^+)$ span V. □

Theorem 5.5 [**dimensions**] *Let V be a finite dimensional vector space.*

(a) *Every spanning list contains a basis for V.*

(b) *Every linearly independent list of vectors in V is contained in a basis for V.*

(c) *V has a basis.*

(d) *Any two bases of V have the same length.*

Proof: (a) Just choose a minimal spanning sublist. It's a basis by 5.3.

(b) Let $\mathfrak{v}$ be a linearly independent list of vectors. Since V is finite dimensional it has (by definition) a spanning list $\mathfrak{w}$. Now (b) follows from 5.4

(c) Follows from (a). (and also from (b) applied to the empty list of vectors)

(d) Let $\mathfrak{v}$ and $\mathfrak{w}$ be bases for V. In particular, $\mathfrak{v}$ is linearly independent and $\mathfrak{w}$ is a spanning set. So by 5.4 the length of $\mathfrak{v}$ does not exceed the length of $\mathfrak{w}$. But we can also apply 5.4 with the roles of $\mathfrak{v}$ and $\mathfrak{w}$ interchanged. So also the length of $\mathfrak{w}$ does to exceed to length of $\mathfrak{v}$ and (d) holds. □

6 Quotient Spaces And The Isomorphism Theorem

Let V be a vector space over $\mathbb{F}$ and W a subspace of V. For $v \in V$ define

$$v + W = \{v + w \mid w \in W\}.$$

$v + W$ is called the coset of W containing v.

We also define

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

Lemma 6.1 [**cosets**] *Let $v, u \in V$ and $W \leq V$. Then the following are equivalent:*

(a) $u \in v + W$.

(b) $u + W = v + W$

(c) $(u + W) \cap (v + W) \neq \emptyset$.

(d) $u - v \in W$.

Proof: (a)$\Longrightarrow$ (b):

If $u \in v + W$ then $u = v + x$ for some $x \in W$. Let $w \in W$. Then $u + w = (v + x) + w = v + (x + w)$. Since W is a subspace $x + w \in W$ and so $u + w \in v + W$. Hence $u + W \subseteq v + W$. Now $v = u + (-x) \in u + W$ and so by symmetry $v + W \subseteq u + W$ and (b) holds .

(b)$\Longrightarrow$ (c):

If $u + W = v + W$ then $v \in (u + W) \cap (v + W)$ and (c) holds.

(c)$\Longrightarrow$ (d):

Let $x \in (u + W) \cap (v + W)$. Then $x = u + w = v + w'$ for some $w, w' \in W$. Thus $u - v = w' - w \in W$ and (d) holds.

(d)$\Longrightarrow$ (a):

Let $w = u - v$. Then by assumption $w \in W$ and so $u = v + w \in v + W$. □

We will now define an addition and scalar multiplication which makes V/W into a vector space over $\mathbb{F}$. For $u, v \in V$ and $a \in \mathbb{F}$ we define

$$(v + W) + (u + W) := (v + u) + W$$

and

$$a \cdot (v + W) := av + W$$

Before we continue we need to verify that the above definitions are well defined, namely that the definitions only depend on $v + W$, $u + W$ and a, and not on the particular choice of u and v. For this we leave it as an exercise to verify that

$$(v+u)+W = \{v'+u' \mid v' \in v+W, u' \in u+W\}$$

and if $a \neq 0$,

$$av+W = \{av' \mid v' \in v+W\}$$

Also $0v+W = W$ for any $v \in V$.

That V/W is a vector space over $\mathbb{F}$ now follows more or less immediately from the definitions and the fact that V is a vector space. Note that $0_{V/W} = 0+W = W$. The vector space V/W is called the quotient space of V by W.

Lemma 6.2 [**linear combinations in quotient spaces**] *Let W be a subspace of V, and $v_i \in V_i$ and $a_i \in \mathbb{F}$ for $1 \leq i \leq n$. Then*

$$\sum_{i=1}^{n} a_i(v_i+W) = (\sum_{i=1}^{n} a_i v_i) + W.$$

Proof: This follows immediately by induction on n using the definition of the vector addition and scalar multiplication on V/W. □

Lemma 6.3 [**Quotients and bases**] *Let W be a finite dimensional subspace of V, let $(w_1, \ldots w_m)$ be basis for W and let $(v_1, \ldots v_k)$ be a list of vectors in V. Then the following three statements are equivalent:*

(a) *$(v_1+W, v_2+W, \ldots, v_k+W)$ is a basis for V/W.*

(b) *$(w_1, w_2, \ldots, w_m, v_1, v_2, \ldots, v_k)$ is a basis for V.*

(c) *$(v_1, v_2, \ldots, v_k)$ is a basis for $U := \mathrm{Span}(v_1, v_2, \ldots v_k)$ and $V = W \oplus U$.*

Proof: (a)$\Longrightarrow$ (b):

Let $v \in V$. Then by (a) there exists $b_1, \ldots b_k \in \mathbb{F}$ with

$$v+W = \sum_{i=1}^{k} b_i(v_i+W).$$

Put

$$u := \sum_{i=1}^{k} b_i v_i.$$

By 6.2 $v+W = u+W$. Thus by 6.1 $w := v-u \in W$. Since $W = \mathrm{Span}(w_1, \ldots, w_m)$,

$$w = \sum_{i=1}^{m} a_i w_i$$

for some $a - i \in \mathbb{F}$. Since $v = w + u$ we conclude that

$$v = \sum_{i=1}^{m} a_u w_i + \sum_{i=1}^{k} b_i v_i$$

Hence $v \in \text{Span}(w_1, \dots, w_m, v_1, \dots, v_k)$. Since $v \in V$ was arbitrary we conclude that $(w_1, \dots, w_m, v_1, \dots, v_k)$ spans V. To show it is also linearly independent suppose that

$$0 = a_1 w_1 + \ldots + a_m w_m + b_1 v_1 + \ldots + v_k \tag{1}$$

Since $a_1 w_1 + \ldots + a_m w_m \in W$, $a_1 w_1 + \ldots + a_m w_m + W = W = 0_{V/W}$ and so

$$0_{V/W} = b_1(v_1 + W) + \ldots + b_k(v_k + W)$$

The linear independence of the $(v_i + W)$'s implies $b_1 = b_2 = \ldots = b_k = 0$. Hence (1) yields

$$0 = a_1 w_1 + \ldots a_m w_m$$

and the linear independence of the $w_i' s$ gives $a_1 = a_2 = \ldots = a_m = 0$. So (b) holds.

(b)$\Longrightarrow$ (c):

Clearly (b) implies that $(v_1, v_2, \ldots v_k)$ is linearly independent and so a basis for U. Since $W + U$ is a subspace of V and contains all w_i and v_i's, (b) implies that $V = W + X$. Let $v \in W \cap U$. Since the w_i's span W, $v = \sum_{i=1}^{m} a_i w_i$ for some $a_i \in \mathbb{F}$. Since the v_i's span U, $v = \sum_{i=1}^{k} b_i v_i$ for some $b_i \in \mathbb{F}$. Hence

$$0 = v - v = \sum_{i=1}^{m} a_i w_i + \sum_{i=1}^{k} (-b_i) v_i$$

The linear independence of $(w_1, \dots, w_m, v_1, \dots, v_k)$ implies that all a_i and b_i are zero. Thus $v = 0$ and $W \cap U = 0$. Thus by 4.2 $V = W \oplus U$.

(c)$\Longrightarrow$ (a):

Let $v \in W$. Then $v = u + w$ with $w \in W$ and $u \in U$. Then $u = \sum_{i=1}^{k} b_i v_i$. Hence $v + W = u + W = \sum_{i=1}^{k} b_i(v_i + W)$ and so $(v_1 + W, \dots, v_k + W)$ spans V/W. Suppose now that

$$W = 0_{V/W} = \sum_{i=1}^{k} b_i(v_i + W)$$

and put $u = \sum_{i=1}^{k} b_i v_i$. Then $u \in W \cap U = \{0\}$ and the linear independence of the v_i implies $b_i = 0$ for all $1 \leq i \leq k$. So (a) holds. $\square$

Lemma 6.4 [**Dimension of quotient spaces**] *Let V be a finite dimensional vector space and W a subspace of V.*

(a)
$$\dim V = \dim W + \dim V/W.$$

(b) *There exists a subspace U of V with $V = W \oplus U$.*

(c) *Let U be any subspace of V with $V = W \oplus U$. Then $\dim U = \dim V/W$ and*
$$\dim V = \dim W + \dim U$$

Proof: (a) & (b) Choose a basis $(w_1, \ldots, w_m)$ of W and extend it to a basis
$$(w_1, \ldots, w_m, v_1, \ldots, v_k).$$
Put $U = \mathrm{Span}(v_1, \ldots, v_k)$. Then $\dim V = m + k$ and $\dim U = k$. Moreover, by 6.3 $\dim V/W = k$ and $V = W \oplus U$. So $\dim V = m + k = \dim W + \dim V/W$.

(c) Suppose that U is a subspace of V with $V = W \oplus U$. Then by 6.3 a basis of W combined with a basis for U gives a basis for V. Thus $\dim V = \dim W + \dim U$ and $\dim U = k = \dim V/W$. □

Theorem 6.5 [**The Isomorphism Theorem**] *Let $T : V \to W$ be a linear map. Then the map*
$$\bar{T} : V/\operatorname{null} T \to \operatorname{range} T, \quad v + \operatorname{null} T \to T(v)$$
is a well defined isomorphism.

Proof: We first need to show that the map is well defined. Suppose that $v + \operatorname{null} T = u + \operatorname{null} T$. Then $u = v + x$ with $T(x) = 0$ and so $T(u) = T(v) + T(x) = T(v) + 0 = T(v)$. Thus $\bar{T}$ is well defined.

That $\bar{T}$ is linear we leave as a routine exercise to the reader.

To show that $\bar{T}$ is injective, let $x + \operatorname{null} T \in \operatorname{null} \bar{T}$. Then $T(x) = 0$, $x \in \operatorname{null} T$ and $x + \operatorname{null} T = \operatorname{null} T = 0_{V/\operatorname{null} T}$. Thus $\bar{T}$ is injective.

To show that $\bar{T}$ is surjective let $w \in \operatorname{range} T$. Then $w = T(v)$ for some $v \in V$ and so $w = \bar{T}(v + \operatorname{null} T)$. Hence $\bar{T}$ is also surjective and thus an isomorphism. □

Proposition 6.6 [**linear maps and dimensions**] *Let V be a finite dimensional vector space and $T : V \to W$ a linear map. Then*
$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

Proof: By 6.4(a) $\dim V = \dim \operatorname{null} T + \dim(V/\operatorname{null} T)$. By the Isomorphism Theorem 6.5 $\dim(V/\operatorname{null} T) = \dim \operatorname{range} T$. (Note here that isomorphic spaces have the same dimension) □

7 Polynomials, the division algorithm and algebraically closed field

Proposition 7.1 [**division algorithm**] *Let $p, q \in \mathcal{P}(\mathbb{F})$ with $p \neq 0$. Then there exist uniquely determined polynomials $s, r \in \mathcal{P}(\mathbb{F})$ such that*

$$q = sp + r$$

and

$$\deg r < \deg p.$$

Proof: Existence of p and q:

We prove this by induction on $\deg q$. If $\deg q < \deg p$, we can choose $s = 0$ and $r = q$.

Suppose now that the existence has been proved for all polynomials of degree less than $\deg q$ and that $\deg p \leq \deg q$. Let $n = \deg p$, $m = \deg q$ and let a and b be the leading coefficients of p and q respectively. That is

$$p = ax^n + \text{ terms of degree less than } n$$

and

$$q = bx^m + \text{ terms of degree less than } m$$

Then both q and $\frac{b}{a}x^{m-n}p$ have leading term bx^m. Thus

$$\tilde{q} := q - \frac{b}{a}x^{m-n}p$$

has degree less than $m = \deg q$. So by the induction assumption there exists $\tilde{s}$ and $\tilde{r}$ in $\mathcal{P}(\mathbb{F})$ such that

$$\tilde{q} = \tilde{s}p + \tilde{r}$$

and

$$\deg \tilde{r} < \deg p.$$

Put $s = \frac{b}{a}x^{m-n} + \tilde{s}$ and $r = \tilde{r}$. Then

$$q = \frac{b}{a}x^{m-n}p + \tilde{q} = \frac{b}{a}x^{m-n}p + \tilde{s}p + \tilde{r} = sp + r$$

and

$$\deg r = \deg \tilde{r} < \deg p.$$

Uniqueness of p and q:

Suppose that $q = sp + r = \tilde{s}p + \tilde{r}$ with $\deg r < \deg p$ and $\deg \tilde{r} < \deg p$. Then $sp - \tilde{s}p = \tilde{r} - r$ and so

$$(s - \tilde{s})p = (\tilde{r} - r)$$

The right hand side is a polynomial of degree less then $\deg p$. The left hand side has degree larger or equal to $\deg p$, unless $s - \tilde{s} = 0$. We conclude that $s = \tilde{s}$ and $r = \tilde{r}$. □

Lemma 7.2 [**dividing polynomials in extension fields**] *Let $p, q \in \mathcal{P}(\mathbb{F})$ with $p \neq 0$. Suppose that $\mathbb{K}$ is a field with $\mathbb{F} \leq \mathbb{K}$ and $q \in \mathcal{P}(\mathbb{K})$ with $q = sp$. Then $s \in \mathcal{P}(\mathbb{F})$, and s is uniquely determined by p and q*

Proof: By 7.1 there exists $\tilde{s}, \tilde{r} \in \mathcal{P}(\mathbb{F})$ with $q = \tilde{s}p + \tilde{r}$ and $\deg \tilde{r} < \deg p$. Put $r = 0$. Then

$$q = sp + r = \tilde{s}p + \tilde{r}$$

By the uniqueness assertion in 7.1 (applied to the field $\mathbb{K}$ in place of $\mathbb{F}$) we get $s = \tilde{s}$ and $r = \tilde{r}$. So $s = \tilde{s} \in \mathcal{P}(\mathbb{F})$. □

We denote the uniquely determined s of the preceding lemma by $\frac{q}{p}$. Note that this is consisted out notations for the field of rational function $\mathcal{R}(\mathbb{F})$.

We say that $a \in \mathbb{F}$ is a root of the polynomial $p \in \mathcal{P}(\mathbb{F})$ if $p(a) = 0$.

Corollary 7.3 [**roots and factorizations**] *Let $p \in \mathcal{P}(\mathbb{F})$ and $a \in \mathbb{F}$. Then a is a root of p if and only if $p = s \cdot (x - a)$ for some $s \in \mathcal{P}(\mathbb{F})$.*

Proof: Suppose that $p(a) = 0$. By 7.1 $q = s \cdot (x - a) + r$ with $s, r \in \mathcal{P}(\mathbb{F})$ and $\deg r < \deg(x - a) = 1$. Thus $r = b$ for some $b \in \mathbb{F}$ and $0 = p(a) = s(a)(a - a) + r(a) = b$. So $q = s \cdot (x - a)$.

Conversely if $p = s \cdot (x - a)$ then $p(a) = s(a)(a - a) = 0$. □

Lemma 7.4 [**factorizing polynomials**] *Let $0 \neq p \in \mathcal{P}(\mathbb{F})$ and $m = \deg p$.*

(a) *There exist $k \in \mathbb{N}$ with $k \leq m$, $a_1, a_2, \ldots a_k \in \mathbb{F}$ and $q \in \mathcal{P}(\mathbb{F})$ such that*

(aa) $p = (x - a_1)(x - a_2) \cdot \ldots \cdot (x - a_k)q$

(ab) *q has no root in $\mathbb{F}$.*

(b) *$a \in \mathbb{F}$ is a root of p if and only if $a = a_i$ for some $1 \leq i \leq k$.*

(c) *q, k and $(a_1, a_2, \ldots, a_k)$ are unique (up to the order of the a_i's).*

(d) *p has at most m distinct roots.*

Proof: (a) Suppose that

$$(*) \quad p = (x - a_1)(x - a_2) \cdot \ldots \cdot (x - a_k)q$$

with $k \in \mathbb{N}$, $a_i \in \mathbb{F}$ and $q \in \mathcal{P}$. Then $\deg p = k + \deg q$ and so $k \leq m$. Also $p = q$ is such a factorization with $k = 0$. Thus we can choose k, a_i's and q as in (*) with k maximal. In particular, (aa) holds. Suppose that q has a root $b \in \mathbb{F}$. Then by 7.3 $q = s \cdot (x - b)$ with $s \in \mathcal{P}(\mathbb{F})$. But then

$$p = (x - a_1)(x - a_2)\dots(x - a_k)q = (x - a_1)(x - a_2)\dots(x - a_k)(x - b)s$$

a contradiction to the maximal choice of k. So (ab) holds.

(b) Clearly each a_i is a root of p. Conversely let a be any root of p. Then by (aa)

$$0 = (a - a_1)(a - a_2) \cdot \ldots \cdot (a - a_k)q(a)$$

By (ab) $q(a) \neq 0$. Since the product of non-zero elements in a field is non-zero, $a - a_i = 0$ for some i. Thus $a = a_i$ and (b) holds.

(c) Suppose that $p = (x - b_1)(x - b_2)\dots(x - b_l)r$ such that $r \in \mathcal{P}(\mathbb{F})$ has no root. If $l = 0$, then $p = r$ has no root, $k = 0$ and $p = q$ and (c) holds in this case. So suppose that $l \geq 1$. Then b_1 is a root of p and so by (b), $b_1 = a_i$ for some $1 \leq i \leq k$. Reordering the a_i we may assume that $b_1 = a_1$. Hence by 7.2

$$(x - a_2) \cdot \ldots \cdot (x - a_k) = \frac{p}{x - a_1} = \frac{p}{x - b_1} = (x - b_2) \cdot \ldots \cdot (x - b_l)r$$

(c) now follows by induction on $\deg p$.

(d) follows from (b) and $k \leq m$. □

Definition 7.5 (a) *A field is called algebraically closed if every non-constant polynomial in $\mathcal{P}(\mathbb{F})$ has a root in $\mathbb{F}$.*

(b) *An algebraic closure of the field $\mathbb{F}$ is a field $\bar{\mathbb{F}}$ such that*

(ba) $\mathbb{F} \leq \bar{\mathbb{F}}$.

(bb) $\bar{\mathbb{F}}$ *is algebraically closed.*

(bc) *Every element in $\bar{\mathbb{F}}$ is the root of a non-zero polynomial in $\mathcal{P}(\mathbb{F})$*

□

The next two theorems can be proved using advanced techniques from analysis and field theory. We state them without proof.

Theorem 7.6 [**fundamental theorem of algebra**] *The field $\mathbb{C}$ of complex numbers is algebraically closed.* □

Theorem 7.7 [**existence of an algebraically closure**] *Every field has an algebraic closure.* □

For example $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$.

Proposition 7.8 [**factorizing polynomials over algebraically closed fields**] *Let p be a non-zero polynomial over the algebraically closed field $\mathbb{F}$. Then p has a unique (up to the order of the factors) factorization of the form*

$$p = c(x - a_1)(x - a_2) \cdot \ldots \cdot (x - a_n)$$

where $n = \deg p$ and $c, a_1, a_2 \ldots, a_n \in \mathbb{F}$.

Proof: Since $\mathbb{F}$ is algebraically closed, every polynomial without roots is a constant. Hence the proposition follows from 7.4 □

A monic polynomial is a polynomial of the form

$$x^n + \text{ lower degree terms.}$$

Lemma 7.9 [**factorizing polynomials over the reals**] *Let $0 \neq p \in \mathcal{P}(\mathbb{R})$. Then there exists a unique (up to order) $c \in \mathbb{R}$, $k \in \mathbb{N}$ and $q_1, q_2, \ldots q_k \in \mathcal{P}(\mathbb{R})$ such that*

(a) $p = cq_1q_2 \ldots q_k$

(b) *For each $1 \leq i \leq k$, q_i is monic and $q_i \leq 2$.*

(c) *If $1 \leq i \leq k$ and $\deg q_i = 2$, then q_i has not root in $\mathbb{R}$.*

Proof: Existence:

If p is a constant we can choose $c = p$ and $k = 0$.

So suppose $\deg p \geq 1$. Let λ be a root for $p \in \mathbb{C}$.

If $\lambda \in \mathbb{R}$, put $q_1 = x - \lambda$. Note that by refroots and factorizations $p = (x - \lambda) \cdot q$ for some $q \in \mathcal{P}(\mathbb{R})$. By induction q and so also p has the appropriate factorization.

So suppose $\lambda \notin \mathbb{R}$. Then (see [Axler, 4.10]) also $\bar{\lambda}$ is a root of p. Hence $p = (x-\lambda)(x-\bar{\lambda})q$ for some $q \in \mathcal{P}(\mathbb{C})$. Put $q_1 = (x - \lambda)(x - \bar{\lambda})$. Then $q_1 \in \mathcal{P}(\mathbb{R})$. By 7.2 $q \in \mathcal{P}(\mathbb{R})$. Also q_1 has no real roots and so we are again done by induction.

Uniqueness:

Let λ be a root of p in $\mathbb{R}$. Then λ is a root of some q_i. It follows that $q_i = x - \lambda$ if λ is real and $q_i = (x - \lambda)(x - \bar{\lambda})$ if λ is not real. So any two factorization have at least one common factor. The uniqueness now follows by induction. □

8 Eigenspaces

Definition 8.1 *Let $T : V \to V$ be an operator and W a subspace in V. Then W is called T-invariant if $T(W) \leq W$, that is $T(w) \in W$ for all $w \in W$.*

Lemma 8.2 [**decomposing operators**] *Let $T : V \to V$ be an operator and W a T-invariant subspace.*

(a) *The map $T|_W : W \to W, \quad w \to T(w)$ is an operator on W.*

(b) *The map $T|_{V/W} : V/W \to V/W \quad v + W \to T(v) + W$ is a well-defined operator on V/W.*

(c) *Suppose that $(w_1, \dots, w_m)$ is basis for W and $(v_1 + W, \dots, v_k + W)$ is a basis for V/W. Let A and B be the matrices of $T|_W$ and $T|_{V/W}$, respectively. The there exists an $m \times k$ matrix C so that the matrix of T with respect to $(w_1, \dots, w_m, v_1, \dots v_k)$ is*

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

Proof: (a) is obvious.

(b) Let $v \in V$ and $w \in W$. Then $T(w) \in W$ and so $T(v+w)+W = T(v)+T(w)+W = T(v)+W$. So $T|_{V/W}$ is well defined. Obviously $T|_{V/W}$ is linear and so (b) holds.

To simplify notation we will often write $T(v + W)$ instead of $T\,|_{V/W}\ (v + W)$. So $T(v + W) = T(v) + W$.

For (c) note first that

$$T(w_j) = \sum_{i=1}^{m} a_{ij} w_i = \sum_{i=1}^{k} a_{ij} w_i + \sum_{j=1}^{k} 0 \cdot v_j.$$

Also

$$T(v_j) + W = T(v_j + W) = \sum_{i=1}^{k} b_{ij}(v_i + W) = (\sum_{i=1}^{k} b_{ij} v_j) + W$$

and so

$$T(v_j) = x_j + \sum_{i=1}^{k} b_{ij} v_k$$

for some $x_j \in W$. Since $(w_1, \dots, w_m)$ is a basis for W there exists $c_{ij} \in \mathbb{F}$ with

$$x_j = \sum_{i=1}^{m} c_{ij} w_i$$

Hence

$$T(v_j) = \sum_{i=1}^{m} c_{ij} w_i + \sum_{j=1} b_{ij} v_i$$

Thus (c) holds with $C = (c_{ij})$. □

Let $T : V \to V$ be an operator and $\lambda \in \mathbb{F}$. We say that λ is an eigenvalue for T if there exists a non-zero vector $v \in V$ with

$$T(v) = \lambda v$$

If λ is an eigenvalue for T, then any vector with $T(v) = \lambda v$ is called an eigenvector of T corresponding to λ. The set of all the eigenvectors of T corresponding to T is called the eigenspace of T corresponding to λ and is denoted by $V_{T,\lambda}$. Note that

$$T(v) = \lambda v \quad \iff \quad (T - \lambda \mathrm{id})(v) = 0$$

So

$$V_{T,\lambda} = \mathrm{null}(T - \lambda \mathrm{id})$$

and $V_{T,\lambda}$ is a subspace of V. A generalized eigenvector for T is vector $v \in V$ such that $(T - \lambda \mathrm{id})^n(v) = 0$ for some $n \in \mathbb{N}$. The height ht v of a generalized eigenvector is the smallest $n \in \mathbb{N}$ with $(T - \lambda)^n(v) = 0$. So 0_V is the only generalized eigenvector of height 0 and the generalized eigenvectors of height 1 are exactly the non-zero eigenvectors. Let $V_{T,\lambda,n}$ be the set of all generalized eigenvector of height at most n.Then

$$V_{T,\lambda,n} = \mathrm{null}(T - \lambda \mathrm{id})^n.$$

Let $V_{T,\lambda,\infty}$ is be set of all generalized eigenvectors of T corresponding to λ. Then

$$V_{T,\lambda,\infty} = \bigcup_{n=1}^{\infty} V_{T,\lambda,n}$$

$V_{T,\lambda,\infty}$ is called a generalized eigenspace.

We claim that generalized eigenspaces are subspaces of V. Indeed, let v, w be generalized eigenvectors and $a \in \mathbb{F}$. Let $n =$ ht v and $m =$ ht w. Put $k = \max(n, m)$. Then both v and w are in the subspace $V_{T,\lambda,k}$ of V. Hence also kv and $v + w$ are in $V_{T,\lambda,k}$ and so in $V_{T,\lambda,\infty}$.

For ease of notation will often write just λ for the operator $\lambda \mathrm{id}_V$. So $T = 0$ means $T(v) = 0$ for all $v - inV$ and $T - \lambda$ is stands for the operator $T - \lambda \mathrm{id}_V$.

If $p = \sum_{i=0}^{m} a_i x^i \in \mathcal{P}(\mathbb{F})$ and T is an operator on V we define $p(T) = \sum_{i=0}^{m} a_i T^i$. Note that $p(T)$ is again an operator on V. Let S and T be operators. We say that S and T commute if $S \circ T = T \circ S$.

Lemma 8.3 [**commuting algebra**]

(a) *R, S and T be operators on V and $a \in \mathbb{F}$. Suppose that R and S commute with T. Then $R + S, R \circ S$ and aS all commute with T.*

(b) *Let T be an operator on V and $p, q \in \mathcal{P}(\mathbb{F})$. Then over $\mathbb{F}$. Then $p(T)$ commutes with $q(T)$.*

Proof: (a)

$$(R+S)\circ T = R\circ T + S\circ T = T\circ R + T\circ S = T\circ (R+S)$$

$$(R\circ S)\circ T = R\circ (S\circ T) = R\circ (T\circ S) = (R\circ T)\circ S = (T\circ R)\circ S = T\circ (R\circ S)$$

$$(aS)\circ T = a(S\circ T) = a(T\circ S) = T\circ (aS)$$

(b) $p(T)\circ q(T) = (pq)(T) = (qp)(T) = q(T)\circ p(T)$. □

Lemma 8.4 [**commuting operators and eigenspaces**]

(a) *Let S and T be operators on V. Then* $\operatorname{null} S$ *and* $\operatorname{range} S$ *are T-invariant.*

(b) *Let T be an operator, $\lambda \in \mathbb{F}$, $n \in \mathbb{N}\cup\{\infty\}$ and $p \in \mathcal{P}(T)$. Then $V_{T,\lambda,n}$ is $p(T)$ invariant. That is, if v is an generalized eigenvector of height at most n, then also $p(T)(v)$ is a generalized eigenvector of height at most n.*

Proof: (a) Let $v \in \operatorname{null} S$. Then $S(T(v)) = T(S(v)) = T(0) = 0$ and so $T(v) \in \operatorname{null} T$. Hence $\operatorname{null} S$ is T-invariant.

Let $v \in \operatorname{range} S$. Then $v = S(w)$ for some $w \in V$. Thus $T(v) = T(S(w)) = S(T(w) \in \operatorname{range} S$. Hence $\operatorname{range} S$ is T-invariant.

(b) Suppose first that $n \in \mathbb{N}$. By 8.3 $p(T)$ commutes with $(T-\lambda)^n$. Hence by (a) $\operatorname{null}(T-\lambda)^n$ is T-invariant. But $\operatorname{null}(T-\lambda)^n = V_{T,\lambda,n}$ and so (b) holds in this case.

Now let $v \in V_{T,\lambda,\infty}$ and let h be the height of v. Then by the "$n \in \mathbb{N}$" case, $p(T)(v)$ is a generalized eigenvector (of height at most h). So $p(T)(v) \in V_{T,\lambda,\infty}$. □

Lemma 8.5 [**quotients of eigenvectors**] *Let T be an operator on V and W a T-invariant subspaces. Let $\lambda \in \mathbb{F}$ on V,and h a generalized eigenvector of height h with respect to T and λ*

(a) *For all $n \in \mathbb{N}\cup\{\infty\}$, $V_{T,\lambda,n} \cap W \leq W_{T\lambda,n}$*

(b) *$v+W$ is a generalized eigenvector of height at most h with respect to $T|_{V/W}$ and λ.*

(c) *If $V_{T,\lambda,m} \leq W$ for some $m \leq h$ then $v+W$ has height at most $h-m$.*

Proof: (a) is obvious. (b) follow from (c) applied with $m=0$. So we only need to prove (c). Let $x \in V_{T,\lambda,n} + W/W$. Then $x = v+W$ for some $v \in V_{T,\lambda,n}$. Put $k = \operatorname{ht} v$. Then $k \leq n$. If $k \leq m$ then $v \in V_T, \lambda, m \leq W$ and so $x = W = 0_{V/W}$. So suppose $k > m$. Then

$$(T-\lambda)^m((T-\lambda)^{k-m}(v)) = (T-\lambda)^k(v) = 0$$

Thus

$$(T-\lambda)^{k-m}(v) \in V_{T,\lambda,m} \leq W$$

and

$$(T-\lambda)^{k-m}(x) = (T-\lambda)^{k-m}(v) + W = W = 0_{V/W}$$

Hence

$$x \in (V/W)_{T,\lambda,k-m} \leq (V/W)_{T,\lambda,n-m}.$$

□

Lemma 8.6 [**polynomials and eigenvectors**] *Let $T \in \mathcal{L}(V)$, $q \in \mathcal{P}(\mathbb{F})$, $\lambda \in \mathbb{F}$ and v an eigenvector of T in V corresponding to λ. Then*

$$q(T)(v) = q(\lambda)v$$

Proof: By induction on $\deg q$. If $q = 0$, both sides equal 0_V. So suppose $q \neq 0$ and let $q = rx + b$ with $\deg r < \deg q$ and $b \in \mathbb{F}$. Then $q(T) = r(T) \circ T + b$. By induction, $r(T)(v) = r(\lambda)v$ and

$$q(T)(v) = r(T)(T(v)) + bv = r(T)(\lambda v) + bv =$$
$$= \lambda r(T)(v) + bv = \lambda r(\lambda)v + bv = (r(\lambda)\lambda + b)v = q(T)v$$

□

Lemma 8.7 [**Eigenspaces are linearly independent**] *Let T be a linear operator on V and $(\lambda_1, \ldots, \lambda_m)$ a list of distinct eigenvalues for T on V. Then the corresponding list of generalized eigenspaces*

$$(V_{T,\lambda_i,\infty} \mid 1 \leq i \leq m)$$

is linearly independent.

Proof: Let $u_i \in V_{T,\lambda_i,\infty}$ with $\sum_{i=1}^m u_i = 0$. Let $h_i = \text{ht } u_i$ and $h = \sum_{i=1}^m h_i$. The proof is by induction on h. If $h = 0$, then $h_i = 0$ for all i and so $u_i = 0$. So we may assume that $h > 0$. Pick $1 \leq j \leq m$ with $h_j \neq 0$ and let $W = V_{T,\lambda_j}$. By 8.4, W is T invariant. By 8.5 $u_i + W$ is a generalized eigenvector corresponding to λ_i of height at most h_i. Moreover u_j has height at most $h_j - 1$. Thus by induction $u_i + W = 0_{V/W}$ for all $1 \leq i \leq n$. So $u_i \in W$ for all $1 \leq i \leq n$ and u_i is an eigenvector of T corresponding to λ_i. Using 8.6 we compute

$$0 = (T-\lambda_i)^{h_i}(u_i) = (\lambda_j - \lambda_i)^{h_i} u_i$$

Let $i \neq j$. Then $\lambda_j - \lambda_i \neq 0$ and thus $u_i = 0$.
Thus $0 = \sum_{i=1}^m u_i = u_j$ and so $u_i = 0$ for all $1 \leq i \leq m$ □

Corollary 8.8 [**Eigenvectors are linearly independent**]

(a) *Let T be a linear operator on V, $\lambda_1, \dots, \lambda_m$ distinct eigenvalues and v_i a non-zeros eigenvector for T corresponding to λ_i. Then*

$$(v_1, v_2, \dots, v_m)$$

is linearly independent.

(b) *Let T be a linear operator on V and suppose that $n = \dim V$ is finite. Then T has at most n distinct eigenvalues.*

Proof: (a) Follows from 8.7 and 4.2.

(b) Let $\lambda_i, 1 \leq i \leq m$ be distinct eigenvalues for T. By definition of an eigenvalue there exists an non-zero eigenvector v_i for T corresponding to λ_i. By (a) $(v_1, \dots, v_m)$ is linearly independent. Thus by 5.4 $m \leq n$ □

Definition 8.9 *$T \in \mathcal{L}(T)$ is called nilpotent if there exists $m \in \mathbb{N}$ with $T^m = 0$. If T is nilpotent the minimal such m is called the height of T.*

Lemma 8.10 [**Eigenvalues of nilpotent operators**] *Let T be a nilpotent operator on the non-zero V. Then 0 is the unique eigenvalue for T. The height of T is the maximal height of a generalized eigenvector in V with respect to T and 0.*

Proof: Let m be the height of T. Then $T^m(v) = 0_V$ for all $v \in V$ and so $V_{T,0,m} = V$. In particular, 0 is an eigenvalue. Let $0 \neq \lambda \in \mathbb{F}$. By 8.7

$$0 = V_{T,0,m} \cap V_{T,\lambda,\infty} = V \cap V_{T,\lambda,\infty} = V_{T,\lambda,\infty}.$$

So T has no non-zero eigenvalues. Let h be the maximal height of a vector in V. Also let v be a vector of height h. Since $T^m = O\mathrm{id}_V$, $T^m(v) = 0$ and so $h \leq m$. Let w be any vector in V. Then w has height less or equal to h and so $T^h(w) = 0$ for all $w \in V$. Thus $T^h = 0\mathrm{id}_V$ and $m \leq h$. So $m = h$. □

Theorem 8.11 [**Jordan Canonical Form for nilpotent operators**] *Let T be a nilpotent operator of height m on the finite dimensional vector space V.*

(a) *There exists non negative integer $n_i, 1 \leq i \leq m$ and a basis*

$$\mathfrak{v} = (v_{i,j,k} \mid 1 \leq i \leq m, 1 \leq j \leq n_i, 1 \leq k \leq i)$$

for v such that

$$T(v_{i,j,k}) = \begin{cases} v_{i,j,k-1} & \text{if } k \neq 1 \\ 0 & \text{if } k = 1 \end{cases}$$

(b) *For $1 \leq i \leq m$ and $1 \leq j \leq n_i$ put $\mathfrak{v}_{i,j} = (v_{i,j,k} \mid 1 \leq k \leq i)$ and $U_{i,j} = \mathrm{Span}(\mathfrak{v}_{i,j})$. Then U_{ij} is a T-invariant subspace of V and*

$$V = \bigoplus_{i=1}^{m} \bigoplus_{j=1}^{n_i} U_{i,j}$$

(c) *The matrix of $T\mid_{U_{ij}}$ with respect to the basis $\mathfrak{v}_{i,j}$ is the $i \times i$ matrix*

$$J(i) := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

(d) *Let $J(i^n)$ be the matrix $ni \times ni$ matrix*

$$J(i^n) := \begin{pmatrix} J(i) & 0 & \dots & 0 & 0 \\ 0 & J(i) & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & J(i) & 0 \\ 0 & 0 & \dots & 0 & J(i) \end{pmatrix}$$

where $J(i)$ appears n-times.

Then the matrix of T with respect to the basis $\mathfrak{v}$ is

$$J(1^{n_1}, 2^{n_2}, \dots, m^{n_m}) := \begin{pmatrix} J(1^{n_1}) & 0 & \dots & 0 & 0 \\ 0 & J(2^{n_2}) & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & J((m-1)^{n_{m-1}}) & 0 \\ 0 & 0 & \dots & 0 & J(m^{n_m}) \end{pmatrix}$$

Proof: (a) We will first show that there exist lists of vectors

$$\mathfrak{w}_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n_i}), 1 \leq i \leq m$$

such that for all $1 \leq i \leq m$

$$(\mathfrak{w}_i, \mathfrak{w}_{i+1}, \dots, \mathfrak{w}_m) \text{ is a basis for } \operatorname{range} T^{i-1} \cap \operatorname{null} T. \tag{1}$$

The proof is by downwards induction on i. Suppose we already found lists of vectors $\mathfrak{w}_{l+1}, \dots, \mathfrak{w}_m$ such that (1) holds for all $l+1 \leq i \leq m$.

Note that $\operatorname{range} T^l = T^{l-1}(T(V)) \leq \operatorname{range} T^{l-1}$. Hence

$$\operatorname{range} T^l \cap \operatorname{null} T \leq \operatorname{range} T^{l-1} \cap \operatorname{null} T.$$

Since (1) holds for $l+1$, $(\mathfrak{w}_{l+1}, \mathfrak{w}_{l+2}, \dots, \mathfrak{w}_m)$ is basis for $\operatorname{range} T^l \cap \operatorname{null} T$. So it is also a set of linearly independent vectors in $\operatorname{range} T^{l-1} \cap \operatorname{null} T$. So by 5.5 it can be extended to a basis

$$(\mathfrak{w}_l, \mathfrak{w}_{l+1}, \mathfrak{w}_{l+2}, \dots, \mathfrak{w}_m)$$

of $\operatorname{range} T^{l-1} \cap \operatorname{null} T$. Thus (1) also holds for l and so by induction for all $1 \leq l \leq m$.

Since $w_{i,j} \in \operatorname{range} T^{i-1}$ we can choose $v_{i,j} \in V$ with

$$T^{i-1}(v_{i,j}) = w_{i,j}$$

For all $1 \leq k \leq i$ we define

$$v_{i,j,k} = T^{i-k}(v_{i,j})$$

Then

$$v_{i,j,1} = T^{i-1}(v_{i,j}) = w_{i,j}$$

and

$$v_{i,j,i} = T^0(v_{i,j}) = v_{i,j}$$

If $k > 1$ then

$$T(v_{i,j,k}) = T(T^{i-k}(v_{i,j})) = T^{i-(k-1)}(v_{i,j}) = v_{i,j,k-1} \tag{2}$$

and as $w_{i,j} \in \operatorname{null} T$

$$T(v_{i,j,1}) = T(w_{i,j}) = 0 \tag{3}$$

So to complete the proof of (a) it remains to show that

$$\mathfrak{v} = (v_{i,j,k} \mid 1 \leq i \leq m, 1 \leq j \leq n_i, 1 \leq k \leq i)$$

is a basis for V. That is we need to verify that $\mathfrak{v}$ is linearly independent and a spanning set.

From (2), (3) and induction we get

$$T^l(v_{i,j,k}) = \begin{cases} v_{i,j,k-l} & \text{if } l < k \\ 0_V & \text{if } l \geq k \end{cases} \tag{4}$$

In particular,

$$T^{k-1}(v_{i,j,k}) = w_{i,j} \tag{5}$$

To show linear independence let $a_{i,j,k} \in \mathbb{F}$ with

$$\sum a_{i,j,k} v_{i,j,k} = 0 \tag{6}$$

Suppose that $a_{i',j',k'} \neq 0$ for some i', j', k'. We choose such an $a_{i',j',k'}$ with k' maximal. Then $a_{i,j,k} = 0$ for all (i, j, k) with $k > k'$.

In particular,

$$T^{k'-1}(a_{i,j,k} v_{i,j,k}) = 0_V \text{ for all } k > k' \tag{7}$$

By (4) $T^{k'-1}(v_{i,j,k}) = 0_V$ if $k' - 1 \geq k$, that is for $k < k'$. Thus

$$T^{k'-1}(a_{i,j,k} v_{i,j,k}) = 0_V \text{ for all } k < k' \tag{8}$$

By (5)

$$T^{k'-1}(a_{i,j,k} v_{i,j,k}) = a_{i,j,k'} w_{i,j} \text{ for } k = k' \tag{9}$$

Applying $T^{k'-1}$ to both sides of (6) and using (7),(8) and (9) we get

$$0_V = \sum_{i,j} a_{i,j,k'} w_{i,j}$$

The linear independence of the $w_{i,j}$'s implies $a_{i,j,k'} = 0$ for all i, j. But this contradicts $a_{i',j',k'} \neq 0$. Hence $\mathfrak{v}$ is linearly independent.

Let $U = \text{Span}(\mathfrak{v})$ and $v \in V$. To show that $\mathfrak{v}$ is a spanning set we need to show that $v \in U$. We do this by induction on the height h of v. If $h = 0$, $v = 0$ and so $v \in U$. Suppose now that $h \geq 1$ and U contains all vectors of height less than h. The strategy is to find a vector $u \in U$ so that $v - u$ has height less than h. Then $v - u \in U$ be induction, and so also $v = (v - u) + u$ is in U. To find u note first that $T(T^{h-1}(v)) = T^h(v) = 0$ and so $T^{h-1}(v) \in \text{range}\, T^{h-1} \cap \text{null}\, T$. Thus by (1)

$$T^{h-1}(v) = \sum_{i=h}^{m} \sum_{j=1}^{n_i} a_{i,j} w_{i,j}$$

for some $a_{i,j} \in \mathbb{F}$. Put

$$u = \sum_{i=h}^{m} \sum_{j=1}^{n_i} a_{i,j} v_{i,j,h}$$

Then by (5)

$$T^{h-1}(u) = \sum_{i=h}^{m} \sum_{j=1}^{n_i} a_{i,j} w_{i,j}$$

Thus $T^{h-1}(v) = T^{h-1}(u)$, $T^{h-1}(v-u) = 0$ and so $v-u$ has height at most $h-1$. So $v-u$ has height less than h and as observed above, $v \in U$. So $\mathfrak{v}$ is a spanning sets and all parts of (a) are proved.

(b) That V is the direct sum follows from exercise 1 from Homework set 4. So it remains to verify that $U_{i,j}$ is T invariant. Let $u \in U_{i,j}$. Since $U_{i,j}$ is spanned by $(v_{i,j,k} \mid 1 \leq k \leq i)$ we have $u = \sum_{i=1}^{k} a_i v_{i,j,k}$ for some $a_k \in \mathbb{F}$. Then by (2) and (3):

$$T(u) = \sum_{i=2}^{k} a_i v_{i,j,k-1}$$

Hence $T(u) \in U_{i,j}$ and $U_{i,j}$ is T-invariant.

(c) Fix i, j with $1 \leq i \leq m$ and $1 \leq j \leq n_i$. Put $u_k = v_{i,j,k}$. Then $\mathfrak{v}_{i,j} = \{u_1, u_2, \ldots, u_i\}$ and by (2) and (3)

$$T(u_1) = 0_V, T(u_2) = u_1, T(u_3) = u_2 \ldots, T(u_i) = u_{i-1}$$

Thus (c) holds.

(d) follows form (b) and (c). □

Lemma 8.12 [**Existence of the minimal polynomial**] *Let T be an operator on the finite dimensional vector space V. Then there exists a non-zero polynomials q with $q(T) = 0$.*

Proof: Let $n = \dim V$. Then $\mathcal{L}(V) \cong \mathcal{M}_n(\mathbb{F})$ and has dimension $m := n^2$. It follows that the list

$$T^0, T^1, T^2, \ldots, T^m$$

cannot be linearly independent. Hence there exists $a_i \in \mathbb{F}$, $(0 \leq i \leq m)$ with

$$\sum_{i=0}^{m} a_i T^i = 0$$

Put $q = \sum_{i=0}^{m} a_i x^i$. Then $q \neq 0$ and $q(T) = 0$. □

Definition 8.13 *Let V be finite dimensional and $T \in \mathcal{L}(V)$. The minimal degree of a constant polynomial q with $q(T) = 0$ is denoted by* $\deg T$. *A minimal polynomial of T is a monic polynomial q with $q(T)$ and* $\deg q = \deg T$.

Lemma 8.14 [**roots of p(T)**] *Let T be an operator of V, $p \in \mathcal{P}(\mathbb{F})$ and $\lambda \in \mathbb{F}$.*

(a) *Let v be non-zero generalized eigenvector of T corresponding λ. If $p(T)(v) = 0$ then $p(\lambda) = 0$.*

(b) *If λ is not a root of p, then $V_{T,\lambda,\infty} \cap \operatorname{null} p(T) = 0$.*

Proof: (a) Let h be the height of v. The proof is by induction on h. Since $v \neq 0$, $h \neq 0$. If $h = 1$ then v is an eigenvector, so by 8.6 $p(T)(v) = p(\lambda)v$. Since $v \neq 0$, but $p(T)(v) = 0$ we get $p(\lambda) = 0$.

So suppose that $h > 1$ and that the lemma is true for all non-zero vectors of height smaller than h. Let $w = (T - \lambda)(v)$. Since $h > 1$, $w \neq 0$. Also w is a generalized eigenvector of height $h - 1$. By 8.4 $\operatorname{null} p(T)$ is $T - \lambda$ invariant. Since $v \in \operatorname{null} p(T)$ we conclude $w \in \operatorname{null} p(T)$. So w fulfils all the assumption on v and so by induction $p(\lambda) = 0$.

(b) Otherwise there exists $0 \neq v \in V_{T,\lambda,\infty} \cap \operatorname{null} p(T)$. Then v is a non-zero generalized eigenvector and $p(T)(v) = 0$. But then by (a) $p(\lambda) = 0$. □

Theorem 8.15 [**Uniqueness of the minimal polynomial**] *Let T be an operator on the finite dimensional vector V.*

(a) *T has a unique minimal polynomial q.*

(b) *Let $p \in \mathcal{P}(\mathbb{F})$. Then $p(T) = 0$ if and only if q divides p.*

Proof: Let $\tilde{q}$ be a non constant polynomial of minimal degree with respect to $\tilde{q}(T) = 0$. Let b be the leading coefficient of $\tilde{q}$ and $q = \frac{1}{b}\tilde{q}$. Then $q(T) = \frac{1}{b}\tilde{q}(T) = 0$ and $\deg q = \deg \tilde{q} = \deg T$. Thus q is a minimal polynomial for T.

Before proving the uniqueness we will show that (b) holds. Let $p = sq + r$ with $\deg s < \deg q$. Then

$$p(T) = s(T)q(T) + r(T) = r(T).$$

Suppose first that q divides p. Then $r = 0$ and so $p(T) = 0$.

Suppose next that $p(T) = 0$. Then $r(T) = 0$. Since $\deg r < \deg q$, the minimality of $\deg q$ implies that $r = 0$. Thus q divides p and (b) holds.

Now suppose that also q^* is a minimal polynomial. Then by (b), $q^* = sq$ for some $s \in \mathcal{P}(\mathbb{F})$. Since q and q^* have the same degree, s is a constant. And since q and q^* are monic we conclude that $s = 1$. Thus $q = q^*$ and q is the unique minimal polynomial for T.

Proposition 8.16 [**eigenvalues are the roots of the minimal polynomial**] *Let V be finite dimensional, $T \in \mathcal{L}(V)$, $\lambda \in \mathbb{F}$ and q the minimal polynomials of T. Let m be the multiplicity of λ as a root of q. Then*

(a) *m equals is the maximal height of a generalized eigenvector corresponding to λ*

(b) *λ is an eigenvalue for T if and only if λ is a root of q*

Proof: (a) Let $q = (x - \lambda^m)p$ with $p \in \mathcal{P}(\mathbb{F})$. Then λ is not a root of p.

Let v be a generalized eigenvector of T with respect to λ and define $w = (T - \lambda)^m(v)$. Then

$$p(T)(w) = p(T)((T - \lambda)^m)(v)) = q(T)(v) = 0.$$

Note also that w is a generalized eigenvector corresponding to λ. So by 8.14 $w = 0$. Thus v has height at most m. It remains to show that there exists a generalized eigenvector of height m.

If $m = 0$, 0_V has height m. So suppose that $m > 0$ a put $r(T) = (x - \lambda)^{m-1}p$. Since q has minimal degree with respect to $q(T) = 0$, $r(T) \neq 0$. Hence there exists $w \in V$ with $r(T)(w) \neq 0$. But $v = p(T)(w)$. Then

$$(T - \lambda)^{m-1}(v) = (T - \lambda)^{m-1}(p(T)(w)) = r(T)(w) \neq 0$$

and

$$(T - \lambda)^m(v) = (T - \lambda)^m(p(T)(w)) = q(T)(w)0$$

thus v is a generalized eigenvector of height m and (a) is proved.

(b) λ is a root if and only if $m \geq 1$. λ is an eigenvalue if and only if T has a generalized eigenvector of height at least 1. So (b) follows from (a). □

Proposition 8.17 [**decomposing V**] *Let V be finite dimensional,$T \in \mathcal{L}(V)$, $\lambda \in \mathbb{F}$ and q the minimal polynomials of T. Let $p \in \mathcal{P}(\mathbb{F})$ and $m \in \mathbb{N}$ with $q = (x - \lambda)^m p$ and $p(\lambda) \neq 0$. Then*

(a) $V = \text{null}\,(T - \lambda)^m \oplus \text{null}\,p(T)$

(b)
$$\text{null}\,(T - \lambda)^m = \text{range}\ p(T) \quad \textit{and} \quad \text{null}\,p(T) = \text{range}\,(T - \lambda)^m$$

(c) *p is the minimal polynomial of the restriction of T to* $\text{null}\,p(T)$.

Proof:

Since $q(T) = 0$, $(T - \lambda)^m(p(T)(v)) = 0$ for all $v \in V$. Thus $\text{range}\,p(T) \leq \text{null}(T - \lambda)^m$. By 8.14

$$\text{null}(T - \lambda)^m \cap \text{null}\,p(T) \neq 0$$

Hence $\dim \operatorname{null}(T-\lambda)^m + \dim \operatorname{null} p(T) \leq \dim V$ But $\operatorname{range} p(T) \leq \operatorname{range} \operatorname{null}(T-\lambda)^m$ implies

$$\dim V = \dim \operatorname{range} p(T) + \dim \operatorname{null} p(T) \leq \dim \operatorname{null}(T-\lambda)^m + \dim \operatorname{null} p(T) \leq \dim V$$

Hence equality must hold. It follows that

$$\operatorname{range} p(T) = \operatorname{null}(T-\lambda)^m \quad V = \operatorname{null} \operatorname{range} \operatorname{null}(T-\lambda)^m \oplus \operatorname{null} q$$

Since $q(T) = 0$ we also get $\operatorname{range}(T-\lambda)^m \leq \operatorname{null} p(T)$ Since

$$\dim \operatorname{range}(T-\lambda)^m = \dim V - \dim \operatorname{null}(T-\lambda)^m = \dim V - \dim \operatorname{range} p(T) = \dim \operatorname{null} p(T)$$

we conclude that

$$\operatorname{range}(T-\lambda)^m = \operatorname{null} p(T)$$

Thus (a) and (b) hold.

Let $W = \operatorname{null} T$ and let r be the minimal polynomial for $T|_W$. Put $s = (x-\lambda)^m r$ and let $v \in V$. By (b) $(T-\lambda)^m(v) \in \operatorname{range}(T-\lambda)^m \leq W$ and so

$$0 = r(T)(T-\lambda)m(v) = s(T)(v)$$

Thus $s(T) = 0$ and so $\deg s \geq \deg q$. Hence $\deg r = \deg s - m \geq \deg q - m = \deg p$.

Also $p(T|_W) = p(T)|_W = 0$. Thus the minimality of $\deg r$ implies $\deg r = \deg p$ and p is the minimal polynomial of $T|_W$. □

Theorem 8.18 [**decomposing V when q splits**] *Let V be finite dimensional,$T \in \mathcal{L}(V)$ and q the minimal polynomials of T. Suppose that there exists distinct $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ and positive integers $m_1, m_2, \ldots, m_k$ such that*

$$q = (x-\lambda_1)^{m_1} \cdot (x-\lambda_2)^{m_2} \cdot \ldots (x-\lambda_k)^{m_k}.$$

(Note that such a factorization always exists if $\mathbb{F}$ is algebraically closed) Then

$$V = \bigoplus_{i=1}^{k} V_{T,\lambda_i,m_i}$$

Proof: By induction on k. If $k = 1$ then $q = (x-\lambda_1)^{m_1}$ and $q(T) = 0$ implies $V = \operatorname{null} q(T) = V_{T,\lambda_1,m_1}$.

Suppose now that $k > 1$ and let $p = \prod_{i=2}^{k}(x-\lambda_i)^{m_i}$. Then $q = (x-\lambda_1)^{m_1} p$ and $p(\lambda_1) \neq 0$. So by 8.17

$$V = \operatorname{null}(T-\lambda_1)^{m_1} \oplus \operatorname{null} p(T).$$

Put $W = \operatorname{null} p(T)$. Since $\operatorname{null}(T-\lambda_1)^{m_1} = V_{T,\lambda_1,m_1}$ we have

$$V = V_{T,\lambda_1,m_1} \oplus W \tag{1}$$

Then 8.17 (c) says that p is the minimal polynomial of $T|_W$. Thus by induction

$$W = \bigoplus_{i=2}^{k} W_{T,\lambda_i,m_i} \tag{2}$$

Let $2 \leq i \leq k$ and $v \in V_{T,\lambda_i,m_i}$. Then $p = r(x-\lambda_i)^{m_i}$ for $r \in \mathcal{P}(\mathbb{F})$ and thus

$$p(T)(v) = r(T)((T-\lambda_i)^{m_i}(v)) = r(T)(0) = 0$$

Thus $v \in \operatorname{null} p(T) = W$ and we conclude that

$$V_{T,\lambda_i,m_i} = W_{T,\lambda_i,m_i} \tag{3}$$

The theorem now follows from (1),(2) and (3). □

Theorem 8.19 [**Jordan canonical form**]*Let V be finite dimensional, $T \in \mathcal{L}(V)$ and q the minimal polynomials of T. Suppose that there exists distinct $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ and positive integers $m_1, m_2, \ldots, m_k$ such that*

$$q = (x-\lambda_1)^{m_1} \cdot (x-\lambda_2)^{m_2} \cdot \ldots (x-\lambda_k)^{m_k}.$$

(Note that such a factorization always exists if $\mathbb{F}$ is algebraically closed) Then there exists a basis $\mathfrak{b}$ for V, positive integers $n_1, n_2 \ldots, n_s$ and (not necessarily distinct) $\mu_1, \mu_2 \ldots, \mu_s \in \mathbb{F}$ such that the matrix for T with respect to $\mathfrak{b}$ has the form

$$J(\mu_1,n_1;\mu_2,n_2;\ldots;\mu_s,n_s) =: \begin{pmatrix} J(\mu_1,n_1) & 0 & \ldots & 0 & 0 \\ 0 & J(\mu_2,n_2) & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & J(\mu_{s-1},n_{s-1}) & 0 \\ 0 & 0 & \ldots & 0 & J(\mu_s,n_s) \end{pmatrix}$$

where $J(\mu,n) = J(n) + \mu I_n$ denotes the matrix

$$J(\mu,n)=\begin{pmatrix} \mu & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mu & 1 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \mu & 1 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & \mu & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & \mu \end{pmatrix}$$

Proof: Suppose first that $V = V_1 \oplus V_2$ for some non-zero T-invariant subspaces V_1, V_2 in V. Then by induction there exists a basis $\mathfrak{b}_i$ for T so that the matrix for $T|_{V_i}$ is

$$A_i = J(\mu_{i,1}, n_{i,1}; \mu_{i,2}, n_{i,2}; \dots ; \mu_{i,s_i}, n_{s_i})$$

Let $\mathfrak{b} = (\mathfrak{b}_1, \mathfrak{b}_2)$. Then $\mathfrak{b}$ is a basis for V and the matrix for T with respect to to T is

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} = J(\mu_{1,1}, n_{1,1}; \dots ; \mu_{1,s_1}, n_{s_1}; \mu_{2,1}, n_{2,1}; \dots ; \mu_{2,s_2}, n_{s_2}))$$

and so we are done in this case.

So we may assume that no such decomposition of V exists. By 8.18

$$V = \bigoplus_{i=1}^{k} V_{T,\lambda_i,m_i}$$

By 8.4b, each of the V_{T,λ_i,m_i} is T-invariant. Since we assume that V cannot be decomposed we conclude that $k = 1$. So $q = (x-\lambda)^m$ where $\lambda = \lambda_1$ and $m = m_1$. It follows that $(T-\lambda)^m = 0$ and so $T - \lambda$ is nilpotent. In particular we can apply 8.11 to $T - \lambda$. Hence by part (b) of 8.11

$$V = \bigoplus_{i,j} U_{i,j}$$

where each $U_{i,j}$ is $T - \lambda$ invariant. Then $U_{i,j}$ is T-invariant and our assumption that V cannot be decomposed implies that $V = U_{i,j}$ for some i, j. Thus by 8.11(c) there exists a basis for $V = U_{i,j}$ so that the matrix of $T - \lambda$ is $J(n)$. Since $T = \lambda + (T - \lambda)$ we conclude that the matrix for T is $\lambda I_n + J(n) = J(\lambda, n)$. □

Lemma 8.20 [**diagonal form**] *Suppose that V is finite dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

(a) *With respect to some basis of V the matrix of T is diagonal, that is of the form*

$$\begin{pmatrix} \mu_1 & 0 & \dots & 0 & 0 \\ 0 & \mu_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \mu_{n-1} & 0 \\ 0 & 0 & \dots & 0 & \mu_n \end{pmatrix}$$

for some not necessarily distinct $\mu_1, \dots, \mu_n$ in $\mathbb{F}$.

(b) *There exists a basis consisting of basis eigenvectors of T.*

(c) *$\lambda_1, \lambda_2, \lambda_k$ be the distinct eigenvalues for T. Then*

$$V = \bigoplus_{i=1}^{l} V_{T,\lambda_i}$$

(d) *Let q be the minimal polynomial for T. Then*

$$q = (x - \lambda_1) \cdot \ldots \cdot (x - \lambda_k)$$

for some, distinct $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{F}$.

Proof: (a)$\Longleftrightarrow$ (b):

Let $(v_1, v_2, \dots, v_n)$ be a basis for V. The matrix for T with respect to this basis is diagonal if and only if $T(v_i) = d_i v_i$ for some $d_i \in \mathbb{F}$ and all $1 \leq i \leq n$. That is its diagonal if and only if each v_i is an eigenvector. So (a) and (b) are equivalent.

(b)$\Longrightarrow$ (c):

Suppose that $(v_i, 1 \leq i \leq n)$ is a basis of eigenvectors. Let $v \in V$ we will first show that $V = \sum_{j=1}^{k} u_j$ where u_j is eigenvector of T corresponding to λ_i. By assumption $v = \sum_{i=1}^{n} a_i v_i$ for some $a_i \in F$. For $1 \leq j \leq k$ let

$$I_j = \{1 \leq i \leq n \mid v_i \text{ is an eigenvector corresponding to} \lambda_j$$

and

$$u_j = \sum_{i \in I_j} a_i v_i$$

Since each i lies in exactly one I_j

$$v = \sum_{i=1}^{n} a_i v_i = \sum_{j=1}^{k} \sum_{i \in I_j} a_i v_i = \sum_{j=1}^{j} u_j$$

Also since eigenspaces are subspaces $u_j \in V_{T,\lambda_j}$.

Hence $V = \sum_{j=1}^{k} V_{T,\lambda_j}$ By 8.8 and 4.2 we conclude that $V = \bigoplus_{j=1}^{k} V_{T,\lambda_j}$.

(c)$\Longrightarrow$ (b):

Just choose a basis for each V_{T,λ_i}. Then problem # 1 on Homework set 4 gives us a basis of V consisting of eigenvectors.

(c)$\Longrightarrow$ (d):

Let $p = \prod_{i=1}^{k}(x - \lambda_i)$. Let $v \in V$. Then $V = \sum_{i=1}^{k} u_i$ with $u_i \in V_{T,\lambda_i}$. By 8.6 $p(T)(u_i) = p(\lambda_i)u_i = 0$. Hence

$$p(T)(v) = \sum_{i=1}^{l} p(T)(u_i) = 0$$

Thus $p(T) = 0$. From 8.15(b), q divides p. By 8.16 each λ_i is a root of q. Hence also p divides q. Since both q and p are monic we conclude that $p = q$.

(d)$\Longrightarrow$ (c):

This follows directly from 8.18 □

9 Hermitian Forms

Throughout this section we assume:

Hypothesis 9.1 (a) $\mathbb{F}$ *is a field.*

(b) V *vector space over* $\mathbb{F}$.

(c) $^{-} : \mathbb{F} \to \mathbb{F}$ *is a function such that*

(ca) $\overline{a+b} = \overline{a} + \overline{b}$ *for all* $a, b \in \mathbb{F}$

(cb) $\overline{ab} = \overline{a}\overline{b}$ *for all* $a, b \in \mathbb{F}$

(cb) $\overline{\overline{a}} = a$ *for all* $a \in \mathbb{F}$.

(d) $\mathbb{K} = \{k \in \mathbb{F} \mid k = \overline{k}\}$.

Note that (cb) implies that $^{-}$ is a bijection. We will refer to $^{-}$ as conjugation. $\overline{a}$ is called the conjugate of a.

For example if $\mathbb{F} = \mathbb{C}$ we can choose $\overline{a+ib} = a - ib$ for all $a, b \in \mathbb{R}$.

If $\mathbb{F}$ is an arbitrary field with $\operatorname{char} F \neq 2$ we can choose $^{-} = \mathrm{id}_{\mathbb{F}}$.

Let $T : V \to W$ be a function. We say that T is semilinear of $T(v+w) = T(v) + T(w)$ and $T(av) = \overline{a}T(v)$ for all $v, w \in W, a \in \mathbb{F}$.

For example $T : \mathbb{F}^2 \to \mathbb{F}, \quad (a, b) \to (2\overline{a} + 3\overline{b})$ is semilinear.

Definition 9.2 *Let V be a vector space. An hermitian form on V is a function*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}, \quad (u, v) \to \langle u, v \rangle$$

such that

(a) $\langle u + v, w \rangle = \langle u, v \rangle + \langle v, w \rangle$ *for all* $u, v, w \in V$

(b) $\langle av, w \rangle = a\langle v, w \rangle$ *for all* $a \in \mathbb{F}$, $v, w \in W$.

(c) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ *for all* $v, w \in V$.

For example $\langle (a_1, \dots, a_n), (b_1, \dots, b_n) \rangle = a_1\overline{b_1} + a_2\overline{b_2} \dots a_n\overline{b_n}$ is a hermitian form on $\mathbb{F}^n$.

An hermitian space is a vector space V together with an hermitian form $\langle \cdot, \cdot \rangle$ From now on $(V, \langle \cdot, \cdot \rangle)$ is an hermitian space.

Lemma 9.3 [**hermitian=sesquilinear**] *Let V be an hermitian space.*

(a) $\langle \cdot, \cdot \rangle$ *is linear in the first coordinate, that is for all $w \in V$ the function:*

$$\langle \cdot, w \rangle : V \to \mathbb{F}, \quad v \to \langle v, w \rangle$$

is linear.

(b) $\langle \cdot, \cdot \rangle$ *is semilinear in the second coordinate, that is for all $v \in V$ the function*

$$\langle v, \cdot \rangle : V \to \mathbb{F}, \quad w \to \langle v, w \rangle$$

is semilinear.

Proof: (a) follows immediately from the definition of a hermitian form.

(b)

$$\langle v, u + w \rangle = \overline{\langle u + w, v \rangle} = \overline{\langle u, v \rangle + \langle w, v \rangle} = \overline{\langle u, v \rangle} + \overline{\langle w, v \rangle} = \langle v, u \rangle + \langle v, w \rangle$$

and

$$\langle v, aw \rangle = \overline{\langle aw, v \rangle} = \overline{a\langle w, v \rangle} = \overline{a}\overline{\langle w, v \rangle} = \overline{a}\langle v, w \rangle$$

□

We say that v, w are orthogonal (or perpendicular) if $\langle v, w \rangle = 0$. We also write $v \perp w$ in this case. Note that $v \perp w$ if and only $w \perp v$. For $v \in V$ let $v^\perp$ the set of vectors orthogonal to v. Note that $v^\perp$ is a subspace of v. Indeed $v^\perp$ is the null space of the linear map $\langle \cdot, v \rangle$. More generally for $U \subseteq V$ define

$$U^\perp = \{v \in V \mid \langle u, v \rangle = 0, \forall u \in U\} = \bigcap_{u \in U} u^\perp$$

Since intersections of subspaces are subspaces $U^\perp$ is a subspace of V.

For $v \in V$ define the (square) norm of v to be $\|v\|^2 = \langle v, v \rangle$. Note that $\overline{\langle v, v \rangle} = \langle v, v \rangle$ so the square norm of a vector is always fixed by $^-$. That is $\|v\|^2 \in \mathbb{K}$ for all $v \in V$. It might be interesting to observe that $\mathbb{K}$ is a subfield of $\mathbb{F}$. Also $\mathbb{K} = \mathbb{F}$ id and only if $^- = \mathrm{id}_\mathbb{F}$. For $\lambda \in \mathbb{F}$ we define $\|\lambda\|^2 = \lambda\overline{\lambda}$. Note that $\|\lambda v\|^2 = \|\lambda\|^2\|v\|^2$.

Lemma 9.4 [**Pythagorean Theorem**] *Let u, v be orthogonal vectors in V. Then*

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof:

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle u,u\rangle = \langle u,u\rangle + \langle v,v\rangle = \|u\|^2 + \|v\|^2$$

□

Lemma 9.5 [**orthogonal projection**] *Let $v \in V$ with $\|v\|^2 \neq 0$.*

(a) *Let $u \in V$ and put $w = u - \frac{\langle u,v\rangle}{\|v\|^2}v$. Then*

$$v \perp w$$

and

$$u = \frac{\langle u,v\rangle}{\|v\|^2}v + w$$

(b) $V = \mathbb{F}v \oplus v^\perp$

Proof: (a)

$$\langle w,v\rangle = \langle u,v\rangle - \frac{\langle u,v\rangle}{\|v\|^2}\langle v,v\rangle = \langle u,v\rangle - \langle u,v\rangle = 0$$

Thus $v \perp w$. The second statement follows directly from the definition of w.

(b) By (a) $V = \mathbb{F}v + v^\perp$. Suppose that $av \in v^\perp$ for some $a \in \mathbb{F}$. Then

$$0 = \langle av, v\rangle = a\|v\|^2$$

Since $\|v\|^2 \neq 0$ we conclude that $a = 0$ and so $\mathbb{F}v \cap v^\perp = 0$ and (b) holds. □

$\frac{\langle u,v\rangle}{\|v\|^2}v$ is called the orthogonal projection of u onto $\mathbb{F}v$. w is called the orthogonal projection of u onto $v^\perp$.

Lemma 9.6 [**Cauchy Schwarz**] *Let $u, v \in V$ with $\|v\|^2 \neq 0$. Let w be the projection of u onto $v^\perp$. Then*

$$\|\langle u,v\rangle\|^2 + \|w\|^2\|v\|^2 = \|u\|^2\|v\|^2$$

Proof: Let w be as in 9.5(a). Then we have $v \perp w$ and $u = \frac{\langle u,v\rangle}{\|v\|^2}v + w$. Thus by the Pythagorean Theorem

$$\|u\|^2 = \|\frac{\langle u,v\rangle}{\|v\|^2}v\|^2 + \|w\|^2 = \frac{\|\langle u,v\rangle\|^2}{\|v\|^2} + \|w\|^2$$

Multiplying both sides with of this equation with $\|v\|^2$ we get the result. □.

Definition 9.7 *A inner product space is an hermitian space such that*

(a) *Either* $\mathbb{F} = \mathbb{C}$ *and* $^-$ *is complex conjugation, or* $\mathbb{F} = \mathbb{R}$ *and* $^- = \mathrm{id}_{\mathbb{R}}$.

(b) $\|v\|^2 \geq 0$ *for all* $v \in V$.

(c) $\|v\|^2 = 0$ *if and only if* $v = 0$.

Note that given (a) we have $\mathbb{K} = \mathbb{R}$ and so $\|v\|^2 \in \mathbb{R}$ for all $v \in V$. So (b) makes sense. In view of (b) we can define $\|v\| = \sqrt{\|v\|^2}$. We also define $\|\lambda\| = \sqrt{\|\lambda\|^2}$.

Theorem 9.8 [**Cauchy Schwarz inequality**] *Let* V *be an inner product space and* $u, v \in V$. *Then*

$$\|\langle u, v\rangle\| \leq \|u\|\|v\|.$$

Equality holds if and only if (u, v) *is linearly dependent.*

Proof: If $\|v\|^2 = 0$ then $v = 0$ and both side of the equality are 0. Note also that (u, v) is linearly dependent in this case.

So suppose that $\|v\|^2 \neq 0$. Then by 9.6

$$\|\langle u, v\rangle\|^2 + \|w\|^2\|v\|^2 = \|u\|^2\|v\|^2$$

But $\|w\|^2\|v\|^2 \geq 0$. So $\|\langle u, v\rangle\|^2 \leq= \|u\|^2\|v\|^2$. Taking square roots gives the Cauchy Schwarz inequality.

Equality holds if and only if $\|w\|^2 = 0$. That is if $w = 0$. By 9.5 this is equivalent to $u \in \mathbb{F}v$. □

Lemma 9.9 [**Triangle inequality**] *Suppose* V *is an inner product space and let* $u, v \in V$. *Then*

$$\|u + v\| \leq \|u\| + \|v\|.$$

Equality holds if and only if one of u *and* v *is a non-negative real multiple of the other.*

Proof: Note first that the following are equivalent:

$$\|u + v\| \leq \|u\| + \|v\|.$$

$$\|u + v\|^2 \leq (\|u\| + \|v\|)^2$$

$$\|u\|^2 + \langle u, v\rangle + \langle v, u\rangle + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2$$

$$\langle u, v\rangle + \overline{\langle u, v\rangle} \leq 2\|u\|\|v\| \tag{1}$$

Let $a = x + iy \in \mathbb{C}$. Then

$$a + \overline{a} = 2x \leq 2\sqrt{x^2 + y^2} = 2\|a\|$$

with equality if and only if $a = x$ is a non-negative real number. Together with the Cauchy Schwarz inequality we obtain

$$\langle u, v\rangle + \overline{\langle u, v\rangle} \leq 2\|\langle u, v\rangle\| \leq 2\|u\|\|v\|$$

This proves the triangle inequality. Moreover, equality holds if and only if $\langle u, v\rangle$ is a non-negative real number and (u, v) is linearly dependent. Note that $\langle \lambda v, v\rangle = \lambda\|v\|^2$ is non-negative if and only if $v = 0$ or $\lambda \geq 0$. So equality holds if and only if $v = 0$ or u is a non-negative real multiple of v. □

Lemma 9.10 *perp of a span Let V be an hermitian space.*

(a) *If $U \subseteq W \subseteq V$, then $W^\perp \leq U^\perp$.*

(b) *If $U \subseteq V$ then $U^\perp = (\operatorname{Span} U)^\perp$.*

Proof: (a) is obvious. For (b) note that $U^{\perp\perp}$ is a subspace of V containing U. So $\operatorname{Span} U \leq U^{\perp\perp}$. Thus

$$U^\perp \leq (\operatorname{Span} U)^\perp$$

Since $U \subseteq \operatorname{Span} U$ we also have

$$(\operatorname{Span} U)^\perp \leq U^\perp$$

□

We say that V is non-degenerate if $V^\perp = \{0_V\}$. Note that V is non-degenerate if and only if for each $0 \neq v \in V$ there exists $w \in V$ with $\langle v, w\rangle \neq 0$.

Lemma 9.11 [**u=v**] *Let V be a non-degenerate hermitian space and $u, v \in V$. Then $u = v$ if and only if*

$$\langle u, w\rangle = \langle v, w\rangle \quad \forall w \in V.$$

Proof: The one direction is obvious. Suppose now that $\langle u, w\rangle = \langle v, w\rangle = 0$ for all $w \in V$. Then $\langle u - v, w\rangle = 0$ for all $w \in V$ and since V is non-degenerate $u - v = 0$. Hence $u = v$.□

Definition 9.12 *A list of vectors $(v_1, v_2, \ldots, v_n)$ is called orthogonal if for all $1 \leq i, j \leq n$*

$$\langle v_i, v_j\rangle = 0 \Longleftrightarrow i \neq j$$

Lemma 9.13 [**orthogonal lists are linearly independent**] *Any orthogonal list of vectors is linearly independent.*

Proof: Let $(v_1, \dots, v_n)$ be an orthogonal list of vectors and suppose that $\sum a_i v_i = 0$ for some $a_i \in \mathbb{F}$. Then

$$0 = \langle 0, v_j \rangle = \langle \sum_i a_i v_i, v_j \rangle = a_j \langle v_j, v_j \rangle$$

By the definition of an orthogonal list $\langle v_j, v_j \rangle \neq 0$. Thus $a_j = 0$ for all j and $(v_1, \dots, v_n)$ is linearly independent. □

Lemma 9.14 [**symplectic forms in char 2**] *Suppose that V is a non-zero, non-degenerate hermitian space such that $\|v\|^2 = 0$ for all $v \in V$. Then either $\langle \cdot, \cdot \rangle = 0$ or $\operatorname{char} F = 2$ and $^- = \mathrm{id}_{\mathbb{F}}$.*

Proof: We may assume that $\langle \cdot, \cdot \rangle \neq 0$. Then there exist $v, w \in V$ with $\langle w, v \rangle \neq 0$. Replacing w by $\langle w, v \rangle^{-1} w$ we may assume that $\langle w, v \rangle = 1$. Then also $\langle v, w \rangle = 1$. Let $a \in F$. Since all vectors have square norm 0 we compute

$$0 = \langle aw + v, aw + v \rangle = a + \bar{a}$$

Thus $\overline{a} = -a$ for all $a \in \mathbb{F}$. Choosing $a = 1$ we see that $\operatorname{char} \mathbb{F} = 2$. Thus $\overline{a} = a$ for all $a \in \mathbb{F}$ and so $^- = \mathrm{id}_{\mathbb{F}}$. □

In view of the preceding lemma we call an hermitian space an hermitian+ space if either $\operatorname{char} \mathbb{F} \neq 2$ or $^- \neq \mathrm{id}_{\mathbb{F}}$. The proceeding lemma now reads:

Corollary 9.15 [**non-singular vectors**] *Let V be a hermitian+ space with $\langle \cdot, \cdot \rangle \neq 0$. Then there exists $v \in V$ with $\|v\|^2 \neq 0$.*

Corollary 9.16 [**equality of hermitian forms**] *Let $\langle \cdot, \cdot \rangle_i$, $i = 1, 2$ be hermitian+ forms on V. Then $\langle \cdot, \cdot \rangle_1 = \langle \cdot, \cdot \rangle_2$ if and only if $\| \cdot \|_1^2 = \cdot_2$.*

Proof: Define $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_1 - \langle \cdot, \cdot \rangle_2$. The is is readily verified that $\langle \cdot, \cdot \rangle$ is an hermitian plus form and $\| \cdot \|^2 = \| \cdot \|_1^2 - \| \cdot \|_2^2$. By 9.15, $\langle \cdot, \cdot \rangle = 0$ if and only if $\| \cdot \|^2 = 0$. But this implies the desired result. □

Theorem 9.17 [**orthogonal basis**] *Every finite dimensional, non-degenerate hermitian+ space has an orthogonal basis.*

Proof: The proof is by induction on $\dim V$. If $\dim V = 0$ the empty list is an orthogonal basis. So suppose that $\dim V > 0$. By 9.15 there exists $v \in V$ with $\|v\|^2 \neq 0$. Let $W = v^\perp$. By 9.5 $V = \mathbb{F}v \oplus W$. We claim that W is non-degenerate. Indeed let $0 \neq w \in W$. Since V is non-degenerate, $\langle u, w \rangle \neq 0$ for some $u \in V$. Then $u = av + x$ for some $a \in \mathbb{F}$ and $x \in W$. Since $v \perp w$ we get

$$\langle x, w \rangle = \langle u, w \rangle \neq 0$$

and so W is non-degenerate. By induction, W has an orthogonal basis $\mathfrak{w}$. Then $(v, \mathfrak{w})$ is an orthogonal basis for V. □

A list of vectors is called orthonormal if its orthogonal and all vectors in the list have square norm 1.

Corollary 9.18 [**orthonormal basis**] *Every finite dimensional inner product space has an orthonormal basis*

Proof: Note that every inner product space is an non-degenerate hermitian+ space. So by 9.17 V has an orthogonal basis $(u_1, \dots, u_n)$. Let $v_i = \frac{u_i}{\|u_i\|}$. Then $\|v_i\|^2 = 1$ and $(v_1, \dots, v_n)$ is an orthonormal basis. □

A list of subspaces $(X_1, X_2, \dots, X_n)$ of V is called orthogonal if each X_i is non-degenerate and $X_i \perp X_j$ for all $i \neq j$. If this is the case we will write

$$\bigoplus_{i=1}^{n} X_i$$

for $\bigoplus_{i=1}^{n} X_i$.

Theorem 9.19 [**orthogonal decomposition**] *Let V be a non-degenerate hermitian+ space and U a finite dimensional, non-degenerate subspace. Then*

$$V = U \oplus U^{\perp}.$$

Proof: By 9.17 U has an orthogonal basis $(u_1, \dots, u_m)$. Let $v \in V$. Let x_i be the orthogonal projection of v onto $\mathbb{F}w_i$. Put $x = \sum_{i=1}^{n} x_i$ and $w = v - x$. Then

$$w = (v - x_i) - \sum_{j \neq i} x_j$$

Note that u_i is orthogonal to $v - x_i$ and to $x_j \in \mathbb{F}u_j$. Thus $u_i \in w^{\perp}$. Since $w^{\perp}$ is a subspace and U is spanned by $(u_1, \dots, u_n)$ we conclude that $U \leq w^{\perp}$. Thus $w \in U^{\perp}$. Also $v = x + w$ and so $V = U + U^{\perp}$. Since U is non-degenerate, $U \cap U^{\perp} = 0$ and so $V = U \oplus U^{\perp}$. It remains to show that $U^{\perp}$ is non-degenerate.

For this let $w \in U^{\perp}$. Then $\langle v, w \rangle \neq 0$ for some $v \in V$. Write $v = u + y$ with $u \in U$, $y \in Y$. Since $u \perp w$ we get

$$\langle y, w \rangle = \langle v, w \rangle \neq 0$$

and $U^{\perp}$ is non-degenerate. □

10 The dual space, adjoint operators and the Spectral Theorem

Let V be a vector space. Define $V^* = \mathcal{L}(V, \mathbb{F})$. V^* is called the dual space of V. The elements of V^* are called linear functional. So a linear functional ϕ on V is a linear map

$$\phi : V \to \mathbb{F}.$$

Lemma 10.1 [**dual basis**] *Suppose that $\mathfrak{v} = (v_1, \ldots, v_n)$ is a basis for the finite dimensional vector space V. Define $v_j^* \in V^*$ by*

$$v_j^*(\sum_i a_i v_i) = a_j.$$

(a)

$$v_j^*(v_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

(b) *Let $\phi \in V^*$, then*

$$\phi = \sum_i \phi(v_i) v_i^*.$$

(c) *$\mathfrak{v}^* = (v_1^*, v_2^*, \ldots, v_n^*)$ is a basis for V^*.*

Proof: (a) follows immediately from the definition of v_j^*.

(b)& (c) Let $a_i \in \mathbb{F}$ for $1 \leq i \leq n$. Then using that $\mathfrak{v}$ spans V and using (a) the following are equivalent:

$$\begin{aligned} \phi &= \textstyle\sum_j a_j v_j^* \\ \phi(v_i) &= (\textstyle\sum_j a_j v_j^*)(v_i) \quad \forall 1 \leq i \leq n \\ \phi(v_i) &= \textstyle\sum_j a_j v_j^*(v_i) \quad \forall 1 \leq i \leq n \\ \phi(v_i) &= a_i \quad \forall 1 \leq i \leq n \end{aligned}$$

So (b) holds. Also each ϕ is a unique linear combination of $\mathfrak{v}^*$ and so (c) holds. □

The basis $\mathfrak{v}^*$ of V is called the basis dual to $\mathfrak{n}v$. Note that (for finite dimensional V) the map

$$\sum_i a_i v_i \to \sum_i a_i v_i^*$$

is an isomorphism form V to V^*. But the reader should realize that this isomorphism depends on the choice of the basis $\mathfrak{v}$. In contrast the next lemma shows that (for finite dimensional V) there exists a canonical isomorphism between V and $V^{**} = (V^*)^*$, the double dual of V.

Lemma 10.2 [**embedding of V in V****] *For $v \in V$ define $\hat{v} \in V^{**}$ by*

$$\hat{v}(\phi) = \phi(v)$$

for all $\phi \in V$.

Then the map

$$\hat{\ } : V \to V^{**}, v \to \hat{v}$$

is linear.

If V is finite dimensional $\hat{\ }$ is an isomorphism.

Proof: We first need to verify that $\hat{v}$ is indeed a linear functional on V^*. Let $S, T \in V^*$ and $a \in \mathbb{F}$. Then

$$\hat{v}(T+S) = (T+S)(v) = T(v) + S(v) = \hat{v}(T) + \hat{v}(S)$$

$$\hat{v}(aT) = (aT)(v) = aT(v) = a\hat{v}(T)$$

So $\hat{v}$ is a linear functional. To show that $\hat{\ }$ is linear, let $u, v \in V$, $a \in \mathbb{F}$ and $T \in V^*$. Then

$$\widehat{u+v}(T) = T(u+v) = T(u) + T(v) = \hat{u}(T) + \hat{v}(T) = (\hat{u} + \hat{v})(T)$$

Since this holds for all $T \in V^*$, the linear functionals $\widehat{u+v}$ and $\hat{u} + \hat{v}$ are equal.

$$\widehat{av}(T) = T(av) = aT(v) = a\hat{v}(T) = (a\hat{v})(T)$$

and so $\widehat{av} = a\hat{v}$. So $\hat{\ }$ is indeed linear.

Suppose now that V is finite dimensional. To show that the map is an isomorphism we explicitly compute it in terms of a basis $\mathfrak{v} = (v_1, \ldots, v_n)$. Then

$$\hat{v_j}(v_i^*) = v_i^*(v_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

It follows that $(\hat{v_1}, \hat{v_2}, \ldots, \hat{v_n})$ is the basis of V^{**} dual to $\mathfrak{v}^*$. In particular, $\hat{\ }$ sends a basis of V to a basis of V^{**} and so is an isomorphism. □

We remark that for infinite dimensional W, $\hat{\ }$ is still one to one but no longer onto. But since the proof depends on the axiom of choice it is not presented.

Lemma 10.3 [**V*perp=0**] *Let V be a finite dimensional vector space and $U \subseteq V$ with $U \neq V$. Then there exists $0 \neq \phi \in V^*$ with $\phi|_U = 0$.*

Proof: Since a basis of U extends to a basis for V we can choose a basis $(v_1, \ldots, v_n)$ of V such that $(v_1, \ldots, v_m)$ is a basis for U. Let $\phi = v_n^*$. Since $U \neq V$, $n \neq m$. Thus $\phi(v_i) = 0$ for all $1 \leq i \leq m$ and $\phi|_U = 0$. □

If A is a matrix then A^{T} denotes the transpose of A, that is the (i,j) coefficient of A^{T} is the (j,i) coefficient of A.

Lemma 10.4 [**dual of a linear map**] *Let $S \in \mathcal{L}(V,W)$. Define*

$$S^* : W^* \to V^*, \quad T \to T \circ S$$

The S^* is a well defined linear map. Moreover, the map

$$^* : \mathcal{L}(V,W) \to \mathcal{L}(W^*,V^*), \quad S \to S^*$$

is linear.

Proof: Since compositions of linear maps are linear, $T \circ S$ is indeed a linear map from $V \to \mathbb{F}$ and so $T \circ S \in V^*$. So S^* is well defined. The rest of the proof is very similar to the proof of 10.2 and we leave the details to the reader. □

Let $\mathfrak{v} = (v_1, \ldots, v_n)$. We denote the dual basis for V^* by $\mathfrak{v}^* = (v_1^*, \ldots, v_n^*)$. We observe

Lemma 10.5 [**matrix of the dual map**] *Let $T : V \to W$ be linear and suppose that $\mathfrak{v}$ and $\mathfrak{w}$ are bases for V and W, respectively. Let A be the matrix of T with respect to $\mathfrak{v}$ and $\mathfrak{w}$ and B the matrix of T^* with respect to $\mathfrak{w}^*$ and $\mathfrak{v}^*$. Then $B = A^{\mathrm{T}}$*

Proof: $T^*(w_j^*)(v_i) = (w_j^* \circ T)(v_i) = w_j^*(T(v_i)) = w_j^*(\sum_k a_{ki} w_k) = a_{ji}$ Thus by 10.1(b)

$$T^*(w_j^*) = \sum_i a_{ji} v_i^*.$$

Hence $b_{ij} = a_{ji}$. □

The next lemma allows to apply most of our proposition about linear maps to semilinear maps.

Lemma 10.6 [**turning semilinear maps into linear maps**] *Let V be a vector space over $\mathbb{F}$.*

(a) *Define a new scalar multiplication*

$$\diamond : \mathbb{F} \times V \to V, \quad (a,v) \to a \diamond v := \overline{a} v$$

Then $V_\diamond = (V, +, \diamond)$ is a vector space over $\mathbb{F}$.

(b) *Let*

$$T : W \to V$$

be semilinear. Then

$$T : W \to V_\diamond$$

is linear.

(c) *Let $U \subseteq V$. Then U is a subspace of V if and only if U is a subspace of $V_\diamond$*

(d) *Let $\mathfrak{v}$ be a list of vectors in V. Then $\mathfrak{v}$ is a basis for V if and only if $\mathfrak{v}$ is a basis for $V_\diamond$*

(e) *If V is finite dimensional, V and $V_\diamond$ are isomorphic.*

Proof: Let $a, b \in \mathbb{F}$ and $v, w \in V$.

(a) We only need to verify the axioms of a vector space which involve scalar multiplication:

$$(ab) \diamond v = (\overline{ab})v = (\overline{a}\overline{b})v = \overline{a}(\overline{b}v) = a \diamond (b \diamond v)$$

$$(a+b) \diamond v = (\overline{a+b})v = (\overline{a} + \overline{b})v = \overline{a}v + \overline{b}v = a \diamond v + b \diamond v$$

$$a \diamond (v+w) = \overline{a}(v+w) = \overline{a}v + \overline{a}w = a \diamond v + b \diamond v$$

Note that $\overline{1} = \overline{1 \cdot 1} = \overline{1} \cdot \overline{1}$ and so $\overline{1} = 1$.

$$1 \diamond v = \overline{1}v = 1v = v$$

(b) $T(aw) = \overline{a}T(w) = a \diamond T(v)$

(c) Since $^-$ is a bijection the following are equivalent:

$$\begin{aligned} au \in U \quad & \forall a \in \mathbb{F}, u \in U \\ \overline{a}u \in U \quad & \forall a \in \mathbb{F}, u \in U \\ a \diamond u \in U \quad & \forall a \in \mathbb{F}, u \in U \end{aligned}$$

Thus (c) holds.

(d) Let $\mathfrak{v} = (v_1, \ldots, v_n)$ and $v \in V$. Then

$$v = \sum_i a_i v_i \iff v = \sum_i \overline{a_i} \diamond v_i$$

Hence v is a unique linear combination of $\mathfrak{v}$ in V if and only if v is a unique linear combination of $\mathfrak{v}$ in $V_\diamond$.

(e) By (d) V and $V_\diamond$ have the same dimension and so are isomorphic. Just for fun let us determine an isomorphism explicitly:

Let $(v_1, v_2, \ldots v_n)$ be a basis. Then

$$\sum_i a_i v_i \to \sum_i \overline{a_i} v_i$$

is an isomorphism from V to $V_\diamond$. Note though that this isomorphism depends an the choice of the basis. □

We noted above that there does not exist a canonical isomorphism between V and V^*. The next proposition says that in the case of non-degenerate hermitian space there does exists a canonical semilinear isomorphism.

Lemma 10.7 [**semilinear isomorphism**] *Let V be a finite dimensional, non-degenerate hermitian space.*

(a) *Define*

$$\Phi : V \to V^*, v \to \langle \cdot, v \rangle.$$

Then $\Phi(v)(u) = \langle u, v \rangle$ and Φ is a semilinear isomorphism.

(b) *For each $\phi \in V^*$ there exists a unique $v \in V$ with*

$$\phi(u) = \langle u, v \rangle$$

for all $u \in V$.

Proof: (a) By 9.3 $\langle \cdot, v \rangle$ is indeed in V^*. That Φ is semilinear follows easily from the fact that $\langle \cdot, \cdot \rangle$ is semilinear in the second coordinate. Let $v \in \operatorname{null} \Phi$. Then $\langle \cdot, v \rangle$ is the zero functional on V, that is $\langle u, v \rangle = 0$ for all $u \in V$. Since V is non degenerate we conclude $v = 0$. So $\operatorname{null} \Phi = \{0_V\}$. Hence Φ is one to one. Since V is finite dimensional we conclude that Φ is a (semilinear) isomorphism.

(c) Note that we can rephrase (c) as: There exists a unique $v \in V$ with $\Phi(v) = \phi$. But this is true since Φ is a bijection. □

We denote the map Φ from the previous theorem by Φ_V.

Lemma 10.8 [**adjoint**] *Let V, W be a finite dimensional, non-degenerate hermitian spaces and $T \in \mathcal{L}(V, W)$. Then there exists a unique $T^{\mathrm{ad}} \in \mathcal{L}(W, V)$ with*

$$\langle T(v), w \rangle = \langle v, T^{\mathrm{ad}}(w) \rangle$$

for all $v \in V, w \in W$. Moreover,

$$T^{\mathrm{ad}} = \Phi_V^{-1} \circ T^* \circ \Phi_W$$

Proof: Note that the following statements are equivalent (if they hold for all $v \in V, w \in W$)

$$\begin{aligned}
\langle T(v), w \rangle &= \langle v, T^{\mathrm{ad}}(w) \rangle \\
\Phi_W(w)(T(v)) &= \Phi_V(T^{\mathrm{ad}}(w))(v) \\
(\Phi_W(w) \circ T)(v) &= (\Phi_V \circ T^{\mathrm{ad}})(w)(v) \\
\Phi_W(w) \circ T &= (\Phi_V \circ T^{\mathrm{ad}})(w) \\
T^*(\Phi_W(w)) &= (\Phi_V \circ T^{\mathrm{ad}})(w) \\
(T^* \circ \Phi_W)(w) &= (\Phi_V \circ T^{\mathrm{ad}})(w) \\
T^* \circ \Phi_W &= \Phi_V \circ T^{\mathrm{ad}} \\
\Phi_V^{-1} \circ \mathrm{T}^* \circ \Phi_W &= T^{\mathrm{ad}}
\end{aligned}$$

Also note that as a composition of (semi)-linear map, $\Phi_V^{-1} \circ \mathrm{T}^* \circ \Phi_W$ is linear. So $T^{\mathrm{ad}} = \Phi_V^{-1} \circ \mathrm{T}^* \circ \Phi_W$ is the unique linear map from $W \to V$ which fulfils the lemma. □

T^{ad} is called the adjoint of T.

Lemma 10.9 [**Tadad**] *Let V be a finite dimensional, non-degenerate hermitian space.*

(a) *If $T \in \mathcal{L}(V)$ then $T^{\mathrm{adad}} = T$*

(b) *If $\lambda \in \mathbb{F}$ then $\lambda^{\mathrm{ad}} = \overline{\lambda}$*

Proof: (a)

$$\langle T^{\mathrm{ad}}(v), w\rangle = \overline{\langle w, T^{\mathrm{ad}}(v)\rangle} = \overline{\langle T(w), v\rangle} = \langle v, T(w)\rangle$$

(b)

$$\langle \lambda v, w\rangle = \lambda\langle v, w\rangle = \langle v, \overline{\lambda}(w)\rangle$$

□

If $A = (a_{ij})$ is a matrix, then $\overline{A}$ denotes the matrix $(\overline{a_{ij}})$.

Lemma 10.10 [**matrix for the adjoint**] *Let V, W be a finite dimensional, non-degenerate hermitian space, $T \in \mathcal{L}(V)$ and $\mathfrak{v}$ and $\mathfrak{w}$ orthonormal bases for V and W, respectively. Let A be the matrix for T and B the matrix of T^{ad}. Then*

$$B = \overline{A}^{\mathrm{T}}$$

Proof:

$$\langle T(v_j), w_i\rangle = \langle \sum_k a_{kj} w_k, w_i\rangle = a_{ij}$$

and

$$\langle v_j, T^{\mathrm{ad}}(w_i)\rangle = \langle v_j, \sum_k b_{ki} v_k\rangle = \overline{b_{ji}}.$$

So by the definition of the adjoint $a_{ij} = \overline{b_{ji}}$ □

Lemma 10.11 [**composition of dual maps**]

(a) *Let $T : V \to W$ and $S : U \to V$ be linear maps. Then*

$$(T \circ S)^* = S^* \circ T^*$$

(b) *Let U, V, W be non-degenerate, finite dimensional hermitian spaces and $T : V \to W$ and $S : U \to V$ linear maps. Then*

$$(T \circ S)^{\mathrm{ad}} = S^{\mathrm{ad}} \circ T^{\mathrm{ad}}$$

Proof: (a) Let $\phi \in W^*$. Then

$$(S^* \circ T^*)(\phi) = S^*(T^*(\phi)) = S^*(\phi \circ T) = (\phi \circ T) \circ S = \phi \circ (T \circ S) = (T \circ S)^*(\phi).$$

(b)

$$(T \circ S)^{\mathrm{ad}} = \Phi_U^{-1} \circ (T \circ S)^* \circ \Phi_W = \Phi_U^{-1} \circ S^* \circ T^* \circ \Phi_W = (\Phi_U^{-1} \circ S^* \circ \Phi_V) \circ (\Phi_V^{-1} \circ T^* \circ \Phi_W) = S^{\mathrm{ad}} \circ T^{\mathrm{ad}}.$$

□

Lemma 10.12 [**Uperp**] *Let V be a non-degenerate, hermitian space and U a finite dimensional subspace of V.*

(a) $\dim U = \dim(V/U^\perp)$.

(b) $U^{\perp\perp} = U$.

Proof: (a) Consider the map

$$\Psi : V \to U^*, v \to \langle \cdot, v \rangle|_U$$

Then Ψ is semilinear and $\Psi(v) = 0$ if and only if $\langle u, v \rangle = 0$ for all $u \in U$. So $\operatorname{null} \Psi = U^\perp$. Suppose that Ψ is not onto. Then by 10.3 there exists $0 \neq \alpha \in U^{**}$ with $\alpha|_{\operatorname{range} \Psi} = 0$. By 10.2 there exists $0 \neq u \in U$ with $\alpha = \hat{u}$. Then for all $v \in V$.

$$0 = \alpha(\Psi(v)) = \hat{u}(\Psi(v)) = \Psi(v)(u) = \langle u, v \rangle$$

But this is impossible as V is non-degenerate and $u \neq 0$.

Thus Ψ is onto. Hence by the isomorphism theorem U and $V/U^\perp$ are (semilinearly) isomorphic. Thus (a) holds.

(b) Note that $U \leq U^{\perp\perp}$. Let $x \in U^\perp \perp$ and put $W = U + \mathbb{F}x$. Then W is finite dimensional and $\dim W \geq \dim U$. Since $W \leq U^{\perp\perp}$, $U^\perp \leq W^\perp$. Thus by (a) applied to W:

$$\dim W = \dim(V/W^\perp) \leq \dim(V/U^\perp) = \dim U \leq \dim W$$

Hence $\dim W = \dim U$ and so $U = W$. Thus $x \in U$ and so $U^{\perp\perp} \leq U$. □

Lemma 10.13 [**null and range for Tad**] *Let V, W be finite dimensional, non-degenerate hermitian spaces and $T : V \to W$ be linear. Then*

(a) $\operatorname{null} T = (\operatorname{range} T^{\mathrm{ad}})^\perp$

(b) $\operatorname{null} T^{\mathrm{ad}} = (\operatorname{range} T)^\perp$

(c) $\operatorname{range} T = (\operatorname{null} T^{\mathrm{ad}})^\perp$

(d) $\operatorname{range} T^{\mathrm{ad}} = (\operatorname{null} T)^{\perp}$

Proof: (a) Let $v \in V$. Using that V is non-degenerate we see that the following are equivalent for

$$\begin{aligned} v &\in \operatorname{null} T \\ T(v) &= 0 \\ \langle T(v), w\rangle &= 0 \qquad \forall w \in W \\ \langle v, T^{\mathrm{ad}}(w)\rangle &= 0 \qquad \forall w \in W \\ \langle v, u\rangle &= 0 \quad \forall u \in \operatorname{range} T^{\mathrm{ad}} \\ v &\in (\operatorname{range} T^{\mathrm{ad}})^{\perp} \end{aligned}$$

So (a) holds.

(d) By (a) and 10.12

$$\operatorname{range} T^{\mathrm{ad}} = (\operatorname{range} T^{\mathrm{ad}})^{\perp\perp} = (\operatorname{null} T)^{\perp}$$

(b) apply (a) to T^{ad} in place of T and use $T^{\mathrm{adad}} = T$.
(c) apply (d) to T^{ad} in place of T and use $T^{\mathrm{adad}} = T$.

Lemma 10.14 [**invariant subspaces for Tad**] *Let V be a finite dimensional, non-degenerate hermitian space and $T \in \mathcal{L}(V)$. Let U be subspace of V. Then U is T-invariant if and only if $U^{\perp}$ is T^{ad}-invariant.*

Proof: The following are equivalent:

$$\begin{aligned} U &\text{ is } T\text{-invariant} \\ T(u) &\in U \qquad \forall u \in U \\ T(u) &\in U^{\perp\perp} \qquad \forall u \in U \\ \langle T(u), w\rangle &= 0 \quad \forall u \in U, w \in U^{\perp} \\ \langle u, T^{\mathrm{ad}}(w)\rangle &= 0 \quad \forall u \in U, w \in U^{\perp} \\ T^{\mathrm{ad}}(w) &\in U^{\perp} \qquad \forall w \in U^{\perp} \\ U^{\perp} &\text{ is } T^{\mathrm{ad}}\text{-invariant} \end{aligned}$$

□

Definition 10.15 *Let T be an operator on the finite-dimensional, non-degenerate hermitian space V.*

(a) *T is called self-adjoint if $T = T^{\mathrm{ad}}$, that is if*

$$\langle T(v), w\rangle = \langle v, T(w)\rangle, \quad \forall v, w \in V$$

(b) *T is called normal of T commutes with T^{ad}, that is if $T \circ T^{\mathrm{ad}} = T^{\mathrm{ad}} \circ T$*

(c) $\langle u, v\rangle_T = \langle T(u), v\rangle$

(d) $\|v\|_T^2 = \langle v, v\rangle_T := \langle T(v), v\rangle$

Note that since T commutes with itself, every self-adjoint operator is normal.

Lemma 10.16 [**alternative definition of self adjoint**] *Let R, S and T be operators on the finite-dimensional, non-degenerate hermitian space V.*

(a) *T is self-adjoint if and only if $\langle \cdot, \cdot\rangle_R$ is a hermitian form on V.*

(b) *R and S on V are equal, if and only if $\langle \cdot, \cdot\rangle_R = \langle \cdot, \cdot\rangle_S$.*

Proof: (a) Since compositions of linear maps are linear $\langle \cdot, \cdot\rangle_T$ is linear in the first coordinate. Let $u, v \in V$. Then

$$\langle v, u\rangle_T = \langle T(v), u\rangle = \langle v, T^{\mathrm{ad}}(u)\rangle = \overline{\langle T^{\mathrm{ad}}(u), v\rangle}$$

and

$$\overline{\langle u, v\rangle_T} = \overline{\langle T(u), v\rangle}$$

Hence (using 9.11) the following are equivalent

$$\begin{aligned} &\langle \cdot, \cdot\rangle_T \text{ is hermitian} \\ \langle T^{\mathrm{ad}}(u), v\rangle &= \langle T^{\mathrm{ad}}(u), v\rangle \quad \forall u, v \in V \\ T^{\mathrm{ad}}(u) &= T(u) \quad \forall u \in V \\ T^{\mathrm{ad}} &= T. \end{aligned}$$

(b) Follows from 9.11. □

Lemma 10.17 [**equality of self-adjoint operators**] *Let R and S be self-adjoint operators on the finite-dimensional, non-degenerate hermitian+ space V. Then $R = S$ if and only if $\|\cdot\|_R^2 = \|\cdot\|_S^2$.*

Proof: By 10.16, $\langle \cdot, \cdot\rangle_R$ and $\langle \cdot, \cdot\rangle$ are hermitian forms and R and S are equal if and only if their forms are. By 9.16 this is the case if and only if $\|\cdot\|_R^2 = \|\cdot\|_S^2$. □

Lemma 10.18 [**another alternative definition of normal**] *LetT be an operator on the finite-dimensional, non-degenerate hermitian space V. Then T is normal and if and only if*

$$\langle T(v), T(w)\rangle = \langle T^{\mathrm{ad}}(v), T^{\mathrm{ad}}(w)\rangle$$

for all $v, w \in V$.

Proof: Since $(TT^{\mathrm{ad}})^{\mathrm{ad}} = T^{\mathrm{adad}}T^{\mathrm{ad}} = TT^{\mathrm{ad}}$, TT^{ad} and $T^{\mathrm{ad}}T$ are self adjoint. Now

$$svw_{TT^{\mathrm{ad}}} = \langle T(T^{\mathrm{ad}}(v)), w\rangle = \langle T^{\mathrm{ad}}(v), T^{\mathrm{ad}}(w)\rangle$$

and

$$svw_{T^{\mathrm{ad}}T} = \langle T^{\mathrm{ad}}(T(v)), w\rangle = \langle T(v), T(w)\rangle$$

So the lemma follows from 10.16(b) □

Lemma 10.19 [**alternative definition of normal**] *Let T be an operator on the finite-dimensional, non-degenerate hermitian+ space V. Then T is normal and if and only if*

$$\|T(v)\|^2 = \|T^{\mathrm{ad}}(v)\|^2$$

for all $v \in V$.

Proof: By the proof of 10.18 we have $\|v\|^2_{TT^{\mathrm{ad}}} = \|T^{\mathrm{ad}}(v)\|^2$ and $\|v\|^2_{T^{\mathrm{ad}}T} = \|T(v)\|^2$. The lemma now follows from 10.17 □

Definition 10.20 *An hermitian space is called definite, if $\|v\|^2 \neq 0$ for all non-zero vectors $v \in V$.*

Note that a definite hermitian space is non-degenerate.

Lemma 10.21 [**eigenvectors for Tad**] *Let T be a normal operator on the finite-dimensional, definite hermitian space V. Let $v \in V$ be an eigenvector of T with eigenvalue λ. Then v is an eigenvector with eigenvalue $\overline{\lambda}$*

Proof: Note that $(T-\lambda)^{\mathrm{ad}} = T^{\mathrm{ad}} - \overline{\lambda}$. In particular, $T-\lambda$ is normal. Since $\|T-\lambda(v)\|^2 = \|0\|^2 == 0$, 10.18implies $\|(T^{\mathrm{ad}} - \overline{\lambda})(v)\|^2 = 0$. Since V is definite we conclude that $(T^{\mathrm{ad}} - \overline{\lambda})(v) = 0$. Hence v is an eigenvector with eigenvalue $\overline{\lambda}$ for T^{ad}. □

Theorem 10.22 [**The Spectral Theorem**] *Let T be a operator on the finite-dimensional, definite hermitian space V. Suppose that $\mathbb{F}$ is algebraically closed. Then T is normal if and only if V has an orthogonal basis consisting of eigenvectors of T.*

Proof: $\Longrightarrow$:

Suppose that $(v_1, \ldots, v_n)$ is an orthogonal basis for V consisting of eigenvectors of T Then

$v_i^{\perp} = \mathrm{Span}(v_j, j \neq i)$ and so $v_i^{\perp}$ is T-invariant. Thus by 10.14 $\mathbb{F}v_i = v_i^{\perp\perp}$ is T^{ad}-invariant. It follows that v_i is an eigenvector for T^{ad}. Thus the matrices for T and for T^{ad} are both diagonal. So T and T^{ad} commute and T is normal.

$\Longleftarrow$:

By induction on $\dim V$. If $\dim V = 0$ there is nothing to prove. So suppose $\dim V \neq 0$. Since $\mathbb{F}$ is algebraically closed, there exists a non -zero eigenvector v for T. By 10.21 v is also an eigenvector for T^{ad}. Hence $\mathbb{F}v$ is T^{ad} invariant. Put $W = v^{\perp} = (\mathbb{F}v)^{\perp}$. Then by 10.14 W is T-invariant. Since $\|v\|^2 \neq 0$, 9.5 implies

$$V = \mathbb{F}v \bigoplus W$$

Note that $T^{\text{ad}}|_W = (T|_W)^{\text{ad}}$. Thus $T|_W$ is normal. Hence by induction there exists a orthogonal basis $\mathfrak{w}$ of W consisting of eigenvectors of T. Then $(v, \mathfrak{w})$ is an orthogonal basis for V consisting of eigenvectors of T. □

References

[Axler] Sheldon Axler , *Linear Algebra Done Right*