Abstract Linear Algebra
Math 350

April 29, 2015

Contents

1 An introduction to vector spaces **2**
1.1 Basic definitions & preliminaries 4
1.2 Basic algebraic properties of vector spaces 6
1.3 Subspaces 7

2 Dimension **10**
2.1 Linear combination 10
2.2 Bases 13
2.3 Dimension 14
2.4 Zorn's lemma & the basis extension theorem 16

3 Linear transformations **18**
3.1 Definition & examples 18
3.2 Rank-nullity theorem 21
3.3 Vector space isomorphisims 26
3.4 The matrix of a linear transformation 28

4 Complex operators **34**
4.1 Operators & polynomials 34
4.2 Eigenvectors & eigenvalues 36
4.3 Direct sums 45
4.4 Generalized eigenvectors 49
4.5 The characteristic polynomial 54
4.6 Jordan basis theorem 56

Chapter 1

An introduction to vector spaces

Abstract linear algebra is one of the pillars of modern mathematics. Its theory is used in every branch of mathematics and its applications can be found all around our everyday life. Without linear algebra, modern conveniences such as the Google search algorithm, iPhones, and microprocessors would not exist. But what is abstract linear algebra? It is the study of vectors and functions on vectors from an abstract perspective. To explain what we mean by an abstract perspective, let us jump in and review our familiar notion of vectors. Recall that a *vector* of length n is a $n \times 1$ array

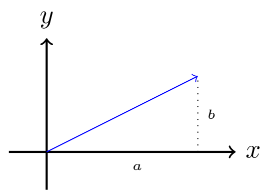
$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},$$

where a_i are real numbers, i.e., $a_i \in \mathbb{R}$. It is also customary to define

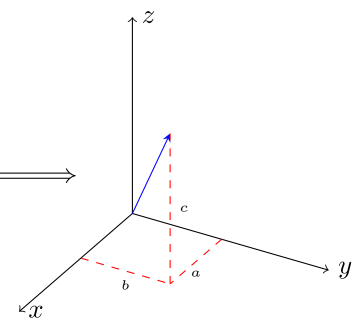
$$\mathbb{R}^n = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \mid a_i \in \mathbb{R} \right\},$$

which we can think of as the set where all the vectors of length n live. Some of the usefulness of vectors stems from our ability to draw them (at least those in $\mathbb{R}^2$ or $\mathbb{R}^3$). Recall that this is done as follows:

$$\begin{bmatrix} a \\ b \end{bmatrix} \Longleftrightarrow$$

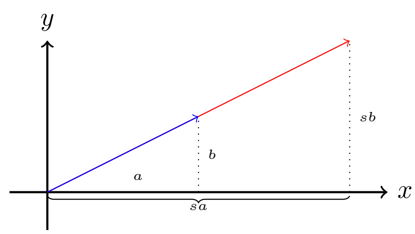


$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \Longleftrightarrow$$



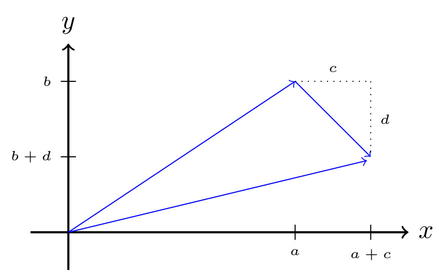
Basic algebraic operations on vectors correspond nicely with our picture of vectors. In particular, if we scale a vector v by a number s then in the picture we either stretch or shrink our arrow.

$$s \cdot \begin{bmatrix} a \\ b \end{bmatrix} \Longleftrightarrow$$



The other familiar thing we can do with vectors is add them. This corresponds to placing the vectors "head-to-tail" as shown in the following picture.

$$\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \qquad \Longleftrightarrow$$



In summary, our familiar notion of vectors can be captured by the following description. Vectors of length n live in the set $\mathbb{R}^n$ that is equipped with two operations. The first operation takes any pair of vectors $u, v \in \mathbb{R}^n$ and gives us a new vector $u + v \in \mathbb{R}^n$. The second operation takes any pair $a \in \mathbb{R}$ and $v \in \mathbb{R}^n$ and gives us a new vector $a \cdot v \in \mathbb{R}^n$.

With this summary in mind we now give a definition which generalizes this familiar notion of a vector. It will be very helpful to read the following in parallel with the above summary.

1.1 Basic definitions & preliminaries

Throughout we let $\mathbb{F}$ represent either the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

Definition. A **vector space over** $\mathbb{F}$ is a set V along with two operation. The first operation is called **addition**, denoted $+$, which assigns to each pair $u, v \in V$ an element $u + v \in V$. The second operation is called **scalar multiplication** which assigns to each pair $a \in \mathbb{F}$ and $v \in V$ an element $av \in V$. Moreover, we insist that the following properties hold, where $u, v, w \in V$ and $a, b \in \mathbb{F}$:

- **Associativity**
$$u + (v + w) = (u + v) + w \qquad \text{and} \qquad a(bv) = (ab)v.$$
- **Commutativity of +**
$$u + v = v + u.$$
- **Distributivity**
$$a(u + v) = au + av \qquad \text{and} \qquad (a + b)v = av + bv$$
- **Multiplicative Identity**
  The number $1 \in \mathbb{F}$ is such that
$$1v = v \qquad \text{for all } v \in V.$$
- **Additive Identity & Inverses**
  There exists an element $0 \in V$, called an **additive identity** or a **zero**, with the property that
$$0 + v = v \qquad \text{for all } v \in V.$$
  Moreover, for every $v \in V$ there exists some $u \in V$, called an **inverse of** v, such that $u + v = 0$.

It is common to refer to the elements of V as **vectors** and the elements of $\mathbb{F}$ as **scalars**. Additionally, if V is a vector space over $\mathbb{R}$ we call it a **real vector space** or an **$\mathbb{R}$-vector space**. Likewise, a vector space over $\mathbb{C}$ is called **a complex vector space** or a **$\mathbb{C}$-vector space**.

Although this definition is intimidating at first, you are more familiar with these ideas than you might think. In fact, you have been using vector spaces in your previous math courses without even knowing it! The following examples aim to convince you of this.

Examples.

1. $\mathbb{R}^n$ is a vector space over $\mathbb{R}$ under the usual vector addition and scalar multiplication as discussed in the introduction.

2. $\mathbb{C}^n$, the set of column vectors of length n whose entries are complex numbers, is a a vector space over $\mathbb{C}$.

3. $\mathbb{C}^n$ is also a vector space over $\mathbb{R}$ where addition is standard vector addition and scalar multiplication is again the standard operation but in this case we limit our scalars to real numbers only. This is NOT the same vector space as in the previous example; in fact, it is as different as a line is to a plane!

4. Let $\mathcal{P}(\mathbb{F})$ be the set of all polynomials with coefficients in $\mathbb{F}$. That is

$$\mathcal{P}(\mathbb{F}) = \{a_0 + a_1 x + \cdots + a_n x^n \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{F}\}.$$

Then $\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$. In this case our "vectors" are polynomials where addition is the standard addition on polynomials. For example, if $v = 1 + x + 3x^2$ and $u = x + 7x^2 + x^5$, then

$$u + v = (1 + x + 3x^2) + (x + 7x^2 + x^5) = 1 + 2x + 10x^2 + x^5.$$

Scalar multiplication is defined just as you might think. If $v = a_0 + a_1 x + \cdots + a_n x^n$, then

$$s \cdot v = sa_0 + sa_1 x + \cdots + sa_n x^n.$$

5. Let $\mathcal{C}(\mathbb{R})$ be the set of continuous functions

$$f : \mathbb{R} \to \mathbb{R}.$$

Then $\mathcal{C}(\mathbb{R})$ is a vector space over $\mathbb{R}$ where addition and scalar multiplication is given as follows. For any functions $f, g \in \mathcal{C}(\mathbb{R})$ we define

$$(f + g)(x) = f(x) + g(x).$$

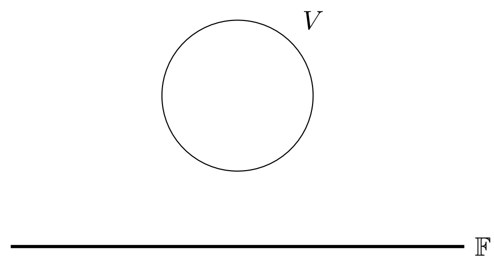
Likewise, for scalar multiplication we define

$$(s \cdot f)(x) = sf(x).$$

The reader should check that these definitions satisfy the axioms for a vector space.

6. Let $\mathcal{F}$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$. Then the set $\mathcal{F}$ is a vector space over $\mathbb{R}$ where addition and scalar multiplication are as given in Example 5.

You might be curious why we use the term "over" when saying that a vector space V is *over* $\mathbb{F}$. The reason for this is due to a useful way to visualize abstract vector spaces. In particular, we can draw the following picture



where our set V is sitting *over* our scalars $\mathbb{F}$.

1.2 Basic algebraic properties of vector spaces

There are certain algebraic properties that we take for granted in $\mathbb{R}^n$. For example, the zero vector

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n$$

is the unique additive identity in $\mathbb{R}^n$. Likewise, in $\mathbb{R}^n$ we do not even think about the fact that $-v$ is the (unique) additive inverse of v. These algebraic properties are so fundamental that we certainly would like our general vector spaces to have these same properties as well. As the next several lemmas show, this is happily the case.

Assume throughout this section that V is a vector space over $\mathbb{F}$.

Lemma 1.1. *V has a unique additive identity.*

Proof. Assume 0 and $0'$ are both additive identities in V. To show V has a unique additive identity we show that $0 = 0'$. Playing these two identities off each other we see that

$$0' = 0 + 0' = 0,$$

where the first equality follows as 0 is an identity and the second follows since $0'$ is also an identity. $\square$

An immediate corollary of this lemma is that now we can talk about *the* additive identity or *the* zero of a vector space. To distinguish between zero, the number in $\mathbb{F}$ and the zero the additive identity in V we will often denote the latter as 0_V.

Lemma 1.2. *Every element $v \in V$ has a unique additive inverse denoted $-v$.*

Proof. Fix $v \in V$. As in the proof of the previous lemma, it will suffice to show that if u and u' are both additive inverses of v, then $u = u'$. Now consider

$$u' = 0_V + u' = (u + v) + u' = u + (v + u') = u + 0_V = u,$$

where associativity gives us the third equality. $\square$

Lemma 1.3 (Cancellation Lemma)**.** *If u, v, w are vectors in V such that*

$$u + w = v + w, \tag{*}$$

then $u = v$

Proof. To show this, add $-w$ to both sides of (*) to obtain $(u + w) + -w = (v + w) + -w$. By associativity,

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

$$u + 0_V = v + 0_V$$

$$u = v.$$

$\square$

Lemma 1.4. *For any $a \in \mathbb{F}$ and $v \in V$, we have*

$$0 \cdot v = 0_V$$

and

$$a \cdot 0_V = 0_V.$$

Proof. The proof of this is similar to the Cancellation Lemma. We leave its proof to the reader. $\square$

The next lemma asserts that $-1 \cdot v = -v$. A natural reaction to this statement is: Well isn't this obvious, what is there to prove? Be careful! Remember v is just an element in an abstract set V endowed with some specific axioms. From this vantage point, it is not clear that the vector defined by the abstract rule $-1 \cdot v$ should necessarily be the additive inverse of v.

Lemma 1.5. $-1 \cdot v = -v$.

Proof. Observe that $-1 \cdot v$ is an additive inverse of v since

$$v + -1 \cdot v = 1 \cdot v + -1 \cdot v = (1 - 1) \cdot v = 0v = 0_V,$$

where the last two equalities follow from the distributive law and the previous lemma respectively. As v has only one additive inverse by Lemma 1.2, then $-1 \cdot v = -v$. $\square$

1.3 Subspaces

Definition. Let V be a vector space over $\mathbb{F}$. We say that a subset U of V is a **subspace** (of V), provided that U is a vector space over $\mathbb{F}$ using the same operations of addition and scalar multiplication as given on V.

Showing that a given subset U is a subspace of V might at first appear to involve a lot of checking. Wouldn't one need to check Associativity, Commutativity, etc? Fortunately, the answer is no. Think about it, since these properties hold true for all the vectors in V they certainly also hold true for some of the vectors in V, i.e., those in U. (The fancy way to say this is that U **inherits** all these properties from V.) Instead we need only check the following:

1. $0_V \in U$
2. $u + v \in U,$ for all $u, v \in U$ (**Closure under addition**)
3. $av \in U,$ for all $a \in \mathbb{F},$ and $v \in U$ (**Closure under scalar multiplication**)

Examples.

1. For any vector space V over $\mathbb{F}$, the sets V and $\{0_V\}$ are both subspaces of V. The former is called a **nonproper** subspace while the latter is called the **trivial** or **zero** subspace. Therefore a proper nontrivial subspace of V is one that is neither V nor $\{0_V\}$.

2. Consider the real vector space $\mathbb{R}^3$. Fix real numbers a, b, c. Then we claim that the subset

$$U = \left\{(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\right\}$$

is a subspace of $\mathbb{R}^3$. To see this we just need to check the three closure properties. First, note that $0_{\mathbb{R}^3} = (0, 0, 0) \in U$, since $0 = a0 + b0 + c0$. To see that U is closed under addition let $u = (x_1, y_1, z_1), v = (x_2, y_2, z_2) \in U$. Since

$$a(x_1+x_2)+b(y_1+y_2)+c(z_1+z_2) = (ax_1+by_1+cz_1)+(ax_2+by_2+cz_2) = 0+0 = 0$$

we see that $u + v = (x_1 + x_2, y_1 + y_2, z_1 + z_2) \in U$. Lastly, a similiar check shows that U is closed under scalar multiplication. Let $s \in \mathbb{R}$, then

$$0 = s0 = s(ax_1 + by_1 + cz_1) = asx_1 + bsy_1 + csz_1.$$

This means that $su = (sx_1, sy_1, sz_1) \in U$.

3. Recall that $\mathcal{P}(\mathbb{R})$ is the vector space over $\mathbb{R}$ consisting of all polynomials whose coefficients are in $\mathbb{R}$. In fact, this vector space is also a subspace of $\mathcal{C}(\mathbb{R})$. To see this note that $\mathcal{P}(\mathbb{R}) \subset \mathcal{C}(\mathbb{R})$. Since the zero function and the zero polynomial are the same function, then $0_{\mathcal{C}(\mathbb{R})} \in \mathcal{P}(\mathbb{R})$. Since we already showed that $\mathcal{P}(\mathbb{R})$ is a vector space then it is certainly closed under addition and scalar multiplication, so $\mathcal{P}(\mathbb{R})$ is a subspace of $\mathcal{C}(\mathbb{R})$.

4. This next examples demonstrates that we can have subspaces within subspaces. Consider the subset $\mathcal{P}_{\leq n}(\mathbb{R})$ of $\mathcal{P}(\mathbb{R})$ consisting of all those polynomials with degree $\leq n$. Then, $\mathcal{P}_{\leq n}(\mathbb{R})$ is a subspace of $\mathcal{P}(\mathbb{R})$. As the degree of the zero polynomial is (defined to be) $-\infty$, then $\mathcal{P}_{\leq n}(\mathbb{R})$. Additionally, if $u, v \in \mathcal{P}_{\leq n}(\mathbb{R})$, then clearly the degree of $u + v$ is $\leq n$, so $u + v \in \mathcal{P}_{\leq n}(\mathbb{R})$. Likewise $\mathcal{P}_{\leq n}(\mathbb{R})$ is certainly closed under scalar multiplication. Combining this example with the previous one shows that we actually have the following sequence of subspaces

$$\mathcal{P}_{\leq 0}(\mathbb{R}) \subset \mathcal{P}_{\leq 1}(\mathbb{R}) \subset \mathcal{P}_{\leq 2}(\mathbb{R}) \subset \cdots \subset \mathcal{P}(\mathbb{R}) \subset \mathcal{C}(\mathbb{R}).$$

5. The subset $\mathcal{D}$ of all differentiable functions in $\mathcal{C}(\mathbb{R})$, is a subspace of the $\mathbb{R}$-vector space $\mathcal{C}(\mathbb{R})$. Since the zero function $f(x) = 0$ is differentiable, and the sum and scalar multiple of differentiable functions is differentiable, it follows that $\mathcal{D}$ is a subspace.

6. Let U be the set of solutions to the differential equation $f''(x) = -f'(x)$, i.e.,

$$U = \{f(x) \mid f''(x) = -f'(x)\}.$$

Then U is a subspace $\mathcal{D}$, the space of differentiable functions. To see this, first note that the zero function is a solution to our differential equation. Therefore U contains our zero vector. To check the closure properties let $f, g \in U$. Therefore $f''(x) = -f'(x)$ and that $g''(x) = -g(x)$ and moreover,

$$(f + g)''(x) = f''(x) + g''(x) = -f'(x) + -g'(x) = -(f + g)'(x).$$

In other words, $f + g \in U$. To check closure under scalar multiplication let $s \in \mathbb{R}$. Now

$$(s \cdot f)''(x) = sf''(x) = -sf'(x) = -(sf)'(x),$$

and so $s \cdot f \in U$.

Chapter 2

Dimension

2.1 Linear combination

Definition. A **linear combination** of the vectors $v_1, \dots, v_m$ is any vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \dots, a_m \in \mathbb{F}$. For a nonempty subset S of V, we define

$$\operatorname{span}(S) = \{a_1 v_1 + \cdots + a_m v_m \mid v_1, \dots, v_m \in S, a_1, \dots, a_m \in \mathbb{F}\},$$

and call this set the **span of** S. If $S = \emptyset$, we define $\operatorname{span}(\emptyset) = \{0_V\}$. Lastly, if $\operatorname{span}(S) = V$, we say that S **spans** V or that S is a **spanning set** for V.

For example, consider the vector space $\mathbb{R}^n$ and let $S = \{e_1, \dots, e_n\}$, where

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix},$$

that is, the vector whose entries are all 0 except the ith, which is 1. Then $\mathbb{R}^n = \operatorname{span}(S)$, since we can express any vector $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}$ as

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a_1 e_1 + \cdots a_n e_n.$$

The vectors $e_1, \dots, e_n$ play a fundamental role in the theory of linear algebra. As such they are named the **standard basis vectors** for $\mathbb{R}^n$.

Now consider the vector space of continuous functions $\mathcal{C}(\mathbb{R})$. For brevity let us write the function $f(x) = x^n$ as x^n and let $S = \{1, x, x^2, \ldots\}$. Certainly

$$\operatorname{span}(S) = \left\{a_0 1 + a_1 x + a_2 x^2 + \cdots + a_n x^n \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{R}\right\} = \mathcal{P}(\mathbb{R}).$$

This example raises a subtle point we wish to make explicit. Although our set S has *infinite* cardinality, each element in $\operatorname{span}(S)$ is a linear combination of a *finite* number of vectors in S. We do not allow something like $1 + x + x^2 + \cdots$ to be an element in $\operatorname{span}(S)$. A good reason for this restriction is that, in this case, such an expression is not defined for $|x| \geq 1$, so it could not possibly be an element of $\mathcal{C}(\mathbb{R})$.

Example 3 in Section 1.3 shows that $\mathcal{P}(\mathbb{R})$ is a subspace of $\mathcal{C}(\mathbb{R})$. The next lemma provides an alternate way to see this fact where we take $S = \{1, x, x^2, \ldots\}$ and $V = \mathcal{C}(\mathbb{R})$. Its proof is left to the reader.

Lemma 2.1. *For any $S \subseteq V$, we have that* $\operatorname{span}(S)$ *is a subspace of V.*

To motivate the next definition, consider the set of vectors from $\mathbb{R}^2$:

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}.$$

Since

$$\begin{bmatrix} a \\ b \end{bmatrix} = b \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (a - b) \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

we see that $\operatorname{span}(S) = \mathbb{R}^2$. That said, the vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is not needed in order to span R^2. It is in this sense that $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is an "unnecessary" or "redundant" vector in S. The reason this occurs is that $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is a linear combination of the other two vectors in S. In particular, we have

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

or

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Consequently, the next definition makes precise this idea of "redundant" vectors.

Definition. We say a set S of vectors is **linearly dependent** if there exists distinct vectors $v_1, \ldots, v_m \in S$ and scalars $a_1, \ldots, a_m \in \mathbb{F}$, *not* all zero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0_V.$$

If S is *not* linearly dependent we say it is **linearly independent**.

As the empty set $\emptyset$ is a subset of every vector space, it is natural to ask if $\emptyset$ is linearly dependent or linearly independent. The only way for $\emptyset$ to be dependent is if there exists some vectors $v_1, \ldots, v_m$ in $\emptyset$ whose linear combination is 0_V. But we are stopped dead in our tracks since there are NO vectors in $\emptyset$. Therefore $\emptyset$ cannot be linearly dependent, hence, $\emptyset$ is linearly independent.

Lemma 2.2 (Linear Dependence Lemma)**.** *If S is a linearly dependent set, then there exists some element $v \in S$ so that*

$$\operatorname{span}(S - v) = \operatorname{span}(S).$$

Moreover, if T is a linear independent subset of S, we may choose $v \in S - T$.

Proof. As S is linearly dependent we know there exist distinct vectors

$$\underbrace{v_1, \ldots, v_i}_{\in T}, \underbrace{v_{i+1}, \ldots, v_m}_{\in S-T}$$

and scalars $a_1, \ldots, a_m$, not all zero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0_V.$$

As T is linearly independent and $v_1, \ldots, v_i$ are distinct, we cannot have $a_{i+1} = \cdots = a_m = 0$. (Why?) Without loss of generality we may assume that $a_m \neq 0$. At this point choose $v = v_m$ and observe that $v \notin T$. Rearranging the above equation we obtain

$$v = v_m = -\left(\frac{a_1}{a_m} v_1 + \ldots + \frac{a_i}{a_m} v_i + \frac{a_{i+1}}{a_m} v_{i+1} + \ldots + \frac{a_{m-1}}{a_m} v_{m-1}\right),$$

which implies that $v \in \operatorname{span}(S - v)$. Moreover, since $S - v \subseteq \operatorname{span}(S - v)$, we see that $S \subset \operatorname{span}(S - v)$. Lemma 2.1 now implies that

$$\operatorname{span}(S) \subseteq \operatorname{span}(S - v) \subseteq \operatorname{span}(S),$$

which yields our desired result. □

Lemma 2.3 (Linear Independence Lemma)**.** *Let S be linearly independent. If $v \in V$ but not in $\operatorname{span}(S)$, then $S \cup \{v\}$ is also linearly independent.*

Proof. If $V - \operatorname{span}(S) = \emptyset$, then there is nothing to prove. Otherwise, let $v \in V$ such that $v \notin \operatorname{span}(S)$ and assume for a contradiction that $S \cup \{v\}$ is linearly dependent. This means that there exists distinct vectors $v_1, \ldots, v_m \in S \cup \{v\}$ and scalars $a_1, \ldots, a_m$, not all zero, such that $a_1 v_1 + a_1 v_1 + \cdots + a_m v_m = 0_V$. First, observe that $v = v_i$ for some i and that $a_i \neq 0$. (Why?) Without loss of generality we may choose $i = m$. Just like the calculation we performed in the proof of the Linear Dependence Lemma, we also have

$$v = v_m = -\left(\frac{a_1}{a_m} v_1 + \cdots + \frac{a_{m-1}}{a_m} v_{m-1}\right) \in \operatorname{span}(S),$$

which contradicts the fact that $v \notin \operatorname{span}(S)$. We conclude that $S \cup \{v\}$ is linearly independent. □

2.2 Bases

Definition. A (possibly empty) subset $\mathcal{B}$ of V is called a **basis** provided it is linearly independent and spans V.

Examples.

1. The set of standard basis vectors $e_1, \ldots, e_n$ are a basis for $\mathbb{R}^n$ and $\mathbb{C}^n$. (This explains their name!)
2. The set $\{1, x, x^2, \ldots, x^n\}$ form a basis for $\mathcal{P}_{\leq n}(\mathbb{F})$.
3. The infinite set $\{1, x, x^2, \ldots\}$ form a basis for $\mathcal{P}(\mathbb{F})$.
4. The emptyset $\emptyset$ forms a basis for the trivial vectors space $\{0_V\}$. This might seem odd at first but consider the definitions involved. First $\emptyset$ was defined to be linearly independent. Additionally, we defined $\text{span}(\emptyset) = \{0_V\}$. Therefore $\emptyset$ must be a basis for $\{0_V\}$.

The proof of the next lemma is left to the reader.

Lemma 2.4. *The subset $\mathcal{B}$ is a basis for V if and only if every vector $u \in V$ is a unique linear combination of the vectors in $\mathcal{B}$.*

Theorem 2.5 (Basis Reduction Theorem)**.** *Assume S is a finite set of vectors such that* $\text{span}(S) = V$. *Then there exists some subset $\mathcal{B}$ of S that is a basis for V.*

Proof. If S happens to be linearly independent we are done. On the other hand, if S is linearly dependent, then, by the Linear Dependence Lemma, there exists some $v \in S$ such that

$$\text{span}(S - v) = \text{span}(S)$$

If $S - v$ is not independent, we may continue to remove vectors until we obtain a subset $\mathcal{B}$ of S which is independent. (Note that since$S$ is finite we cannot continue removing vectors forever, and since $\emptyset$ is linearly independent this removal process must result in an independent set.) Additionally,

$$\text{span}(\mathcal{B}) = V,$$

since the Linear Dependence Lemma guarantees that the subset of S obtained after each removal spans V. We conclude that $\mathcal{B}$ is a basis for V. $\square$

Theorem 2.6 (Basis Extension Theorem)**.** *Let L be a linearly independent subset of V. Then there exists a basis $\mathcal{B}$ of V such that $L \subset \mathcal{B}$.*

We postpone the proof of this lemma to Section 2.4.

Corollary 2.7. *Every vector space has a basis.*

Proof. As the empty set $\emptyset$ is a linearly independent subset of any vector space V, the Basis Extension Theorem implies that V has a basis. $\square$

2.3 Dimension

Lemma 2.8. *If L is any finite independent set and S spans V, then $|L| \leq |S|$.*

Proof. Of all the sets that span V and have cardinality $|S|$ choose S' so that it maximizes $|L \cap S'|$. If we can prove that $L \subset S'$ we are done, since

$$|L| \leq |S'| = |S|.$$

For a contradiction, assume L is not a subset of S'. Fix some vector $u \in L - S'$. As S' spans V and does not contain u, then $D = S' \cup \{u\}$ is linearly dependent. Certainly, $\text{span}(D) = V$. Now define the linearly independent subset

$$T = L \cap D,$$

and observe that $u \in T$. By the Linear Dependence Lemma there exists some $v \in D - T$ so that

$$\text{span}(D - v) = \text{span}(D) = V.$$

Observe $u \neq v$. This immediately yields our contradiction since $|D - v| = |S'| = |S|$ and $D - v$ has one more vector from L (the vector u) than S' does. As this contradicts our choice of S', we conclude that $L \subset S'$ as needed. $\square$

Theorem 2.9. *Let V be a vector space with at least one finite basis $\mathcal{B}$. Then every basis of V has cardinality $|\,\mathcal{B}\,|$.*

Proof. Fix any other basis $\mathcal{B}_0$ of V. As $\mathcal{B}_0$ spans V, Lemma 2.8, with $S = \mathcal{B}_0$ and $L = \mathcal{B}$, implies that $|\,\mathcal{B}\,| \leq |\,\mathcal{B}_0\,|$. Our proof will now be complete if we can show that $|\,\mathcal{B}\,| \not< |\,\mathcal{B}_0\,|$. For a contradiction, assume that $n = |\,\mathcal{B}\,| < |\,\mathcal{B}_0\,|$ and let L be any $n+1$ element subset of $\mathcal{B}_0$ and $S = \mathcal{B}$. (As $\mathcal{B}_0$ is a basis, L is linearly independent.) Lemma 2.8 then implies that $n + 1 = |L| \leq |S| = n$, which is absurd. $\square$

Definition. A vector space V is called **finite-dimensional** if it has a finite basis $\mathcal{B}$. As all bases in this case have the same cardinality, we call this common number the **dimension** of V and denote it by $\dim(V)$.

A beautiful consequence of Theorem 2.9, is that in order to find the dimension of a given vector space we need only find the cardinality of some basis for that space. Which basis we choose doesn't matter!

Examples.

1. The dimension of $\mathbb{R}^n$ is n since $\{e_1, \ldots, e_n\}$ is a basis for this vector space.
2. Recall that $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathcal{P}_{\leq n}(\mathbb{R})$. Therefore $\dim \mathcal{P}_{\leq n}(\mathbb{R}) = n + 1$.

3. Consider the vector space $\mathbb{C}$ over $\mathbb{C}$. A basis for this space is $\{1\}$ since every element in $\mathbb{C}$ can be written uniquely as $s \cdot 1$ where s is a scalar in $\mathbb{C}$. Therefore, we see that this vector space has dimension 1. We can write this as $\dim_{\mathbb{C}}(\mathbb{C}) = 1$, where the subscript denotes that we are considering $\mathbb{C}$ as a vector space over $\mathbb{C}$.

   On the other hand, recall that $\mathbb{C}$ is also a vector space over $\mathbb{R}$. A basis for this space is $\{1, i\}$ since, again, every element in $\mathbb{C}$ can be uniquely expressed as

   $$a \cdot 1 + b \cdot i$$

   where $a, b \in \mathbb{R}$. It now follows that this space has dimension 2. We write this as $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

4. What is the dimension of the trivial vector space $\{0_V\}$? A basis for this space is the emptyset $\emptyset$, since by definition it is linearly independent and $\mathrm{span}(\emptyset) = \{0_V\}$. Therefore, this space has dimension $|\emptyset| = 0$.

We now turn our attention to proving some basic properties about dimension.

Theorem 2.10. *Let V be a finite-dimensional vector space. If L is any linearly independent set in V, then $|L| \leq \dim(V)$. Moreover, if $|L| = \dim(V)$, then L is a basis for V.*

Proof. By the Basis Extension Theorem, we know that there exists a basis $\mathcal{B}$ such that

$$L \subseteq \mathcal{B}\,. \tag{*}$$

This means that $|L| \leq |\,\mathcal{B}\,| = \dim(V)$. In the special case that $|L| = \dim(v) = |\,\mathcal{B}\,|$, then $(*)$ implies $L = \mathcal{B}$, i.e., L is a basis. $\square$

A useful application of this theorem is that whenever we have a set S of $n+1$ vectors sitting inside an n dimensional space, then we instantly know that S must be dependent. The following corollary, is another useful consequence of this theorem.

Corollary 2.11. *If V is a finite-dimensional vector space, and U is a subspace, then $\dim(U) \leq \dim(V)$.*

It might occur to the reader that an analogous statement for spanning sets should be true. That is, if we have a set S of $n-1$ vectors which sits inside an n-dimensional vectors space V can we conclude that $\mathrm{span}\, S \neq V$? As the next theorem shows, the answer is yes.

Theorem 2.12. *Let V be a finite-dimensional vector space. If S is any spanning set for V, then $\dim(V) \leq |S|$. Moreover, if $|S| = \dim(V)$ then S is a basis for V.*

To prove this theorem, we would like to employ similar logic as in the proof of the previous theorem but with Theorem 2.5 in place of Theorem 2.6. The problem with this is that S is not necessarily finite. Instead, we may use the following lemma in place of Theorem 2.6. Both its proof, and the proof of this lemma are left as exercises for the reader.

Lemma 2.13. *Let V be a finite-dimensional vector space. If S is any spanning set for V, then there exists a subset $\mathcal{B}$ of S, which is a basis for V.*

2.4 Zorn's lemma & the basis extension theorem

Definition. Let $\mathfrak{X}$ be a collection of sets. We say a set $B \in \mathfrak{X}$ is **maximal** if there exists no other set $A \in \mathfrak{X}$ such that $B \subset A$. A **chain** in $\mathfrak{X}$ is a subset $\mathfrak{C} \subseteq \mathfrak{X}$ such that for any two sets $A, B \in \mathfrak{C}$ either

$$A \subseteq B \qquad \text{or} \qquad B \subseteq A.$$

Lastly, we say $C \in \mathfrak{X}$ is an **upper bound** for a chain $\mathfrak{C}$ if $A \subseteq C$, for all $A \in \mathfrak{C}$.

Observe that if $\mathfrak{C}$ is a chain and $A_1, \ldots, A_m \in \mathfrak{C}$, then there exists some $1 \leq k \leq m$, such that

$$A_k = A_1 \cup A_2 \cup \cdots \cup A_m.$$

This observation follows by a simple induction, which we leave to the reader.

Zorn's Lemma. *Let $\mathfrak{X}$ be a collection of sets such that every chain $\mathfrak{C}$ in $\mathfrak{X}$ has an upper bound. Then $\mathfrak{X}$ has a maximal element.*

Lemma 2.14. *Let V be a vector space and fix a linearly independent subset L. Let $\mathfrak{X}$ be the collection of all linearly independent sets in V that contain L. If B is a maximal element in $\mathfrak{X}$, then B is a basis for V.*

Proof. By definition of the set $\mathfrak{X}$, we know that B is linearly independent. It only remains to show that $\operatorname{span}(B) = V$. Assume for a contradiction that it does not. This means there exists some vector $v \in V - \operatorname{span}(B)$. By the Linear Independence Theorem, $B \cup \{v\}$ is linearly independent and hence must be an element of $\mathfrak{X}$. This contradicts the maximality of B. We conclude that $\operatorname{span}(B) = V$ as desired. $\square$

Proof of Theorem 2.6. Let L be a linearly independent subset of V and define $\mathfrak{X}$ to be the collection of all linearly independent subsets of V containing L. In light of Lemma 2.14, it will suffice to prove that $\mathfrak{X}$ contains a maximal element. An application of Zorn's Lemma, assuming its conditions are met, therefore completes our proof. To show that we can use Zorn's Lemma, we need to check that every chain $\mathfrak{C}$ in $\mathfrak{X}$ has an upper bound. If $\mathfrak{C} = \emptyset$, then $L \in \mathfrak{X}$ is an upper bound. Otherwise, we claim that the set

$$C = \bigcup_{A \in \mathfrak{C}} A,$$

is an upper bound for our chain $\mathfrak{C}$. Clearly, $A \subset C$, for all $A \in \mathfrak{C}$. It now remains to show that $C \in \mathfrak{X}$, i.e., $L \subset C$ and C is independent. As $\mathfrak{C} \neq \emptyset$, then for any $A \in \mathfrak{C}$, we have

$$L \subseteq A \subseteq C.$$

To show C is independent, assume

$$a_1 v_1 + \cdots + a_m v_m = 0_V,$$

for some (distinct) $v_i \in C$ and $a_1 \in \mathbb{F}$. By construction of C each vector v_i must be an element of some A_i. As $\mathfrak{C}$ is a chain the above remark implies that

$$A_k = A_1 \cup A_2 \cup \cdots \cup A_m,$$

for some $1 \leq k \leq m$. Therefore all the vectors $v_1, \ldots, v_m$ lie inside the linearly independent set $A_k \in \mathfrak{X}$. This means our scalars $a_1, \ldots, a_m$ are all zero. We conclude that C is an independent set. $\square$

Chapter 3

Linear transformations

In this chapter, we study functions from one vector space to another. So that the functions of study are linked, in some way, to the operations of vector addition and scalar multiplication we restrict our attention to a special class of functions called linear transformations. Throughout this chapter V and W are always vector spaces over $\mathbb{F}$.

3.1 Definition & examples

Definition. We say a function $T : V \to W$ is a **linear transformation** or a **linear map** provided

$$T(u + v) = T(u) + T(v)$$

and

$$T(av) = aT(v) \text{ for all } v \in V, a \in \mathbb{F}$$

for all $u, v \in V$ and $a \in \mathbb{F}$. We denote the set of all such linear transformations, from V to W, by $\mathcal{L}(V, W)$.

To simplify notation we often write Tv instead of $T(v)$. It is not a coincidence that this simplified notation is reminiscent of matrix multiplication; we expound on this in Section 3.4.

Examples.

1. The function $T : V \to W$ given by $Tv = 0_W$ for all $v \in V$ is a linear map. Appropriately, this is called the **zero map**.
2. The function $I : V \to V$, given by $Iv = v$, for all $v \in V$ is a linear map. It is called the **identity map**.
3. Let A be an $m \times n$ matrix with real coefficients. Then $A : \mathbb{R}^n \to \mathbb{R}^m$ given by matrix-vector product is a linear map. In fact, we show in Section 3.4, that, in some sense, all linear maps arise in this fashion.

4. Recall the vector space $\mathcal{P}_{\leq n}(\mathbb{R})$. Then the map $T : \mathcal{P}_{\leq n}(\mathbb{R}) \to \mathbb{R}^n$ defined by
$$T(a_0 + a_1 x + \cdots a_n x^n) = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$
is a linear map.

5. Recall the space of continuous functions $\mathcal{C}$. An example of a linear map on this space is the function $T : \mathcal{C} \to \mathcal{C}$ given by $Tf = xf(x)$.

6. Recall that $\mathcal{D}$ is the vector space of all differentiable function $f : \mathbb{R} \to \mathbb{R}$ and $\mathcal{F}$ is the space of all function $g : \mathbb{R} \to \mathbb{R}$. Define the map $\partial : \mathcal{D} \to \mathcal{F}$, so that $\partial f = f'$. We see that ∂ is a linear map since
$$\partial(f + g) = (f + g)' = f' + g' = \partial f + \partial g$$
and
$$\partial(af) = (af)' = af' = a\partial f.$$

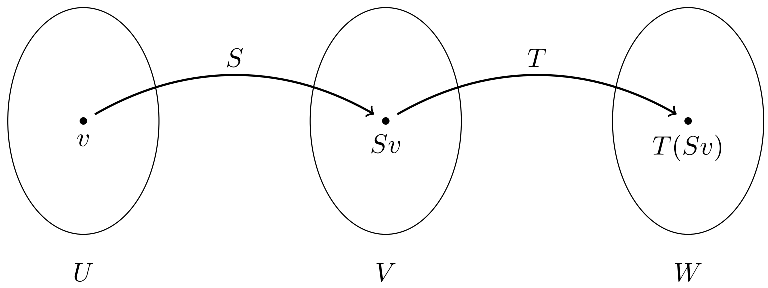
7. From calculus, we obtain another linear map $T : \mathcal{C} \to \mathbb{R}$ given by
$$Tf = \int_0^1 f \; dx.$$
The reader should convince himself that this is indeed a linear map.

Although the above examples draw from disparate branches of mathematics, all these maps have the property that they map the zero vector to the zero vector. As the next lemma shows, this is not a coincidence.

Lemma 3.1. *Let $T \in \mathcal{L}(V, W)$. Then $T(0_V) = 0_W$.*

Proof. To simplify notation let $0 = 0_V$. Now
$$T(0) = T(0 + 0) = T(0) + T(0).$$
Adding $-T(0)$ to both sides yields
$$T(0) + -T(0) = T(0) + T(0) + -T(0).$$
Since all these vectors are elements of W, simplifying gives us $0_W = T(0)$. □

It is often useful to "string together" existing linear maps to obtain a new linear map. In particular, let $S \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(V, W)$ where U is another $\mathbb{F}$-vector space. Then the function defined by
$$ST(v) = S(Tv)$$
is clearly a linear map in $\mathcal{L}(U, W)$. (The reader should verify this!) We say that TS is the **composition** or **product** of S with T. The reader may find the following figure useful for picturing the product of two linear maps.



There is another important way to combine two existing linear maps to obtain a third. If $S, T \in \mathcal{L}(V, W)$ and $a, b \in \mathbb{F}$, then we may define the function

$$(aS + bT)(v) = aSv + bTv$$

for any $v \in V$. Again we encourage the reader to check that this function is a linear map in $\mathcal{L}(V, W)$.

Before closing out this section, we first pause to point out a very important property of linear maps. First, we need to generalize the concept of a line in $\mathbb{R}^n$ to a line in an abstract vector space. Recall, that any two vector $v, u \in \mathbb{R}^n$ define a line via the expression $av + u$, where $a \in \mathbb{R}$. As this definition requires only vector addition and scalar multiplication we may "lift" it to the abstract setting. Doing this we have the following definition.

Definition. Fix vectors $u, v \in V$. We define a **line** in V to be all points of the form

$$av + u, \quad \text{where} \quad a \in \mathbb{F}.$$

Now consider applying a linear map $T \in \mathcal{L}(V, W)$ to the line $av + u$. In particular, we see that

$$T(av + u) = aT(v) + T(v).$$

In words, this means that the points on our line in V map to points on a new line in W, defined by the vectors $T(v), T(u) \in W$. In short we say that linear transformations have the property that they map lines to lines.

In light of the preceding lemma, even more is true. Observe that any line containing 0_V is of the form $av + 0_V$. (We think of such lines as analogues to lines through the origin in $\mathbb{R}^n$.) Lemma 3.1 now implies that such lines are mapped to lines of the form

$$T(av + 0_V) = aT(v) + T(0_W) = aT(v) + 0_W.$$

In other words, linear transformations actually map lines through the origin in V to lines through the origin in W.

3.2 Rank-nullity theorem

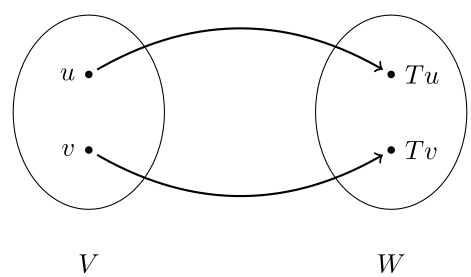
The aim of this section is to prove the Rank-Nullity Theorem. This theorem describes a fundamental relationship between linear maps and dimension. An immediate consequence of this theorem, will be an beautiful proof to the fact that a homogeneous system of equations with more variables than equations must have an infinite number of solutions.

We begin with the following definition.

Definition. Let $T : V \to W$ be a linear map. We say T is **injective** if $Tv = Tu$ implies that $u = v$. On the other hand, we say that T is **surjective** provided that for every $w \in W$, there exists some $v \in V$ such that $Tv = w$. A function that is both injective and surjective is called **bijective**.

It might help to think of T as a cannon that shoots shells (elements in V) at targets (elements of W)[1]. From this perspective there is an easy way to think about surjectivity and injectivity.

- Surjectivity means that the cannon T hits every element in W.
- Injectivity means that every target in W is hit at most once.
- Bijectivity means that every target is hit exactly once. In this case we can think of T as "matching up" the elements in T with the elements in W.



As is always the case in mathematics, it will be beneficial to have more than one description of a single idea. Our next lemma provides this alternative description of injectivity and surjectivity.

Definition. Fix $T \in \mathcal{L}(V, W)$. Define the **null space** of T to be

$$\operatorname{null} T = \{v \in V \mid Tv = 0_W\}$$

and the **range** of T to be

$$\operatorname{ran} T = \{Tv \mid v \in V\}.$$

[1]D. Saracino, A first course in abstract algebra.

Observe that the null space is a subset of V where as the range is a subset of W.

Lemma 3.2. *Let $T \in \mathcal{L}(V, W)$.*

1. *T is surjective if and only if $\operatorname{ran} T = W$.*
2. *T is injective if and only if $\operatorname{null} T = \{0_V\}$.*

Proof. Saying that T is surjective is equivalent to saying that for any $w \in W$ there exists some $v \in V$ such that $Tv = w$. In other words,

$$\operatorname{ran} T = \{Tv \mid v \in V\} = W,$$

as claimed. For the second claim, begin by assume T is injective. This means that $u = v$ whenever $Tu = Tv$. Consequently,

$$\operatorname{null} T = \{v \in V \mid Tv = 0_W = T(0_V)\} = \{0_V\},$$

as desired. For the other direction, assume $\{0_V\} = \operatorname{null} T$ and consider any two vectors $u, v \in V$ such that $Tu = Tv$. To prove T is injective we must show that $u = v$. To this end, the linearity of T yields

$$T(u - v) = 0_W.$$

This means that $u - v \in \operatorname{null} T = \{0_V\}$. Hence $u - v = 0_V$ or $u = v$ as desired. $\square$

The next lemma, whose proof we leave to the reader, states that our two sets $\operatorname{null} T$ and $\operatorname{ran} T$ are no ordinary sets; they are vector spaces in their own right.

Lemma 3.3. *Let $T \in \mathcal{L}(V, W)$. Then $\operatorname{null}(T)$ is a subspace of V and $\operatorname{ran}(T)$ is a subspace of W.*

Examples.

1. Consider the linear map $T : \mathbb{R}^{n+1} \to \mathcal{P}_{\leq n}$ given by

   $$T(a_0, \dots, a_{n+1}) = a_0 + a_1 x + \cdots a_n x^n.$$

   Then $\operatorname{null} T = \{(0, \dots, 0)\} = \{0_{\mathbb{R}^2}\}$ and $\operatorname{ran} T = \mathcal{P}_{\leq n}$.

2. The $T : \mathcal{P} \to \mathcal{P}$ given by $T(f)(x) = f'(x)$ is linear with $\operatorname{null} T =$ constant polynomials and $\operatorname{ran} T = \mathcal{P}$. Therefore, T is surjective but not injective.

3. Consider the map $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by

   $$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ -(x+y) \end{bmatrix}.$$

Then $\operatorname{null} T = \{0_{\mathbb{R}^2}\}$, but $\operatorname{ran} T \neq \mathbb{R}^3$. In fact,

$$\operatorname{ran} T = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \;\mid x + y + z = 0 \right\}$$

which is the equation of a plane in $\mathbb{R}^3$. So this map if injective but not surjective.

Before reading any further, can you spot a relation among the dimensions of the domain, range and null space in example 3)?

As the next theorem shows, there is one and an important one at that! The reader should check that in fact, both examples 1) and 3) do indeed satisfy this theorem.

Theorem 3.4 (Rank-Nullity)**.** *Assume V is finite-dimensional. For any $T \in \mathcal{L}(V, W)$,*

$$\dim V = \dim(\operatorname{null} T) + \dim(\operatorname{ran} T).$$

Proof. As $\operatorname{null} T$ is a subspace of V, and hence a vector space in its own right, it has a basis. Let $\{e_1, \ldots, e_k\}$ be such a basis for $\operatorname{null} T$. By the Basis Extension Theorem (Theorem 2.6), there exists vectors $f_1, \ldots, f_m$ so that

$$\mathcal{B} = \{e_1, \ldots e_k, f_1, \ldots, f_m\}$$

is a basis for V. Since

$$\dim V = k + m = \dim \operatorname{null} T + m,$$

we must show $\dim \operatorname{ran} T = m$. To this end, it suffices to show that $S = \{Tf_1, \ldots, Tf_m\}$ is a basis for $\operatorname{ran} T$. We do this in two parts.

We first show that $\operatorname{span}(S) = \operatorname{ran} T$. This readily follows from the following:

$$\begin{aligned} \operatorname{ran} T &= \{Tv \mid v \in V\} \\ &= \{T(a_1 e_1 + \cdots + a_k e_k + b_1 f_1 + \cdots + b_m f_m) \mid a_i, b_i \in \mathbb{F}\} \\ &= \{a_1 T e_1 + \cdots + a_k T e_k + b_1 T f_1 + \cdots + b_m T f_m) \mid a_i, b_i \in \mathbb{F}\} \\ &= \{b_1 T f_1 + \cdots + b_m T f_m) \mid b_i \in \mathbb{F}\} \\ &= \operatorname{span}(S), \end{aligned}$$

where the second equality uses the fact that $\mathcal{B}$ is a basis for V and the third equality follows since $e_1, \ldots, e_k \in \operatorname{null} T$.

We now turn our attention to showing that S is linearly independent. To do this we must show that if $a_1 T f_1 + \cdots + a_m T f_m = 0_W$, then all the scalars are zero. By the linearity of T we have

$$0_W = T(a_1 f_1 + \cdots + a_m f_m),$$

which means that $a_1 f_1 + \cdots + a_m f_m \in \operatorname{null} T$. As $e_1, \ldots, e_k$ are a basis for $\operatorname{null} T$ it follows that

$$b_1 e_1 + \cdots + b_k e_k = a_1 f_1 + \cdots + a_m f_m,$$

for some scalars b_i. Rearranging we see that

$$0_V = -(b_1 e_1 + \cdots + b_k e_k) + a_1 f_1 + \cdots + a_m f_m.$$

The linear independence of $\mathcal{B} = \{e_1, \ldots e_k, f_1, \ldots, f_m\}$ forces all the scalars to be zero. In particular, $a_1 = a_2 = \cdots = 0$ as needed. As we have shown that S is independent and spans $\operatorname{ran} T$, we may conclude that it is a basis for $\operatorname{ran} T$ as desired. $\square$

To motivate our first corollary recall the following fact about plain old sets. If X and Y are sets and $f : X \to Y$ is injective then $|X| \leq |Y|$. On the other hand if f is surjective, then $|X| \geq |Y|$. As dimension measures the "size" of a vector space, the following is the vector space analogue to this set theory fact.

Corollary 3.5. *Let V and W be finite-dimensional vector spaces and let T be an arbitrary linear map in $\mathcal{L}(V, W)$.*

1. *If $\dim V > \dim W$, then T is not injective.*
2. *If $\dim V < \dim W$, then T is not surjective.*

Proof. Fix $T \in \mathcal{L}(V, W)$. To prove the first claim, assume $\dim V > \dim W$. By the Rank-Nullity Theorem we see that

$$\dim(\operatorname{null} T) = \dim V - \dim(\operatorname{ran} T) \leq \dim V - \dim W > 0,$$

where the second inequality follows since $\operatorname{ran} T$ is a subspace of W. Consequently, $\operatorname{null} T$ is not the trivial space, so by Lemma 3.2 T is not injective. Likewise, if $\dim V < \dim W$, then

$$\dim(\operatorname{ran} T) = \dim V - \dim(\operatorname{null} T) \leq \dim V < \dim W.$$

Consequently, $\operatorname{ran} T$ cannot be all of W, i.e., T is not surjective. $\square$

In general mathematical function can be injective without being surjective and vice versa. For example consider the functions $f, g : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ and

$$g(n) = \begin{cases} n+1 & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ n+2 & \text{if } n < 0. \end{cases} .$$

Then f is injective but not surjective and g is surjective but not injective since $g(-2) = 0 = g(0)$. Consequently the next theorem is quiet amazing. It states that if two vector spaces have the same dimension, then a linear map between them is surjective if and only if it is injective!

Corollary 3.6. *Let V and W be finite-dimensional vector spaces with the same dimension. For any linear map $T \in \mathcal{L}(V, W)$ we have that T is surjective if and only if T is injective.*

Proof. By the Rank-Nullity Theorem states we always have

$$\dim V = \dim(\operatorname{null} T) + \dim(\operatorname{ran} T).$$

Now, observe that

$$\begin{aligned}
T \text{ is surjective } &\iff \operatorname{ran} T = W \\
&\iff \dim(\operatorname{ran} T) = \dim W \\
&\iff \dim(\operatorname{ran} T) = \dim V \\
&\iff \dim(\operatorname{null} T) = 0 \\
&\iff \operatorname{null} T = \{0_V\} \\
&\iff T \text{ is injective,}
\end{aligned}$$

where the third equivalence is the fact that $\dim V = \dim W$ and the fourth equivalence follows from the Rank-Nullity Theorem.

□

Before closing this section, let demonstrate two beautiful applications of the Rank-Nullity Theorem. First, consider a homogeneous system of linear equations

$$\begin{aligned}
a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\
a_{21}x_1 + \ldots + a_{2n}x_n &= 0 \\
&\vdots \\
a_{m1}x_1 + \ldots + a_{1n}x_n &= 0.
\end{aligned}$$

A standard result from any elementary linear algebra course is that if this system has more variable than equations ($n > m$), then a non-trivial solution to the system exists, i.e., one other than $x_1 = \cdots x_n = 0$. We are now in a position to give an elegant proof of this fact. First, rewrite this system in matrix form as $A\mathbf{x} = \mathbf{0}$, where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Recall that $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map given by matrix vector multiplication. As $n > m$, Corollary 3.5 states that T is not injective and hence $\{0_V\} \subsetneq \operatorname{null} T$. Therefore there exists some nonzero $\mathbf{x} \in \operatorname{null} T$. As $\mathbf{x}$ is nonzero and $A\mathbf{x} = \mathbf{0}$, we see that our system has a nontrivial solution as claimed.

For our second application, consider a system of (not necessarily homogeneous) linear equations

$$\begin{aligned}
a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + \ldots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + \ldots + a_{1n}x_n &= b_m.
\end{aligned}$$

In matrix form this becomes $A\mathbf{x} = \mathbf{b}$. Another standard result from elementary linear algebra is that if our system has more equations than unknowns, i.e., $n < m$, then there exists some choice of $\mathbf{b} \in \mathbb{R}^m$ so that our system is inconsistent (has no solutions). To prove this, again think of $A : \mathbb{R}^n \to \mathbb{R}^m$ as a linear map. Corollary 3.5 tell us that A is not surjective. This means there exists some $\mathbf{b} \in \mathbb{R}^m$ so that no choice of $\mathbf{x} \in \mathbb{R}^n$ gives $A\mathbf{x} = \mathbf{b}$. In other words, for this b, our system is inconsistent.

3.3 Vector space isomorphisims

Our next goal is to identify when two vectors spaces are essential the same. Mathematically, we say they are *isomorphic* which is latin for "same shape". To see what we mean by this, imagine you are given an $\mathbb{F}$-vector space V and you paint all its elements red to obtain a new space W. Although this new space W "looks" different (all its vectors are red!), it still has the same algebraic structure as V.

A more concrete example of isomorphic vector spaces has been in front of us almost since page one! In fact, it might have already occurred to you that as vector spaces $\mathbb{R}^{n+1}$ and $\mathcal{P}_{\leq n}$ were strikingly similar. Certainly as sets they are very different – one is a set of vectors while the other is a set of polynomials! In terms of their vector space structure this is just a cosmetic difference. To convince you note that an arbitrary vector in $\mathbb{R}^{n+1}$ looks like

$$\begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$$

while an arbitrary vector in $\mathcal{P}_{\leq n}$ looks like $a_0 + a_1 x + \cdots + a_n x^n$. In either case, a vector is just a list of $n+1$ numbers $\{a_0, \ldots, a_n\}$. Moreover, the operation of addition and scalar multiplication are essentially the same in both spaces too. For example, addition in $\mathbb{R}^{n+1}$ is given by

$$\begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_0 + b_0 \\ \vdots \\ a_n + b_n \end{bmatrix}$$

which is really no different then addition in $\mathcal{P}_{\leq n}$ which looks like

$$(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n)$$
$$= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n.$$

. Intuitively, these two spaces are the "same". With this example in mind consider the formal definition for vector spaces to be isomorphic.

Definition. We say two vector spaces V and W are **isomorphic** and write $V \cong W$, if there exists $T \in \mathcal{L}(V, W)$ which is both injective and surjective. We call such a T an **isomorphism**.

Theorem 3.7. *Two finite-dimension vector spaces V and W are isomorphic if and only if they have the same dimension.*

Proof. Assume V and W are isomorphic. This means there exists a linear map $T : V \to W$ that is both surjective and injective. Corollary 3.5 immediately implies that $\dim V = \dim W$. For the reverse direction, let $\mathcal{B}_V = \{v_1, \ldots, v_n\}$ be a basis for V and $\mathcal{B}_W = \{w_1, \ldots, w_n\}$ be a basis for W. As every vector $v \in V$ can be written (uniquely) as

$$v = a_1v_1 + \cdots + a_nv_n$$

for $a_i \in \mathbb{F}$, we may define a function $T : V \to W$ by

$$Tv = a_1w_1 + \cdots + a_nw_n.$$

Observe that the uniqueness of our representation of v implies that T is a well-defined function. Moreover, a straightforward check reveals that T is indeed a linear map. It only remains to show that T is an isomorphism. To see that T is injective, let that $v \in \operatorname{null} T$ and let $b_i \in \mathbb{F}$ be such that $v = b_1v_1 + \cdots + b_nv_n$. This means

$$0_W = Tv = b_1w_1 + \cdots + b_nw_n.$$

Since $\mathcal{B}_W$ is an independent set, it follows that all our scalars b_i must be 0 and, in turn, $v = 0$. This shows that $\operatorname{null} T = \{0_V\}$, i.e., T is injective.

To see that T is also surjective, note that any vector $w \in W$ can be written as

$$w = c_1w_1 + \cdots + c_mw_m,$$

for some choice of scalars c_i (why?). Now consider the vector $c_1v_1 + \cdots + c_mv_m \in V$ and observe that

$$T(c_1v_1 + \cdots + c_mv_m) = c_1w_1 + \cdots + c_mw_m = w.$$

This shows that T is surjective. □

Definition. Let $T \in \mathcal{L}(V, W)$. We say T is **invertible** provided there exists some $S \in \mathcal{L}(W, V)$ so that $ST : V \to V$ is the identity map on V and $TS : W \to W$ is the identity map on W. We call S an **inverse** of T.

As a consequence of the next lemma, we are able to refer to *the* inverse of T which we denote by T^{-1}.

Lemma 3.8. *Let $T \in \mathcal{L}(V, W)$. If T is invertible, then its inverse is unique.*

Proof. Assume S and S' are both inverses for T. Then

$$S = SI_W = STS' = I_V S' = S'$$

where I_V and I_W are the identity maps on V and W respectively. □

Lemma 3.9. *Let $T \in \mathcal{L}(V, W)$. Then T is invertible if and only if T is an isomorphism.*

Proof. Let us first assume that T is invertible. We must prove that T is both injective and surjective. To see injectivity, let $u \in \operatorname{null} T$, then

$$u = I_V u = T^{-1}Tu = T^{-1}0_W = 0_V,$$

where I_V is the identity map on V. We conclude that $\operatorname{null} T = \{0_V\}$ and hence T is injective. To see that T is also surjective fix $w \in W$. Observe that T maps the vector $T^{-1}w \in V$ onto w since

$$T(T^{-1}w) = TT^{-1}w = I_W w = w.$$

We conclude that T is surjective.

Now assume $T : V \to W$ is an isomorphism. As T is both injective and surjective, then for every $w \in W$ there exists exactly one $v \in V$ so that $Tv = w$. Now define the function $S : W \to V$ by $S(w) = v$. We claim $T^{-1} = S$. By definition we have $S(Tv) = v$ for all $v \in V$ and $TS(w) = w$ for all $w \in W$. It only remains to show that S is linear. Let $w_1, w_2 \in W$ and let v_1, v_2 be the unique vectors in V so that $Tv_i = w_i$. As T is linear then $T(v_1 + v_2) = w_1 + w_2$. By definition of S we now have

$$S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2).$$

Likewise,

$$S(aw_1) = aw_1 = aS(w_1).$$

We may now conclude that $T^{-1} = S$ and hence T is invertible. □

3.4 The matrix of a linear transformation

In this section we study a striking connection between linear transformations and matrices. In fact, we will see that linear transformations and matrices are really two sides of the same coin! Before beginning let us review the basics of matrix multiplication. Let $A = [\vec{a}_1 \cdots \vec{a}_n]$ where $\vec{a}_i \in \mathbb{R}^m$, so that A is an $m \times n$ matrix whose ith column is the vector $\vec{a}_i$. For any $\vec{b} \in \mathbb{R}^n$ we define

$$A\vec{b} := [\vec{a}_1 \cdots \vec{a}_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b_1\vec{a}_1 + \cdots + b_n\vec{a}_n.$$

So $A\vec{b}$ is the linear combintion of the columns of A using the coefficients in $\vec{b}$.

Example.

$$\begin{bmatrix} 1 & 4 & 5 \\ 2 & 3 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 4 \\ 3 \end{bmatrix} + c \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

Moreover, if $B = [\vec{b}_1 \cdots \vec{b}_k]$ where each $\vec{b}_i \in \mathbb{R}^n$, then we define

$$AB = A[\vec{b}_1 \cdots \vec{b}_k] = [A\vec{b}_1 \cdots A\vec{b}_k],$$

so that the ith column of AB is $A\vec{b}_i$.

With this review under our belt, let us begin our study. As usual, fix finite-dimensional $\mathbb{F}$-vector spaces V and W with bases $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V and $\mathcal{C} = \{w_1, \ldots, w_m\}$ for W.

Definition. For any $v \in V$, we define

$$[v]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

where $v = b_1v_1 + \cdots + b_nv_n$. We call this vector the **coordinates** of v with respect to the basis $\mathcal{B}$.

The next fact follows directly from the definition of coordinates. We leave its proof to the reader.

Lemma 3.10 (Linearity of Coordinates)**.** *For any $u, v \in V$ and $a, b \in \mathbb{F}$, we have*

$$[au + bv]_{\mathcal{B}} = a[u]_{\mathcal{B}} + b[v]_{\mathcal{B}}$$

Now fix $T \in \mathcal{L}(V, W)$ and consider the action of T on some $v \in V$. First let

$$[v]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Now we see that

$$\begin{aligned} Tv &= T(b_1v_1 + \cdots + b_nv_n) \\ &= b_1Tv_1 + \cdots + b_nTv_n. \end{aligned}$$

As $Tv_i \in W$ and $\mathcal{C}$ is a basis for W we must have

$$Tv_i = b_{1i}w_1 + \cdots + b_{mi}w_m$$

for some scalars $b_{ij} \in \mathbb{F}$ where $1 \le i \le m$ and $1 \le j \le n$. Solving for the coordinates of Tv with respect to $\mathcal{C}$ we have

By the linearity of coordinates we now have

$$[Tv]_{\mathcal{C}} = [b_1Tv_1 + \cdots + b_nTv_n]_{\mathcal{C}} = b_1[Tv_1]_{\mathcal{C}} + \cdots + b_n[Tv_n]_{\mathcal{C}}.$$

If we set

$$[Tv_i]_{\mathcal{C}} = \begin{bmatrix} c_{1i} \\ \vdots \\ c_{mi} \end{bmatrix}$$

then we see that

$$[Tv]_{\mathcal{C}} = b_1 \begin{bmatrix} c_{11} \\ \vdots \\ c_{m1} \end{bmatrix} + b_2 \begin{bmatrix} c_{12} \\ \vdots \\ c_{m2} \end{bmatrix} + \cdots + b_n \begin{bmatrix} c_{1n} \\ \vdots \\ c_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

This discussion proves the following theorem.

Theorem 3.11. *Fix $T \in \mathcal{L}(V, W)$. Then*

$$[Tv]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}\ [v]_{\mathcal{B}},$$

where $[T]_{\mathcal{B}}^{\mathcal{C}}$ is the matrix whose columns are $[Tv_i]_{\mathcal{C}}$. That is

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \end{bmatrix} \quad \textit{where} \quad [Tv_i]_{\mathcal{C}} = \begin{bmatrix} c_{1i} \\ \vdots \\ c_{mi} \end{bmatrix}.$$

The reader should take note that $[T]_{\mathcal{B}}^{\mathcal{C}}$ is an $m \times n$ matrix and T is a linear map from an n-dimensional space V to an m-dimensional space W.

Having developed this connection let consider the following concrete example.

Examples.

1. Define the linear map
$$T : \mathbb{R}^3 \to \mathcal{P}_{\leq 2}$$
by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = (a+b)x + (a+c)x^2$. Now let
$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$
be our basis for $\mathbb{R}^3$ and take $\mathcal{C} = \{1, x, x^2\}$ as our basis for $\mathcal{P}_{\leq 2}$. As
$$T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = 2x+x^2, \quad T\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = x+2x^2, \quad T\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = x+x^2,$$
we have
$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}.$$
Next, consider the vector $v = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \in \mathbb{R}^3$ so that $[v]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. By Theorem 3.11 we have
$$[Tv]_{\mathcal{C}} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 7 \\ 8 \end{bmatrix},$$
which is consistent as $Tv = 7x + 8x^2$.

2. For any n-dimensional vector space V, let $I : V \to V$ be the identity map. If $\mathcal{B}$ is any basis for V, the reader should check that
$$[I]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$
where the matrix on the right has size $n \times n$. For obvious reasons this matrix is often called the **identity** matrix and is denoted by I_n.

In the past, you may have wondering why matrix multiplication is defined in the way it is. Why not just multiply corresponding entries? Why is this definition so great? We are now in a position to answer this question! Consider our next theorem.

Theorem 3.12. *Let U be another finite-dimensional vector space with basis $\mathcal{D}$. For any $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$, we have*

$$[ST]_{\mathcal{B}}^{\mathcal{D}} = [S]_{\mathcal{C}}^{\mathcal{D}} \cdot [T]_{\mathcal{B}}^{\mathcal{C}}$$

where the operation $\cdot$ on the left is the usual matrix multiplication.

In other words, matrix multiplication is precisely the operation needed in order to obtain the matrix $[ST]_{\mathcal{B}}^{\mathcal{D}}$ from the matrices $[S]_{\mathcal{C}}^{\mathcal{D}}$ and $[T]_{\mathcal{B}}^{\mathbb{C}}$. Another way to look at this is that composition of functions on the linear transformation side corresponds to matrix multiplication on the matrix side. Let us now prove this pretty result.

Proof. To simplify notation let us denote $[T]_{\mathcal{B}}^{\mathcal{C}}$ by $[T]$ and $[S]_{\mathcal{C}}^{\mathcal{D}}$ by $[S]$. Theorem 3.11 tells us that

$$[STv_i]_{\mathcal{D}} = [S][Tv_i]_{\mathcal{C}}.$$

As $ST \in \mathcal{L}(U, W)$, then

$$[ST]_{\mathcal{B}}^{\mathcal{D}} = \Big[[STv_1]_{\mathcal{D}} \cdots [STv_n]_{\mathcal{D}}\Big] = \Big[[S][Tv_1]_{\mathcal{C}} \cdots [S][Tv_n]_{\mathcal{C}}\Big] = [S][T],$$

where the last equality is our definition of matrix multiplication. $\square$

We are now in a position to answer our last fundamental question regarding this correspondence between linear transformations and matrices. Assume $\mathcal{B}'$ and $\mathcal{C}'$ are also bases for V and W respectively. For any $T \in \mathcal{L}(V, W)$ how are the matrices $[T]_{\mathcal{B}}^{\mathcal{C}}$ and $[T]_{\mathcal{B}'}^{\mathcal{C}'}$ related to each other?

Theorem 3.13 (Change of Basis)**.**

$$[T]_{\mathcal{B}'}^{\mathcal{C}'} = [J]_{\mathcal{C}}^{\mathcal{C}'} \; [T]_{\mathcal{B}}^{\mathcal{C}} \; [I]_{\mathcal{B}'}^{\mathcal{B}},$$

where J and I are the identity maps on V and W respectively.

Proof. Certainly we have the following equality

$$T = JTI.$$

Several applications of Theorem 3.12 now yields

$$[T]_{\mathcal{B}'}^{\mathcal{C}'} = [JTI]_{\mathcal{B}'}^{\mathcal{C}'} = [J]_{\mathcal{C}}^{\mathcal{C}'} \; [TI]_{\mathcal{B}'}^{\mathcal{C}} = [J]_{\mathcal{C}}^{\mathcal{C}'} \; [T]_{\mathcal{B}}^{\mathcal{C}} \; [I]_{\mathcal{B}'}^{\mathcal{B}}.$$

$\square$

You might be mislead into thinking that $[I]_{\mathcal{B}'}^{\mathcal{B}}$ must be the identity matrix since I is the identity map plus we know this to be true when $\mathcal{B} = \mathcal{B}'$ (see Example 2 above). The following simple example shows that this is NOT generally true. Consider the identity map $I : \mathbb{R}^3 \to \mathbb{R}^3$ where $\mathcal{B}'$ is the standard basis and $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$. Since

$$e_1 = \frac{1}{2}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right), \quad e_2 = \frac{1}{2}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right),$$

and

$$e_3 = \frac{1}{2}\left(- \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right)$$

We see that

$$[I]_{\mathcal{B}'}^{\mathcal{B}} = \Big[[e_1]_{\mathcal{B}} \ [e_2]_{\mathcal{B}} \ [e_3]_{\mathcal{B}} \Big] = \begin{bmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{bmatrix}$$

is certainly not the identity matrix. We hope this example dispels any possible confusion.

Chapter 4

Complex operators

In this chapter we study linear transformations that map a vector space V into itself. Such maps, formally called operators, are the foundation for a rich theory. This chapter aims to provide an introduction to this theory.

4.1 Operators & polynomials

Definition. We call a linear map T from V to V an **operator**. If V is a complex vector space we say T is a **complex operator**. Likewise, if V is a real vector space we say T is a **real operator**. We denote the set of operators $\mathcal{L}(V)$ instead of $\mathcal{L}(V, V)$.

As the domain and range of an operator $T \in \mathcal{L}(V)$ is the same, we may compose such function with themselves an arbitrary number of times. In particular, we write

$$T^i = \underbrace{TT \cdots T}_{i},$$

for integers $i > 0$ and define $T^0 = I_V$ to be the identity map on V. If $p(x) = a_0 + a_1 x + \cdots a_n x^n$ is any polynomial with coefficients in $\mathbb{F}$ we define the function $p(T) : V \to V$ given by

$$p(T)u = a_0 u + a_1 Tu + \cdots + a_n T^n u,$$

for any $u \in V$. A straightforward check, which we leave to the reader, shows that $p(T)$ is an operator. For example consider the operator $T : \mathbb{C} \to \mathbb{C}$ given by

$$T\left(\left[\begin{array}{c} a \\ b \end{array}\right]\right) = \left[\begin{array}{c} 2a \\ 3b \end{array}\right].$$

If $p(x) = 2 + 3x + x^2$, then

$$\begin{aligned} p(T)\begin{bmatrix} a \\ b \end{bmatrix} = (2 + 3T + T^2)\begin{bmatrix} a \\ b \end{bmatrix} &= 2\begin{bmatrix} a \\ b \end{bmatrix} + 3T\begin{bmatrix} a \\ b \end{bmatrix} + T^2\begin{bmatrix} a \\ b \end{bmatrix} \\ &= \begin{bmatrix} 2a \\ 2b \end{bmatrix} + 3\begin{bmatrix} 2a \\ 3b \end{bmatrix} + \begin{bmatrix} 4a \\ 9b \end{bmatrix} \\ &= \begin{bmatrix} 12a \\ 20b \end{bmatrix}. \end{aligned}$$

Continuing in this vain we note that if $f(x), g(x)$, and $h(x)$ are polynomials such that $f(x) = g(x)h(x)$, then $f(T) = g(T)h(T)$. We encourage the reader to check this fact for themselves. Moreover, since $f(x) = g(x)h(x) = h(x)g(x)$ we see that

$$g(T)h(T) = h(T)g(T). \tag{4.1}$$

Returning to the example above, observe that $p(x) = (1 + x)(2 + x)$. Computing we obtain

$$\begin{aligned} p(T)\begin{bmatrix} a \\ b \end{bmatrix} = (1 + T)(2 + T)\begin{bmatrix} a \\ b \end{bmatrix} &= (1 + T)\left(2\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} 2a \\ 3b \end{bmatrix}\right) \\ &= (1 + T)\begin{bmatrix} 4a \\ 5b \end{bmatrix} \\ &= \begin{bmatrix} 4a \\ 5b \end{bmatrix} + \begin{bmatrix} 8a \\ 15b \end{bmatrix} = \begin{bmatrix} 12a \\ 20b \end{bmatrix}, \end{aligned}$$

which agrees with our computation above. The reader is encouraged to check that one gets the same answer by computing

$$(2 + T)(1 + T)\begin{bmatrix} a \\ b \end{bmatrix}.$$

For the remainder of this section we review some basic facts about polynomials that will be needed throughout the remainder of this chapter. We start with one of the most famous theorems in algebra.

Theorem 4.1. *(Fundamental Theorem of Algebra) Every non-constant $p(z) \in \mathcal{P}(\mathbb{C})$ has at least one root. Consequently, $p(z)$ factors completely into linear terms as*

$$p(z) = a(z - \lambda_1) \cdots (z - \lambda_n),$$

for some $c, \lambda_i \in \mathbb{C}$.

Recall that for integers a, b, we say a divides b, and write $a|b$ provided there exists some integer c such that $b = ac$. We now give an analogous definition for polynomials.

Definitions. For polynomials $p(x), q(x)$ we say $q(x)$ **divides** $p(x)$ and write $q(x)|p(x)$, provided there exists some polynomials $s(x)$ such that $p(x) = q(x)s(x)$.

We define the **greatest common divisor** or **GCD** of $p(x)$ and $q(x)$ to be the monic polynomials of greatest degree, written (p, q), that divides both p and q. We can easily extend this definition to a finite number of polynomials $p_1, \dots, p_k$ by defining their GCD $(p_1, \dots, p_k)$ to be the monic polynomials of largest degree that divides each p_i.

Lastly, we say $p_1, \dots, p_k$ are **relatively prime** provided their GCD is 1.

For example, observe that $x-1|x^3-1$ since $(x-1)(x^2+x+1) = x^3-1$. We also have that that the polynomials $(x-a)^n$ and $(x-b)^m$ are relatively prime provided $a \neq b$ and if $a = b$ then their GCD is $(x-a)^k$, where $k = \min(n, m)$.

The next lemma, whose proof can be readily found in many algebra books, will serve as a critical ingredient in Section 4.4.

Lemma 4.2 (Bézout). *Let $p_1(x), \dots, p_k(x) \in \mathcal{P}(\mathbb{C})$ with GCD $d(x)$. Then there exists polynomials, $a_1(x), \dots, a_k(x) \in \mathcal{P}(\mathbb{C})$ such that*

$$a_1(x)p_1(x) + \cdots + a_k(x)p_k(x) = d(x).$$

4.2 Eigenvectors & eigenvalues

Definition. Let $T \in \mathcal{L}(V)$. We say a subspace U of V if **T-invariant** provided that $T(U) = \{Tu \mid u \in U\} \subseteq U$.

The utility of this definition is that if U is T-invariant, then we can restrict T to the "smaller" linear map

$$T : U \to U.$$

A particularly, important T-invariant subspace is given by the following definition.

Definition. Let $T \in \mathcal{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is said to be an **eigenvalue**, if there is a nonzero vector $v \in V$, such that

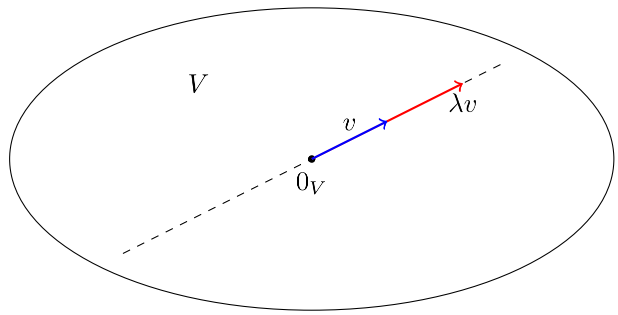
$$Tv = \lambda v.$$

In this case, we say v is an **eigenvector** for λ and refer to the tuple (λ, v) as an **eigenpair** for T.

Observe that if (λ, v) is an eigenpair for T then $\text{span}(v)$ is T-invariant and

$$T : \text{span}(v) \to \text{span}(v)$$

is given by multiplication by λ. Geometrically, this means that along the line generated by v, i.e., $\text{span}(v)$, T acts by simply stretching the line by a factor of λ as if the line were made of rubber. To illustrate consider the following picture where the dashed line is $\text{span}(v)$.



The significance of this is that although T may be wildly complicated in general, the existence of an eigenpair (λ, v) means that at least in the v direction T acts in a very simple way – it just stretches the line by a factor of λ.

Examples.

1. For the identity operator, I_V, every nonzero vector v is an eigenvector with corresponding eigenvalue 1. What are the eigenvalues/vectors for the zero map on V?

2. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$. As an operator $A : \mathbb{R}^2 \to \mathbb{R}^2$ we see (by inspection) that
$$\left(2, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \quad \text{and} \quad \left(-1, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$$
are its two eigenpairs.

3. Now define the operator T on $\mathbb{R}^2$ by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+2y \\ 4y-x \end{bmatrix}$. Unlike the previous example, it is a bit harder to just spot the eigenpairs for this operator. Instead, let us appeal to some simple algebra. We seek values x, y and λ such that
$$\begin{bmatrix} x+2y \\ 4y-x \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$
Equating the first and second coordinates we obtain two equations:
$$x + 2y = \lambda x \qquad \text{and} \qquad 4y - x = \lambda y. \tag{$\star$}$$
Using the second equation to obtain an expression for x we then plug this expression into the first equation to obtain (after simplifying)
$$0 = y(\lambda^2 - 5\lambda + 6) = y(\lambda - 3)(\lambda - 2). \tag{4.2}$$
Therefore $y = 0$ or $\lambda = 3, 2$. We see from above that when $y = 0$, x is also 0. Since the zero vector cannot be an eigenvector we are not interested in this solution. If $\lambda = 2$, then solving ($\star$) yields $x = 2, y = 1$. Hence

$\left(2, \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$ is an eigenpair for T. Likewise, when $\lambda = 3$, we get the eigenpair $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$.

4. Let T be the operator on the space of continuous function $\mathcal{C}(\mathbb{R})$ given by

$$Tf = \int f(x)\; dx.$$

Now consider the function $e^{\lambda x}$ with $\lambda \neq 0$. Since

$$T(e^{\lambda x}) = \int e^{\lambda x}\; dx = \frac{1}{\lambda} e^{\lambda x}$$

we see that $\left(\frac{1}{\lambda}, e^{\lambda x}\right)$ is an eigenpair for this operator.

Lemma 4.3. *Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then the nonzero vectors in* null$(T - \lambda)$ *are precisely the eigenvectors for T corresponding to λ.*

Proof. To prove this lemma we must show that $0_V \neq v \in \operatorname{null}(T - \lambda)$ if and only if (λ, v) is an eigenpair for T. This follows directly from the definitions involved. In particular, we have

$$\begin{aligned} 0_V \neq v \in \operatorname{null}(T - \lambda) &\Leftrightarrow (T - \lambda)v = 0_V \text{ and } v \neq 0_V \\ &\Leftrightarrow Tv - \lambda v = 0_V \text{ and } v \neq 0_V \\ &\Leftrightarrow Tv = \lambda v \text{ and } v \neq 0_V \\ &\Leftrightarrow (\lambda, v) \text{ is an eigenpair for } T. \end{aligned}$$

□

We first remark that a straightforward consequence of this lemma is that λ is an eigenvalue for T if and only if null$(T - \lambda)$ is not trivial.

This lemma also suggests that the subspace null$(T - \lambda)$ is intimately connected to the eigenvectors for λ. As such, it makes sense to formally name this subspace.

Definition. Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$ be an eigenvalue for T. Then null$(T - \lambda)$ is called the **eigenspace** of T corresponding to λ.

Theorem 4.4. *Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T with corresponding eigenvectors $v_1, \ldots, v_m$. Then the vectors $v_1, \ldots, v_m$ are linearly independent.*

Proof. Let a_i be scalars such that

$$a_1 v_1 + \cdots + a_m v_m = 0_V.$$

Next define the new operator

$$T_j = (T - \lambda_1) \cdots (T - \lambda_{j-1})(T - \lambda_{j+1}) \cdots (T - \lambda_m).$$

As we can write the factors of T_j in any order, it follows that $T_j v_i = 0_V$, provided $i \neq j$. Next we determine the value of $T_j v_j$. To compute this, first observe that

$$(T - \lambda_k)v_j = Tv_j - \lambda_k v_j = \lambda_j v_j - \lambda_k v_j = (\lambda_j - \lambda_k)v_j.$$

Using this repeatedly, we obtain

$$\begin{aligned} T_j v_j &= (T - \lambda_1) \cdots (T - \lambda_{j-1})(T - \lambda_{j+1}) \cdots (T - \lambda_m)v_j \\ &= \underbrace{(\lambda_j - \lambda_1) \cdots (\lambda_j - \lambda_{j-1})(\lambda_j - \lambda_{j+1}) \cdots (\lambda_j - \lambda_m)}_{\Lambda} v_j. \end{aligned}$$

It now follows that

$$0_V = T_j(a_1 v_1 + \cdots + a_m v_m) = a_j \Lambda v_j.$$

As our eigenvalues are distinct, then $\Lambda \neq 0$. Dividing through by Λ now yields $0_V = a_j v_j$. Therefore $a_j = 0$, since v_j is an eigenvector and cannot be the zero vector. As j was arbitrary, this shows that all our scalars must indeed be zero, completing our proof. $\square$

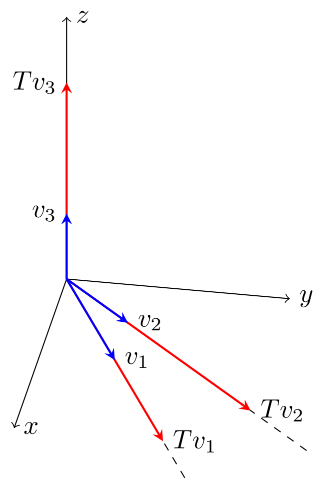
Corollary 4.5. *Any operator on a finite-dimensional vector space V has at most $\dim V$ distinct eigenvalues.*

Proof. If V has dimension n, then any set of linearly independent vectors must have at most n vectors. The result now follows immediately from the preceding theorem. $\square$

What can we say if T has exactly $\dim V$ linearly independent eigenvectors. In fact, we can say a lot! To motivate consider the operator T on $\mathbb{R}^3$ given by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \begin{bmatrix} a+2b \\ 4b-a \\ 3c \end{bmatrix}$. Straightforward calculations show that T has eigenpairs

$$\left(2, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}\right), \quad \left(3, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) \quad \text{and} \quad \left(3, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right),$$

where we denote the ith vector by v_i. In other words, T simply stretches the lines given by v_1, v_2, and v_3 as shown in the following illustration.



Since that these three eigenvectors are linearly independent, they form a basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

for $\mathbb{R}^3$. Consequently, T is completely determined by these eigenvectors. Moreover, observe that

$$[T]_{\mathcal{B}} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

As we have just seen, operators that have "enough" eigenvectors are easy to understand – they essentially just stretch space. Such operators also have a very nice description in terms of their matrix representations as described by our last theorem. We leave the details to the reader.

Definition. Let T be any operator on an $\mathbb{F}$-vector space. We say T is **diagonalizable** provided there exists a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}$ is a diagonal matrix, i.e.,

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & \text{\Large 0} & \\ & \text{\Large 0} & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

for some scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$.

Theorem 4.6. *Let $T \in \mathcal{L}(V)$. Then T is diagonalizable if and only if V has a basis consisting entirely of eigenvectors for T.*

We now turn our attention to the following question: When do operators have eigenvectors? We would hope that all operators have eigenvectors but, as the next example shows, this is unfortunately not true! To illustrate such an operator consider $T \in \mathcal{L}(\mathbb{R}^2)$ given by $T(x, y) = (-y, x)$. If T had eigenvectors, then we would be able to find values x, y, and λ such that

$$(-y, x) = T(x, y) = \lambda(x, y) \tag{$\star$}$$

Equating the first and second coordinates yields

$$-y = \lambda x \qquad \text{and} \qquad x = \lambda y.$$

Combining we obtain $y(\lambda^2 + 1) = 0$. Since $y = 0$ cannot be a solution (why?) and the equation

$$(\lambda^2 + 1) = 0 \tag{$\dagger$}$$

has no real solutions, we conclude that T has no eigenvectors.

As you might already suspect, what if we work over $\mathbb{C}$ instead of over $\mathbb{R}$? In that case, the equation ($\dagger$) has a solution, namely, $\lambda = \pm i$. If we indeed do this by considering T as an operator on $\mathbb{C}^2$ then we obtain the eigenpairs:

$$(i, (i, 1)) \qquad \text{and} \qquad (-i, (-i, 1)),$$

since

$$T(i, 1) = (-1, i) = i(i, 1) \qquad \text{and} \qquad T(-i, 1) = (-1, -i) = -i(-i, 1).$$

More generally, since polynomials always have roots over $\mathbb{C}$, then it stands to reason that any complex operator should have at least one eigenvector. Not only does the next theorem assert this but its proof relies solely on the Fundamental Theorem of Algebra!

Theorem 4.7. *Every operator T on a non-trivial finite-dimensional complex vector space V has at least one eigenvalue.*

Proof. Set $n = \dim V$ and fix any nonzero vector $u \in V$. Now consider the $n+1$ vectors

$$u, Tu, \ldots, T^n u.$$

As the dimension of V is n, these $n+1$ vectors must be dependent and so there exists scalars $a_i \in \mathbb{C}$ not all zero such that

$$\begin{aligned} 0_V &= a_0 u + a_1 T u + \cdots + a_n T^n u \\ &= (a_0 + a_1 T + \cdots + a_n T^n) u \\ &= p(T) u, \end{aligned}$$

where $p(z) = a_0 + a_1 z + \cdots + a_n z^n$. Let m be the largest index such that $a_m \neq 0$. As u is a nonzero vector we must have $0 < m \leq n$. In particular, $p(z)$ is a nonconstant polynomial. By the Fundamental Theorem of Algebra $p(z)$ factors as $p(z) = a_m(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_m)$, where $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. Plugging this into the above equation yields,

$$0_V = a_m(T - \lambda_1)(T - \lambda_2)\cdots(T - \lambda_m)u.$$

It now follows that one of the factors $(T - \lambda_i)$ must have a nontrivial null space. In other words, λ_i is an eigenvalue for T. $\square$

The above theorem only guarantees that a complex operator has at least one eigenvalue. It does not guarantee that we have lots of distinct eigenvalues, say $\dim V$ of them! To see an example of this, consider the matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ thought of as an operator $A : \mathbb{C}^2 \to \mathbb{C}^2$. To find its eigenpairs we set up the usual eigenvector/value equation

$$\begin{bmatrix} w \\ 0 \end{bmatrix} = A \begin{bmatrix} z \\ w \end{bmatrix} = \lambda \begin{bmatrix} z \\ w \end{bmatrix},$$

to obtain the equations $w = \lambda z$ and $0 = \lambda w$. We see that the only eigenpairs for A are $\left(0, \begin{bmatrix} z \\ 0 \end{bmatrix}\right)$ where $z \neq 0$. Additionally, this example demonstrates that complex operators may not even have $\dim V$ linearly independent eigenvectors. (In this case all the eigenvectors for A are multiplies of $(1, 0)$.) This means, thanks to Theorem 4.6, that not every complex-operator is diagonalizable.

That said, all is not lost! The next theorem demonstrates that, for the right choice of basis, every complex operator can be represented by an "upper-triangular" matrix. We start with a few definitions.

Definition. Let $T \in \mathcal{L}(V)$. We say an basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ for V is ***T*-triangularizing** provided that

$$Tv_k \in \operatorname{span}(v_1, \ldots, v_k),$$

for each $k \leq n$.

The next definition and lemma reveal why the term "triangularizing" is used.

Definition. We say a square matrix M is **upper triangular** provided it is of the form

$$\begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix},$$

so that the values weakly above the diagonal are arbitrary (indicated by the $*$) but the values strictly below the diagonal must be zero.

Lemma 4.8. *Let T be an operator on V with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Then the following are equivalent.*

(a) The basis $\mathcal{B}$ is T-triangularizing.

(b) For all $k \leq n$, the subspaces $\operatorname{span}(v_1, \ldots, v_k)$ *are T-invariant.*

(c) The matrix $[T]_{\mathcal{B}}$ is upper triangular.

Proof.
`Case:` (a) $\Rightarrow$ (b)

Assume $\mathcal{B}$ is T-triangularizing and fix $k \leq n$. Now for any $i \leq k$ we have

$$Tv_i \in \text{span}(v_1, \ldots, v_i) \subset \text{span}(v_1, \ldots, v_k).$$

To see that this implies T-invariance, consider an arbitrary $u \in \text{span}(v_1, \ldots, v_k)$. We know that $u = a_1 v_1 + \cdots + a_k v_k$ for some scalars a_i. Thus

$$Tu = T(a_1 v_1 + \cdots + a_k v_k) = a_1 T v_1 + \cdots + a_k T v_k \in \text{span}(v_1, \ldots, v_k),$$

where the last step follows since subspaces are closed under linear combinations. This shows $\text{span}(v_1, \ldots, v_k)$ is T-invariant.

`Case:` (b) $\Rightarrow$ (c)

Assume $\text{span}(v_1, \ldots, v_k)$ is T-invariant for each $k \leq n$. To show that

$$[T]_{\mathcal{B}} = \{[Tv_1]_{\mathcal{B}} \cdots [Tv_k]_{\mathcal{B}} \cdots [Tv_n]_{\mathcal{B}}\}$$

is upper triangular, it will suffice to show that the bottom $n-k$ rows of the kth column $[Tv_k]_{\mathcal{B}}$ are all zeros. To see this, observe that $Tv_k \in \text{span}(v_1, \ldots, v_k)$, as $\mathcal{B}$ is T-triangularizing. This means there exists scalars $a_1, \ldots, a_k$ so that

$$Tv_k = a_1 v_1 + \cdots + a_k v_k + 0 v_{k+1} + \cdots + 0 v_n.$$

Hence the coordinates of this vector and hence the kth column of our matrix is

$$[Tv_k]_{\mathcal{B}} = \begin{array}{cl} \left[\begin{array}{c} a_1 \\ \vdots \\ a_k \\ 0 \\ \vdots \\ 0 \end{array}\right] & \begin{array}{l} \\ \\ \leftarrow k\text{th row} \\ \\ \\ \\ \end{array} \end{array}.$$

As the bottom $n-k$ rows are indeed all zero, we conclude that $[T]_{\mathcal{B}}$ is upper triangular.

`Case:` (c) $\Rightarrow$ (a)

In this case we know that $[T]_{\mathcal{B}}$ is an upper triangular of the form

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ & & a_{3,3} & & \\ & & & \ddots & \vdots \\ 0 & & & & a_{n,n} \end{bmatrix}.$$

We must show that $Tv_k \in \text{span}(v_1, \ldots, v_k)$. To do this lets consider the coordinates of Tv_k (with respect to $\mathcal{B}$) first.

$$[Tv_k]_{\mathcal{B}} = [T]_{\mathcal{B}}\ [v_k]_{\mathcal{B}} = [T]_{\mathcal{B}}\ \mathbf{e}_k = \begin{bmatrix} a_{1,k} \\ \vdots \\ a_{k,k} \\ \vdots \\ 0 \end{bmatrix},$$

where the first equality follows from Theorem 3.11 and $\mathbf{e}_k$ is the kth standard vector. By definition we thus have

$$Tv_k = a_{1,k}v_k + \cdots + a_{k,k}v_k \in \text{span}(v_1, \ldots, v_k).$$

So $\mathcal{B}$ is T-triangularizing as needed. □

Theorem 4.9. *Let T be a complex operator on V. Then there exists a T-triangularizing basis $\mathcal{B}$ for V.*

Proof. We proceed by induction on $\dim V$. The result is clear when $\dim V = 1$ since basis in this case is T-triangularizing.

Now let T be an operator on an $n > 1$ dimensional space V. By Theorem 4.7, we know there exists some eigenpair (λ, v_1) for T. As v_1 is nonzero (why?) then we can extend this single vector to a basis $\{v_1, u_2, \ldots, u_n\}$ for all of V. Set $U = \text{span}(u_2, \ldots, u_n)$. Observe $\dim U = n - 1$ and $v_1 \notin U$. Next, define the function $S : U \to U$ by

$$Su = a_2u_2 + \cdots a_nu_n$$

where $Tu = a_1v_1 + a_2u_2 + \cdots + a_nu_n$. A straightforward check shows that S is indeed an operator. By induction we know that U has a S-triangularizing basis $\mathcal{B}_U = \{v_2, \ldots, v_n\}$.

We now claim that

$$\mathcal{B} = \{v_1\} \cup \{v_2, \ldots, v_n\}$$

is a T-triangularizing basis for all of V. Certainly, $\mathcal{B}$ is a basis for V as it is linearly independent (why?) and $|\,\mathcal{B}\,| = n = \dim V$. It only remains to see that $\mathcal{B}$ is T-triangularizing. When $k = 1$, we see that $Tv_1 = \lambda v_1 \in \text{span}(v_1)$, since (λ, v_1) is an eigenpair for T.

For $k \geq 2$ we observe that as $\mathcal{B}_U$ is S-triangularizing, then

$$Sv_k \in \text{span}(v_2, \ldots, v_k)$$

where $Tv_k = a_1v_1 + Sv_k$. So, $Tv_k \in \text{span}(v_1, \ldots, v_k)$. This concludes our proof. □

4.3 Direct sums

To motivate this section consider the vector space $\mathbb{R}^2$ thought of as the xy-plane. From this perspective the two subspaces

$$X = \{(x, 0) \mid x \in \mathbb{R}\} \qquad \text{and} \qquad Y = \{(0, y) \mid y \in \mathbb{R}\}$$

of $\mathbb{R}^2$ have a more familar description – they are just the x- and y-axes. The fact that these axes only intersect at the origin is reflected in the fact that

$$X \cap Y = \{(0, 0)\}.$$

We also know that every vector in $\mathbb{R}^2$ can be written (uniquely) as vector in the x-axis plus a vector in the y-axis. This is mirrored by the fact that

$$\mathbb{R}^2 = \{u + v \mid u \in X, v \in Y\}.$$

It now makes sense to denote the set on the right by $X + Y$, in which case we write $\mathbb{R}^2 = X + Y$. That is $\mathbb{R}^2$ is the "sum" of two of its subspaces. We can perform a similar decomposition to $\mathbb{R}^3$ as well. Redefining X, Y and Z to be the subspaces given by the x-, y-, and z-axis, then we certainly have

$$\mathbb{R}^3 = X + Y + Z.$$

Our first definition formalizes these ideas.

Definition. Let $U_1, \ldots, U_m$ be subspaces of V. We define their **sum** to be the set

$$U = U_1 + \cdots + U_m = \{u_1 + \cdots + u_m \mid u_i \in U_i\}.$$

We say U is the **direct sum** of the U_i provided that the only choice of $u_i \in U_i$ that gives

$$u_1 + \cdots + u_m = 0_V,$$

is $u_i = 0_V$. In this case we write,

$$U = U_1 \oplus \cdots \oplus U_m.$$

A straightforward check shows that U is a subspace of V. As it is clear that U_i is a subset of U (check!), it now follows that U_i is a subspace of U. There is also an important analogy that is not to be missed. The subspaces U_i are behaving a lot like vectors. Saying that $U = U_1 + \cdots + U_m$ is reminiscent of a spanning set of vectors, while the definition of a direct sum is reminiscent of linear independence. The next few lemmas make this analogy stronger. We leave the proof of the first lemma to the reader.

Lemma 4.10. *Let $U_1, \ldots, U_m$ be subspaces of V and consider the space*

$$U = U_1 + \cdots + U_m.$$

Then, U is a direct sum of the U_i's if and only if for every $u \in U$ there is a unique choice of $u_i \in U_i$ so that vector $u = u_1 + \cdots + u_m$.

```
make this an iff, so that we can create direct sums by
partitioning a basis.
```

ToDo

Lemma 4.11. *Let $U_1, \dots, U_m$ be subspaces of V such that*

$$V = U_1 \oplus \cdots \oplus U_m.$$

If $\mathcal{B}_i$ is a basis for each U_i, then

$$\mathcal{B} = \mathcal{B}_1 \bigcup \cdots \bigcup \mathcal{B}_m$$

is a basis for V. Consequently, $\dim V = \dim U_1 + \cdots + \dim U_m$.

Proof. To simplify notation we prove the claim when $m = 2$. As the reader should check, the same proof works in general. Let $U = U_1$ and $W = U_2$ and fix bases $\mathcal{B}_U = \{u_1, \dots, u_k\}$ and $\mathcal{B}_W = \{w_1, \dots, w_m\}$ for U and W respectively. A simple computation shows that $\mathcal{B}$ spans all of V:

$$\begin{aligned}
\operatorname{span}(\mathcal{B}) &= \{(a_1u_1 + \cdots + a_ku_k) + (b_1w_1 + \cdots + b_mw_m) \mid a_i, b_i \in \mathbb{F}\} \\
&= \{u + w \mid u \in U, w \in W\} \\
&= U + W \\
&= V
\end{aligned}$$

where the second equality follows since $\mathcal{B}_U$ spans U and $\mathcal{B}_W$ spans W. To see that $\mathcal{B}$ is linearly independent, assume a_i and b_i are such that

$$(a_1u_1 + \cdots + a_ku_k) + (b_1w_1 + \cdots + b_mw_m) = 0_V.$$

By definition of our sum being direct, we must have

$$a_1u_1 + \cdots + a_ku_k = 0_V \qquad \text{and} \qquad b_1w_1 + \cdots + b_mw_m = 0_V.$$

By the independence of the u_i's and the w_i's we see that all the a_i's and all the b_i's are zero. □

Continuing our analogy between direct sums and bases, the next lemma is a generalization of the Basis Extension Theorem. In fact the main ingredient in its proof is exactly this theorem.

```
FIX the wording of this lemma
```

ToDo

Lemma 4.12. *Fix vector space V and subspace U. Let $S = \{v_1, \dots, v_k\}$ be a set of vectors such that $U \cap \operatorname{span}(S) = \{0_V\}$. Then there exists a subspace W such that*

$$V = U \oplus W$$

and $S \subset W$.

Proof. Let $W' = \operatorname{span}(S)$. If $V = U \oplus W'$ we are done. Otherwise, let $\mathcal{B}_U$ be a basis for U and $\mathcal{B}_{W'}$ be one for W'. By the previous lemma $\mathcal{B}_U \cup \mathcal{B}_{W'}$ is a basis for $U \oplus W'$. Now extend $\mathcal{B}_U \cup \mathcal{B}_{W'}$ to a basis for all of V by adding vectors $v_1, \ldots, v_\ell$. Then

$$W = \operatorname{span}(\mathcal{B}_{W'} \cup \{v_1, \ldots, v_\ell\})$$

is the desired space. $\square$

A nice consequence of this lemma is that if U is a subspace of V, so that $V = U \oplus W$. Simply choose $S = \{0_V\}$ in the lemma!

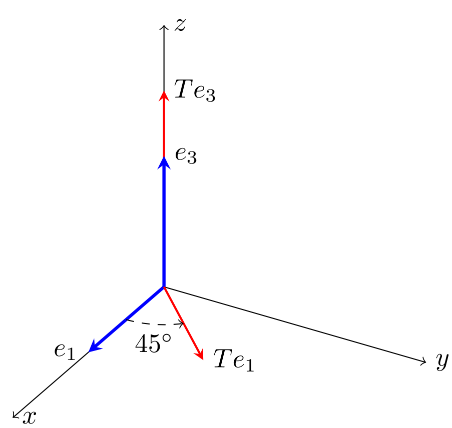
As we have already seen, direct sums decompose our vector space into (smaller) vector spaces. Interestingly, we can also use direct sums to decompose operators as well. The one requirement is that the operator must behave "nicely" with respect to the subspaces involved. Our next definition makes this precise.

Definition. Let $T \in \mathcal{L}(V)$. We say a subspace U of V if T**-invariant** provided that $T(U) = \{Tu \mid u \in U\} \subseteq U$.

To motive what is to come, let us consider an example. Let $T \in \mathcal{L}(\mathbb{R}^3)$ be given by

$$T(x, y, z) = \left(\frac{x-y}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}, 3z\right).$$

Observe that T decomposes into two parts. The first is a rotation by 45° in the xy-plane and the second is scaling by 3 along the z-axis. To illustrate consider the following picture.



Now define the following subspaces of $\mathbb{R}^3$

$$U = \{(x, y, 0) \mid x, y \in \mathbb{R}\} \qquad \text{and} \qquad W = \{(0, 0, z) \mid z \in \mathbb{R}\}.$$

It is clear from our description of T that U and W are both T-invariant subspaces. Moreover, it is clear that $\mathbb{R}^3 = U \oplus W$. Now consider $[T]_{\mathcal{B}}$, where

$\mathcal{B} = \{e_1, e_2\} \cup \{e_3\}$. That is $\mathcal{B}$ is the union of a basis for U and a basis for W. Computing we see that

$$[T]_{\mathcal{B}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Looking closely, we see this matrix consists of a 2×2 block (encodes the rotation) and 1×1 block (encodes the scaling) along its diagonal and all the other entries are zero.

The significance of all this, is that if we can decompose our vector space into the direct sum of T-invariant subspace, then this means that our operator also decomposes into operators on smaller spaces as seen in this example. We consider this idea in complete generality. The reader is encouraged to keep our specific example in mind when reading the general ideas. We begin with an expected definition.

Definition. We say a square matrix M is **block diagonal** provided it is of the form

$$M = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{bmatrix},$$

where each of the A_i are square matrices lying along the diagonal of M and all other entries in M are zero. In the case each block is just a 1×1 matrix, we say M is **diagonal**.

As suggested by our motivating example, we now have the following lemma.

Lemma 4.13. *Assume $U_1, \ldots, U_m$ are T-invariant subspaces of V such that*

$$V = U_1 \oplus \cdots \oplus U_m.$$

If $\mathcal{B}_i$ is a basis for U_i, then

$$[T]_{\mathcal{B}} = \begin{bmatrix} [T|_{U_1}]_{\mathcal{B}_1} & & & 0 \\ & [T|_{U_2}]_{\mathcal{B}_2} & & \\ & & \ddots & \\ 0 & & & [T|_{U_m}]_{\mathcal{B}_m} \end{bmatrix},$$

where $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_m$. Conversely, if $\mathcal{B}$ is any basis for V such that $[T]_{\mathcal{B}}$ is block diagonal, then V is the direct sum of T-invariant subspaces.

The reader should pause to convince themselves that this lemma is a direct generalization of Theorem 4.6. In fact, it further strengthens our analogy between linear independent vectors and direct sums. This should be somewhat expected since the 1-dimensional space spanned by an eigenvector is just an invariant subspace. The proof of this lemma is left to the reader.

4.4 Generalized eigenvectors

In Section 4.2, we saw that the (real) operator that rotates $\mathbb{R}^2$ by $90°$ fails to have a single eigenvector. From this we concluded that, in general, not all real operators decompose into their eigenspaces, i.e., they are not all diagonalizable. In the same section, we also showed that the complex operator $A : \mathbb{C}^3 \to \mathbb{C}^3$, where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

only has eigenspaces

$$\mathrm{null}(A-1) = \mathrm{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) \qquad \text{and} \qquad \mathrm{null}(A-3) = \mathrm{span}\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right).$$

Consequently, it follows that not all complex operators are diagonalizable either. Since the final goal of this chapter is to decompose (complex) operators into invariant subspaces, these examples demonstrate that eigenspaces are the wrong invariant subspaces to consider. The question now is: Which are the right invariant subspaces? To help answer this question, let us revisit the operator $A : \mathbb{C}^3 \to \mathbb{C}^3$ from above. In particular, observe that

$$\mathrm{null}(A-1)^2 = \mathrm{null} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}^2 = \mathrm{null} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

It now follows that

$$\mathrm{null}(A-1)^2 = \mathrm{span}(\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right),$$

and moreover, that

$$\mathrm{null}(A-1)^2 \oplus \mathrm{null}(A-3) = \mathbb{C}^3.$$

This example suggests the following definition.

Definition. Let λ be an eigenvalue of T and define the **generalized eigenspace** (corresponding to λ) to be the subspace

$$\mathcal{G}_\lambda = \mathrm{null}(T-\lambda)^{\dim V}.$$

We call $0_V \neq v \in \mathcal{G}_\lambda$ a **generalized eigenvector** (corresponding to λ).

Just like eigenspaces, we next show that generalized eigenspaces are also invariant subspaces.

Lemma 4.14. *Let λ be an eigenvalue of T. Then $\mathcal{G}_\lambda$ is a T-invariant.*

Proof. Fix $n = \dim V$ and let $u \in \mathcal{G}_\lambda = \operatorname{null}(T-\lambda)^n$. To prove invariance, we must show that $Tu \in \operatorname{null}(T-\lambda)^n$. Computing, we see that

$$(T-\lambda)^n Tu = T(T-\lambda)^n u = T0_V = 0_V,$$

where the first equality follows as T commutes with itself. Hence $Tu \in \operatorname{null}(T-\lambda)^n$ as needed. □

As suggested by the above example, our goal now is to prove the following structure theorem for complex operators.

Theorem 4.15. *Let T be an operator on a finite-dimensional complex vector space V. If the distinct eigenvalues of T are $\lambda_1, \dots, \lambda_m$, then*

$$V = \mathcal{G}_{\lambda_1} \oplus \cdots \oplus \mathcal{G}_{\lambda_m}.$$

In order to prove this theorem, we will need to establish a few lemmas and a deeper understanding of generalized eigenvectors. We do this first, postponing the proof of this theorem to the end of the section. We begin with a technical lemma.

Lemma 4.16. *Assume U is a T-invariant subspace of V with dimension m. Then*

$$T^m(U) = T^{m+1}(U) = T^{m+2}(U) = \cdots.$$

Proof. First observe that $T^i(U) \supseteq T^{i+1}(U)$. To see this note that

$$T^{i+1}u = T^i(Tu) \in T^i(U),$$

since $Tu \in U$ as U is T-invariant. This gives us the following sequence of inclusions

$$U \supseteq T(U) \supseteq T^2(U) \supseteq \cdots.$$

Observe we must have some i such that $T^i(U) = T^{i+1}(U)$. If not then

$$m = \dim U > \dim T(U) > \dim T^2(U) > \cdots$$

would be an infinite decreasing sequence of positive integers! Therefore there must exist a smallest $i \leq m$ so that $T^i(U) = T^{i+1}(U)$. (Why must $i \leq m$?) Now observe that

$$T^{i+1}(U) = TT^i(U) = TT^{i+1}(U) = T^{i+2}(U).$$

Doing this again we get

$$T^{i+2}(U) = TT^{i+1}(U) = TT^{i+2}(U) = T^{i+3}(U).$$

Repeating this argument indefinitely yields

$$T^i(U) = T^{i+1}(U) = T^{i+2}(U) = \cdots$$

and completes the proof. □

We now turn our attention to exploring an important connection between T-triangularizing bases and generalized eigenspaces. To warm-up, let T be a complex operator on V with T-triangularizing basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ so that

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & & \ldots & a_{1,k} & & * \\ & \lambda_2 & \ldots & a_{2,k} & & \\ & & \ddots & \vdots & & \\ & & & \lambda_k & & \\ & & & & \ddots & \\ 0 & & & & & \lambda_n \end{bmatrix}.$$

As suggested by our choice of notation, we claim that every entry on the diagonal is actually an eigenvalue. To see this it will suffice to show that $\operatorname{null}(T - \lambda_k)$ is **not** trivial. This follows almost immediately. As $\mathcal{B}$ is T-triangularizing, we know that

$$(T - \lambda_k)v_i \in \operatorname{span}(v_1, \ldots, v_i) \subset \operatorname{span}(v_1, \ldots, v_{k-1})$$

for all $i < k$. Additionally,

$$(T - \lambda_k)v_k = (a_{1,k}v_1 + \cdots a_{k-1,k}v_{k-1} + \lambda_k v_k) - \lambda_k v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

This means $(T - \lambda_k)$ is a map from the k-dimensional space $\operatorname{span}(v_1, \ldots, v_k)$ to the $k-1$-dimensional space $\operatorname{span}(v_1, \ldots, v_{k-1})$. Consequently, it must have a nontrivial nullspace. (Why?) We now see that all the values on the diagonal of $[T]_{\mathcal{B}}$ are eigenvalues for T.

In fact, the next lemma states an even deeper connection between eigenvalues and triangularizing bases.

Lemma 4.17. *Fix a complex operator T on the n-dimensional vector space V and let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be any T-triangularizing basis for V. Then every scalar on the diagonal of $[T]_{\mathcal{B}}$ is an eigenvalue for T.*

Moreover, if $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues for T, then λ_i appears on the diagonal of $[T]_{\mathcal{B}}$ precisely $\dim \mathcal{G}_{\lambda_i}$ times. Consequently,

$$\dim V = \dim \mathcal{G}_{\lambda_1} + \cdots + \dim \mathcal{G}_{\lambda_m}.$$

Proof. As we have already proved the first claim, we concentrate on the second. To this end, it suffices to prove the following special case:

$$\boxed{\text{If } S \text{ is an arbitrary operator on } V \text{ with } S\text{-triangularizing basis } \mathcal{B}, \text{ then zero appears on the diagonal of } [S]_{\mathcal{B}} \text{ exactly } \dim \operatorname{null} S^n \text{ times.}}$$

The general result follows from this special case as follows. Let λ be an eigenvalueIf

$$d = \dim \operatorname{null}(\underbrace{T - \lambda_i}_{S})^n = \dim \mathcal{G}_{\lambda_i},$$

then by the special case d is the number of times zero appears on the diagonal of

$$[S]_{\mathcal{B}} = [T - \lambda_i]_{\mathcal{B}} = [T]_{\mathcal{B}} - [\lambda_i]_{\mathcal{B}} = [T]_{\mathcal{B}} - \begin{bmatrix} \lambda_i & & & 0 \\ & \lambda_i & & \\ & & \ddots & \\ 0 & & & \lambda_i \end{bmatrix}$$

which, we now see, is the same as the number of times λ_i appears on the diagonal of $[T]_{\mathcal{B}}$. This proves the general result.

We now concentrate on proving the special case. For concreteness, set

$$[S]_{\mathcal{B}} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ & & a_{3,3} & & \\ & & & \ddots & \vdots \\ 0 & & & & a_{n,n} \end{bmatrix},$$

and let d be the number of zero entries along the diagonal of $[S]_{\mathcal{B}}$. The number of nonzero entries along this diagonal is then $n - d$. The proof will certainly be complete if we can establish that

$$\dim \operatorname{ran} S^n = \dim S^n(V) = n - d,$$

since an application of Rank-Nullity shows that

$$\begin{aligned} \dim \operatorname{null} S^n &= \dim V - \dim \operatorname{ran} S^n \\ &= n - (n - d) \\ &= d, \text{ the number of zeros on the diagonal of } [S]_{\mathcal{B}}, \end{aligned}$$

as needed. To this end set

$$V_i = \operatorname{span}(v_1, \ldots, v_i),$$

for $0 < i$, and $V_0 = \{0_V\}$. As $V_i \subseteq V_{i+1}$, for $0 \leq i$, we obtain the following sequence of inclusions

$$\{0_V\} \subseteq S^n(V_1) \subseteq S^n(V_2) \subseteq \cdots \subseteq S^n(V_n) = S^n(V). \qquad (\star)$$

Instead of proving directly that $\dim S^n(V) = n - d$, it will be easier to show that

$$\dim S^n(V) = \# \text{ of strict inclusions in } (\star) = n - d.$$

At this point, we have reduced the problem to establishing these two equalities. The first equality will be proved by establishing the claim that

$$\dim S^n V_i + 1 = \dim S^n V_{i+1} \qquad (1)$$

provided $S^n V_i \subsetneq S^n V_{i+1}$. Likewise, the second equality will follow from the claim that

$$S^n V_i \subsetneq S^n V_{i+1} \text{ if and only if } a_{i+1,i+1} \neq 0. \qquad (2)$$

We now turn our attention to proving claims (1) and (2). To prove claim (1), observe that since

$$S^n(V_i) = \operatorname{span}(S^n v_1, \ldots, S^n v_i),$$

it follows that the spanning sets, for any two consecutive spaces in ($\star$), differ by exactly one vector. This proves claim (1).

To prove claim (2), let us first assume that $a_{i+1,i+1} = 0$. As we already know that $S^n(V_i) \subseteq S^n(V_{i+1})$, we need only establish the reverse inclusion. To this end observe that since $a_{i+1,i+1} = 0$, our matrix representation of S implies that

$$Sv_{i+1} = a_{1,i}v_1 + \cdots + a_{i,i}v_i \in V_i.$$

Additionally, as $\mathcal{B}$ is S-triangularizing, we also have $Sv_1, \ldots, Sv_i \in V_i$. This implies that

$$S(V_{i+1}) = \operatorname{span}(Sv_1, \ldots, Sv_{i+1}) \subseteq V_i.$$

Now consider,

$$S^n(V_{i+1}) = S^{n-1}S(V_{i+1}) \subseteq S^{n-1}(V_i) = S^n(V_i),$$

where the last equality follows from Lemma 4.16 since $\dim V_i = i < n$. It only remains to consider the case when $a = a_{i+1,i+1} \neq 0$. In this case, we see that

$$S^n v_{i+1} = b_1 v_1 + \cdots + b_i v_i + a^n v_{i+1},$$

for some scalars $b_1, \ldots, b_i$. As $a \neq 0$ and the v_j are linearly independent, we see that $S^n v_{i+1} \notin \operatorname{span}(v_1, \ldots, v_i) = V_i$. Consequently, $S^n v_{i+1} \notin S^n(V_i) \subseteq V_i$, where the inclusion follows as $\mathcal{B}$ is S-triangularizing. So, $S^n V_i \subsetneq S^n V_{i+1}$, which completes our proof. □

`fix working of statement; awk`

ToDo

Lemma 4.18. *Let T be a complex operator whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_m$. Assume $u_i \in \mathcal{G}_{\lambda_i}$ are such that*

$$0_V = u_1 + \cdots + u_m,$$

each $u_i = 0_V$. Consequently, if W is the sum of the subspaces $\mathcal{G}_{\lambda_i}$, then

$$W = \mathcal{G}_{\lambda_1} \oplus \cdots \oplus \mathcal{G}_{\lambda_m}.$$

Proof. For brevity set $n = \dim V$ and $\mathcal{G}_i = \mathcal{G}_{\lambda_i}$. Assume for a contradiction that there exists $u_i \in \mathcal{G}_i$, not all zero, so that

$$0_V = u_1 + \cdots + u_m. \tag{$\star$}$$

Without loss of generality let us assume $u_1 \neq 0_V$. We now argue that we can further assume that u_1 is an eigenvector. To show this observe that since $(T - \lambda_1)^n u_1 = 0_V$, then there must exists a largest $k \geq 0$ so that

$$(T - \lambda_1)^k u_1 \neq 0_V \qquad \text{and} \qquad (T - \lambda_1)^{k+1} u_1 = 0_V$$

This means the $(T-\lambda_1)^k u_1$ is an eigenvector for T. As each of the $\mathcal{G}_i$ are T-invariant, and hence $(T-\lambda_1)^k$-invariant, applying the operator $(T-\lambda_1)^k$ to both sides of $(\star)$, and relabeling the ith vector as u_i, allows us to assume u_1 is an eigenvector.

We now model the remainder of our proof on the proof in Theorem 4.4. As such define the operators

$$S = (T-\lambda_2)^n \cdots (T-\lambda_m)^n$$

As in the proof of Theorem 4.4 we have

$$Su_i = \begin{cases} 0_V & \text{if } i \neq 1 \\ \Lambda u_1 \neq 0_V & \text{if } i = 1 \end{cases}$$

for some $\Lambda \neq 0$ in $\mathbb{C}$. Multiplying $(\star)$ through by S now reveals that $0_V = \Lambda u_1$. This is a contradiction since $\Lambda \neq 0$ and $u_1 \neq 0_V$.

$\square$

Proof of Theorem 4.15. It follows from the previous lemma that

$$\mathcal{G}_{\lambda_1} \oplus \cdots \oplus \mathcal{G}_{\lambda_m} \subseteq V.$$

To complete the proof we must show this is actually an equality. The easiest way to do this is to show the dimension of the direct sum is $n = \dim V$. Now consider

$$\dim(\mathcal{G}_{\lambda_1} \oplus \cdots \oplus \mathcal{G}_{\lambda_m}) = \sum_{i=1}^{m} \dim(\mathcal{G}_{\lambda_i}) = \dim V,$$

where the second equality follows directly from Lemma 4.17.

$\square$

4.5 The characteristic polynomial

Throughout this section let T be a complex operator on V with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Additionally set $d_i = \dim \mathcal{G}_{\lambda_i}$. We then have the following definition.

Definition. The **characteristic polynomial** of a complex operator T is the polynomial

$$\rho_T(x) = (x-\lambda_1)^{d_1} \cdots (x-\lambda_m)^{d_m}.$$

For example, consider the operator $A : \mathbb{C}^3 \to \mathbb{C}^3$, where

$$A = \begin{bmatrix} 1 & 5 & 5 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{bmatrix}.$$

Then by Lemma 4.17, it follows that

$$\rho_A(x) = (x-1)^2(x-3).$$

Now let us compute the operator $\rho_A(A)$. We see that

$$\rho_A(A) = (A-1)^2(A-3) = \begin{bmatrix} 0 & 5 & 5 \\ 0 & 0 & 5 \\ 0 & 0 & 2 \end{bmatrix}^2 \begin{bmatrix} -2 & 5 & 5 \\ 0 & -2 & 5 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Our next (famous) theorem states that this always occurs.

Theorem 4.19 (Cayley-Hamilton)**.** *Let T be any complex operator, then $\rho_T(T) = 0$.*

Proof. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a T-triangularizing basis for V so that

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & & \ldots & b_1 & & * \\ & \ell_2 & \ldots & b_2 & & \\ & & \ddots & \vdots & & \\ & & & \ell_k & & \\ & & & & \ddots & \\ 0 & & & & & \ell_n \end{bmatrix}.$$

Now set $V_i = \operatorname{span}(v_1, \ldots, v_i)$. First observe that $Tv_1 = \ell_1 v_1$ or equivalently,

$$(T - \ell_1) : V_1 \to \{0_V\}.$$

Next we claim that for any $k > 1$ we have

$$(T - \ell_k) : V_k \to V_{k-1}.$$

If $i < k$, then, as $\mathcal{B}$ is T-triangularizing, we have

$$(T - \ell_k)v_i = Tv_i - \ell_k v_i \in V_k.$$

Additionally, we have

$$(T - \ell_k)v_k = (b_1 v_1 + \cdots + b_{k-1}v_{k-1} + \ell_k v_k) - \ell_k v_k \in V_{k-1}.$$

We now have the following sequence of maps

$$V = V_n \xrightarrow{T-\ell_n} V_{n-1} \xrightarrow{T-\ell_{n-1}} \quad \cdots \quad \xrightarrow{T-\ell_2} V_1 \xrightarrow{T-\ell_1} \{0_V\}.$$

For any $u \in V$ we now get

$$0 = (T - \ell_1) \cdots (T - \ell_n)u = (T - \lambda_1)^{d_1} \cdots (T - \lambda_m)^{d_m} u = \rho_T(T)u,$$

where the second equality follows from Lemma 4.17. As $u \in V$ is arbitrary, $\rho_T(T)$ has to be the zero map as claimed. $\square$

4.6 Jordan basis theorem

Definition. We say an operator N on V is **nilpotent** if there exists some positive integer k such that $N^k = 0$.

Examples.

1. The zero operator is nilpotent.

2. The operator $N : \mathbb{C}^3 \to \mathbb{C}^3$ given by $N(a, b, c) = (0, a, b)$ is nilpotent since $N^3 = 0$.

3. The matrix operator $A : \mathbb{C}^3 \to \mathbb{C}^3$ where
$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$
is nilpotent as A^3 is the zero matrix. Do the nilpotent operators A and N look similar? They should since they are the same! In fact, you should convince yourself that
$$[N]_{\{e_1, e_2, e_3\}} = A,$$
so that A is just the matrix representation of N with respect to the standard basis for $\mathbb{C}^3$. A good way to think about N is that it forces the vectors e_1, e_2, e_3 to "walk the plank":
$$\begin{tikzcd} e_1 \arrow[r, bend left, "N"] & e_2 \arrow[r, bend left, "N"] & e_3 \arrow[r, bend left, "N"] & 0_{\mathbb{C}^3} \end{tikzcd}$$

4. The operator $M : \mathbb{C}^5 \to \mathbb{C}^5$, given by
$$M(a, b, c, d, e) = (0, a, 0, c, d)$$
is nilpotent as the reader should check that $M^3 = 0$. In this case M forces our standard vectors to walk two planks:
$$\begin{tikzcd} e_1 \arrow[r, bend left, "M"] & e_2 \arrow[r, bend left, "M"] & 0_{\mathbb{C}^3} \end{tikzcd}$$
$$\begin{tikzcd} e_3 \arrow[r, bend left, "M"] & e_4 \arrow[r, bend left, "M"] & e_5 \arrow[r, bend left, "M"] & 0_{\mathbb{C}^3} \end{tikzcd}$$
From this we see that $\mathrm{span}(e_1, e_2)$ and $\mathrm{span}(e_3, e_4, e_5)$ are M-invariant spaces such that $\mathbb{C}^5 = \mathrm{span}(e_1, e_2) \oplus \mathrm{span}(e_3, e_4, e_5)$. Consequently, Lemma 4.13 gives us
$$[M]_{\{e_1, e_2\} \cup \{e_3, e_4, e_5\}} = \left[\begin{array}{c|c} \begin{matrix} 0 & 0 \\ 1 & 0 \end{matrix} & \Large 0 \\ \hline \Large 0 & \begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{matrix} \end{array}\right].$$

5. Let T be any operator with eigenvalue λ. Then $(T - \lambda)$ restricted to the generalized eigenspace corresponding to λ is nilpotent.

Definition. Let N be a nilpotent operator on V. We say a basis $\mathcal{J}$ is a **Jordan basis** with respect to N, provided there exists vectors $v_1, \ldots, v_s \in \mathcal{J}$ such that

$$\mathcal{J} = \{v_1, Nv_1, \ldots, N^{k_1}v_1\} \cup \cdots \cup \{v_s, Nv_s, \ldots, N^{k_s}v_s\}$$

and $N(N^{k_i}v_i) = 0_V$. We say the vectors v_i are **$\mathcal{J}$-generating** vectors.

Definition. Let $k \geq 0$. We define a **Jordan block** of size $k+1$ to be the $(k+1) \times (k+1)$ matrix

$$J(k) = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & 1 & \ddots & \\ & & & \\ & & 1 & 0 \end{bmatrix},$$

so that $J(k)$ has $k+1$ zeros down the diagonal and k ones on the off diagonal.

As a warm-up for the proof of our next lemma, assume that N is a nilpotent operator on V and that $\mathcal{J} = \{v, Nv, \ldots, N^k v\}$ is a basis for V where $N^{k+1}v = 0_V$. Before continuing, the reader should convince themselves that

$$[N]_{\mathcal{J}} = J(k).$$

Lemma 4.20. *Let N be a nilpotent operator on V and assume $\mathcal{J}$ is a Jordan basis for V with respect to N. Then*

$$[N]_{\mathcal{J}} = \begin{bmatrix} J(k_1) & & & 0 \\ & J(k_2) & & \\ & & \ddots & \\ 0 & & & J(k_s) \end{bmatrix},$$

where $k_1, \ldots, k_s$ are as defined in our definition of a Jordan basis.

Proof. As

$$\mathcal{J} = \{v_1, Nv_1, \ldots, N^{k_1}v_1\} \cup \cdots \cup \{v_s, Nv_s, \ldots, N^{k_s}v_s\}$$

we see that

$$V = \operatorname{span}(v_1, Nv_1, \ldots, N^{k_1}v_1) \oplus \cdots \oplus \operatorname{span}(v_s, Nv_s, \ldots, N^{k_s}v_s).$$

Certainly the ith summand is N-invariant. Moreover, denoting the ith summand U_i and its basis $\mathcal{B}_i = \{v_i, Nv_i, \ldots, N^{k_i}v_i\}$, we see from our warm-up that

$$\big[N|_{U_i}\big]_{\mathcal{B}_i} = J(k_i).$$

An application of Lemma 4.13 now yields the final result. □

Theorem 4.21. *Let V be a nontrivial finite-dimensional vector space and let N be a nilpotent operator on V. Then there exists a Jordan basis for V with respect to N.*

Proof. Begin by setting $V_0 = \{0_V\}$ and defining $V_{i+1} = \{v \in V \mid Nv \in V_i\}$. A straightforward argument shows that $V_i = \operatorname{null} N^i$ and

$$V_0 \subset V_1 \subset \cdots \subset V_m = V,$$

for some m. In fact, as V is nontrivial we must have $m \geq 1$.

We now aim to prove the following, slightly stronger, result. Given any set S of linearly independent vectors such that

$$\operatorname{span}(S) \cap V_{m-1} = \{0_V\}, \qquad (\star)$$

there exists a Jordan basis $\mathcal{J}$ for V such that S is a set of (not necessarily all) $\mathcal{J}$-generating vectors. We proceed by induction on m. If $m = 1$, then $V = \operatorname{null} N$, and the result follows by the Basis Extension Theorem. Now assume $m > 1$ and let $S = \{v_1, \ldots, v_k\}$ be such that $(\star)$ holds. By Lemma 4.12 we see that there exists a subset W such that

$$V = V_{m-1} \oplus W$$

and $S \subset W$. Extending S, if necessary, we may assume it is a basis for W.

Now observe that $S' = \{Nv_1, \ldots, Nv_k\} \subset V_{m-1}$. We claim that $(\star)$ holds with S' in place of S and V_{m-2} in place of V_{m-1}. For the moment, let us assume this to be true. Applying our inductive hypothesis to the vector space V_{m-1} and the set S' then gives us a Jordan basis $\mathcal{J}'$ for V_{m-1} such that S' is a set of $\mathcal{J}'$-generating vectors. It now follows that $\mathcal{J} = \mathcal{J}' \cup S$, is a Jordan basis for all of V and S is a set of $\mathcal{J}$-generating vectors. In light of this, it only remains to prove our claim.

First we show $\operatorname{span}(S') \cap V_{m-2}$ is trivial. To this end, assume a_i are such that

$$a_1 Nv_1 + \cdots + a_k Nv_k \in V_{m-2}.$$

By linearity, we see that $N(a_1v_1 + \cdots + a_kv_k) \in V_{m-2}$. Therefore $a_1v_1 + \cdots + a_kv_k \in V_{m-1} \cap W = \{0_V\}$. As the v_i are independent, we must have all $a_i = 0$. We conclude $\operatorname{span}(S') \cap V_{m-2}$ is trivial as needed.

A similar argument shows that S' is independent. If the a_i are now such that

$$0_V = a_1Nv_1 + \cdots a_kNv_k = N(a_1v_1 + \cdots a_kv_k),$$

then we see that

$$a_1 v_1 + \cdots + a_k v_k \in W \ \cap \ \operatorname{null} N \subset W \ \cap \ V_{m-1} = \{0_V\}.$$

As the v_i are independent, then all the a_i's are zero. This shows S' is a independent. $\square$

Corollary 4.22. *(Jordan Form) Let T be a complex operator on an n-dimensional space V with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Then there exists a basis $\mathcal{B}$ for V so that*

$$[T]_{\mathcal{B}} = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{bmatrix},$$

and A_i is some $\dim \mathcal{G}_i \times \dim \mathcal{G}_i$ block diagonal matrix whose blocks are of the form

$$\begin{bmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & 1 & \ddots & \\ & & 1 & \lambda_i \end{bmatrix}.$$

Proof. By Theorem 4.15 we know that

$$V = \mathcal{G}_{\lambda_i} \oplus \cdots \oplus \mathcal{G}_{\lambda_m}.$$

As each $\mathcal{G}_i$ is T invariant, it will suffice to show, thanks to Lemma 4.13, that there exists a basis $\mathcal{B}_i$ for $\mathcal{G}_i$ so that A_i may be taken to be

$$\left[T|_{\mathcal{G}_i}\right]_{\mathcal{B}_i}.$$

From Example 5 above, we know that $T - \lambda_i$ is nilpotent when restricted to $\mathcal{G}_i$. Theorem 4.21 now provides us with a Jordan basis $\mathcal{B}_i$ for $\mathcal{G}_i$ with respect to $(T - \lambda_i)$. This indeed finishes the proof since Lemma 4.20 tells us that the matrix

$$\left[T|_{\mathcal{G}_i}\right]_{\mathcal{B}_i} = \left[(T - \lambda_i)|_{\mathcal{G}_i}\right]_{\mathcal{B}_i} + \left[\lambda_i\right]_{\mathcal{B}_i}$$

has the desired form. $\square$