Axler Notes

Chapters 1–6

Elvin Lo

Summer–Fall 2023

2024.6.8

Preface

These notes follow the third edition of Axler, *Linear Algebra Done Right*. I worked through the first four chapters of Axler over Summer 2023, intending to dip my feet in proof-based mathematics and linear algebra before beginning college. The remaining sections I skimmed throughout Fall 2023 to accompany Math 55a lecture.

Below are solutions to exercises:

- https://linearalgebras.com;
- https://github.com/celiopassos/linear-algebra-done-right-solutions;
- https://github.com/jubnoske08/linear_algebra (odd exercises only).

And below are some other resources that may be helpful to accompany Axler:

- MIT's 18.700 "Linear Algebra" from Fall 2016, with problem sets and exams roughly corresponding to every few chapters of Axler;
- Axler's corresponding YouTube playlist, with videos explaining each subsection of the text;
- Homework assignments for Brown's Mathematics 540 "Honors Linear Algebra" from Spring 2020, including selected exercises from Axler.

Contents

1 Vector Spaces

1.A $\mathbb{R}^n$ and $\mathbb{C}^n$

Remark 1.1.
Throughout Axler, $\mathbb{F}$ stands for either $\mathbb{R}$ or $\mathbb{C}$. Thus if we prove a theorem involving $\mathbb{F}$, it will hold when $\mathbb{F}$ is replaced by $\mathbb{R}$ and when $\mathbb{F}$ is replaced by $\mathbb{C}$. Further, many of the definitions, theorems, and proofs in linear algebra that will work for $\mathbb{R}$ and $\mathbb{C}$ also will work for arbitrary fields.

1.B Definition of vector space

Definition 1.2. Vector space
A vector space is a set V along with an addition on V (assigning an element $u + v \in V$ for every pair $u, v \in V$) and a scalar multiplication on V (assigning an element $\lambda v \in V$ to each $\lambda \in \mathbb{F}$ and each $v \in V$) s.t. the following properties hold:

- Commutativity of addition
- Associativity of addition and scalar multiplication
- Additive identity: There exists an element $0 \in V$ s.t. $v + 0 = v$ for all $v \in V$
- Additive inverse: For every $v \in V$, there exists $w \in V$ s.t. $v + w = 0$
- Multiplicative identity: $1v = v$ for all $v \in V$
- Distributive properties

Notation 1.3. $\mathbb{F}^S$
If S is a set, then $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$.

1.C Subspaces

Definition 1.4. Subspace
A subset U of V is called a subspace of V if U is also a vector space (using the same addition and scalar multiplication as on V).

Proposition 1.5. Conditions for a subspace
That is, a subset U of V is a subspace of V iff U satisfies three conditions:

- Additive identity
- Closed under addition
- Closed under scalar multiplication

Proof. If U is a subspace of V, then the three conditions are satisfied by the definition of a vector space.

Conversely, suppose U satisfies the three conditions above. The first condition ensures the additive identity of V is in U. The second condition ensures addition makes sense on U. The third condition ensures that scalar multiplication makes sense on U and that every $u \in U$ has an additive inverse in U, namely $(-1)u$. Finally, associativity, commutativity, and distributive properties are inherited from V. Thus, U satisfies all properties defining a vector space and is a subspace of V. □

Definition 1.6. Sum of subsets
Given subsets $U_1, \ldots, U_m$ of V, define the sum

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Definition 1.7. Direct sum
Given subspaces $U_1, \ldots, U_m$ of V, the sum $U_1 + \cdots + U_m$ is called a direct sum if each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j. If $U_1 + \cdots + U_m$ is a direct sum, then it may be denoted $U_1 \oplus \cdots \oplus U_m$ to indicate that it is a direct sum.

Proposition 1.8. Condition for a direct sum
Given subspaces $U_1, \ldots, U_m$ of V, the sum $U_1 + \cdots + U_m$ is a direct sum iff the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each $u_j = 0$.

Proof. First suppose $U_1 + \cdots + U_m$ is a direct sum. Then by definition, there is only one way to write 0 as a sum $u_1 + \cdots + u_m$, namely, taking each $u_j = 0$.

Now suppose the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each $u_j = 0$. To show that $U_1 + \cdots + U_m$ is a direct sum, consider some element $v \in U_1 + \cdots + U_m$. We can write $v = u_1 + \cdots + u_m$ for some $u_1 \in U_1, \ldots, u_m \in U_m$, and to show this representation is unique, suppose we have $v = v_1 + \cdots + v_m$ where $v_1 \in U_1, \ldots, v_m \in U_m$. Then subtracting, we have

$$0 = (u_1 - v_1) + \cdots (u_m - v_m).$$

Because $u_1 - v_1 \in U_1, \ldots, u_m - v_m \in U_m$, then from our assumption that 0 may be written only by taking each $u_j = 0$, we know that each $u_j - v_j$ equals 0, i.e., $u_1 = v_1, \ldots, u_m = v_m$ as desired. □

Proposition 1.9. Direct sum of two subspaces
Suppose U and W are subspaces of V. Then $U + W$ is a direct sum iff $U \cap W = \{0\}$.

Proof. First suppose that $U + W$ is a direct sum. Then if $v \in U \cap W$, note that $-v \in U \cap W$. Thus $0 = v + (-v)$, where $v \in U$ and $-v \in W$. By the unique representation of 0 as a sum of a vector in U and a vector in W, we have $v = 0$ and thus $U \cap W = \{0\}$.

Conversely, now suppose $U \cap W = \{0\}$. To prove $U + W$ is a direct sum, suppose $u \in U$, $w \in W$, and $u + w = 0$. By 1.8, we now need only show that $u = w = 0$. This implies $w = -u \in U$, thus $w \in U \cap W$. Hence $w = 0$ and therefore $u = 0$, as desired. □

2 Finite-Dimensional Vector Spaces

Notation 2.1. V
Standing notation throughout this section:

- V denotes a vector space over $\mathbb{F}$.

2.A Span and linear independence

Definition 2.2. Linear combination
Defined intuitively.

Definition 2.3. Span
Defined intuitively. Note that the span of the empty list () is defined to be $\{0\}$.

Proposition 2.4. Span is the smallest containing subspace
The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Proof. Suppose $v_1, \ldots, v_m \in V$. First we show that $\operatorname{span}(v_1, \ldots, v_m)$ is a subspace of V. The additive identity is contained because we can take the linear combination with all coefficients equal to zero. Also, $\operatorname{span}(v_1, \ldots, v_m)$ is closed under addition because

$$(a_1 v_1 + \cdots + a_m v_m) + (c_1 v_1 + \cdots + c_m v_m) = (a_1 + c_1)v_1 + \cdots + (a_m + c_m)v_m.$$

Finally, $\operatorname{span}(v_1, \ldots, v_m)$ is closed under scalar multiplication by a similar argument. Thus, we have shown $\operatorname{span}(v_1, \ldots, v_m)$ is a subspace of V.

Now we show that $\operatorname{span}(v_1, \ldots, v_m)$ is the smallest subspace containing $v_1, \ldots, v_m$. Trivially, $\operatorname{span}(v_1, \ldots, v_m)$ contains each v_j. Conversely, because subspaces are closed under scalar multiplication and addition, every subspace of V containing each v_j contains $\operatorname{span}(v_1, \ldots, v_m)$. Thus $\operatorname{span}(v_1, \ldots, v_m)$ is the smallest subspace of V containing all the vectors $v_1, \ldots, v_m$. □

Definition 2.5. Finite-dimensional vector space
A vector space is finite-dimensional if some list of vectors in it spans the space. Otherwise, the vector space is infinite-dimensional.

Notation 2.6. $\mathcal{P}_m(\mathbb{F})$
For m a nonnegative integer, $\mathcal{P}_m(\mathbb{F})$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m. Note that $\mathcal{P}_m(\mathbb{F})$ contains the polynomial 0, because the polynomial that is identically 0 is said to have degree $-\infty$.

Definition 2.7. Linearly independently
A list $v_1, \ldots, v_m$ of vectors in V is called linearly independent if the only choice of $a_1, \ldots, a_m \in \mathbb{F}$ that makes $a_1v_1 + \cdots + a_mv_m$ equal 0 is $a_1 = \cdots = a_m = 0$. Otherwise, the list is called linearly dependent.
Note that the empty list () is also declared to be linearly independent.

Lemma 2.8. Linear Dependence Lemma
Suppose $v_1, \ldots, v_m$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, \ldots, m\}$ such that follow hold:

(a) $v_j \in \text{span}(v_1, \ldots, v_{j-1})$;

(b) if the j^{th} term is removed from $v_1, \ldots, v_m$, the span of the remaining list equals $\text{span}(v_1, \ldots, v_m)$.

Proof. Because the list $v_1, \ldots, v_m$ is linearly dependent, there exists numbers $a_1, \ldots, a_m \in \mathbb{F}$, not all 0, such that

$$a_1v_1 + \cdots + a_mv_m = 0.$$

Let j be the largest element of $\{1, \ldots, m\}$ such that $a_j \neq 0$. Then we have

$$v_j = -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1}, \tag{2.9}$$

hence $v_j \in \text{span}(v_1, \ldots, v_{j-1})$, proving (a).

To prove (b), suppose we have $u \in \text{span}(v_1, \ldots, v_m)$. Then we have

$$u = c_1v_1 + \ldots + c_mv_m.$$

Again considering v_j, we substitute the RHS from our above equation 2.9. This yields that u is a linear combination of the vectors in the list obtained by removing v_j from $v_1, \ldots, v_m$, and thus u is in their span, proving (b). □

Proposition 2.10. Length of linearly independent list $\leq$ length of spanning list
In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof. Suppose $u_1, \ldots, u_m$ is linearly independent in V. Suppose also that $w_1, \ldots, w_n$ spans V. We need to prove that $m \leq n$. We do so through the multi-step process described below; in each step we add one of the u's and remove one of the w's.

Step 1: Let B be the list $w_1, \ldots, w_n$ which spans V. Thus adjoining any vector in V to B produces a linearly dependent list (because the new vector can be written as a linear combination of the other vectors). In particular, $u_1, w_1, \ldots, w_n$ is linearly dependent. Thus by the Linear Dependence Lemma (2.8), we can remove one of the w's so that the new list B (of length n), consisting of u_1 and the remaining w's, spans V.

Step j: The list B after step $j - 1$ spans V. Thus adjoining any vector to this list produces a linearly dependent list. In particular, we add u_j to B, placing it just after $u_1, \ldots, u_{j-1}$. By the Linear Dependence Lemma (2.8), one of the vectors in the list is in the span of the previous vectors,

and because $u_1, \ldots, u_j$ are linearly independent, this vector is one of the w's. We can remove that w from B so that the new list B (of length n) consisting of $u_1, \ldots, u_j$ and the remaining w's spans V.

After step m, we will have added all of $u_1, \ldots, u_m$ and the process ends. At each step as we add a u to B, the Linear Dependence Lemma implies that there is some w to remove. Thus there are at least as many w's as u's. □

Proposition 2.11. Finite-dimensional subspaces
Every subspace of a finite-dimensional vector space is finite-dimensional.

Proof. Suppose V is finite-dimensional and U is a subspace of V. We need to prove that U is finite-dimensional. We do this through the following multi-step construction.

Step 1: If $U = \{0\}$, then U is finite-dimensional and we are done. Otherwise, choose a vector $v_1 \in U$.

Step j: If $U = \text{span}(v_1, \ldots, v_{j-1})$, then U is finite-dimensional and we are done. Otherwise, adjoin a vector $v_j \in U$ such that $v_j \notin \text{span}(v_1, \ldots, v_{j-1})$.

After each step, our list of vectors will be linearly independent by the Linear Dependence Lemma (2.8), because no vector is in the span of the previous vectors. By 2.10, the length of this list cannot be longer than any spanning list of V. Thus the process eventually terminates, meaning U is finite-dimensional. □

Example 2.12. Axler, page 38, exercise 2.A.14
Prove that V is infinite-dimensional iff there is a sequence $v_1, v_2, \ldots$ of vectors in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m.

Proof. If there is a sequence $v_1, v_2, \ldots$ of vectors in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m, then obviously V is infinite-dimensional.

Conversely, if V is infinite-dimensional, then we may obtain such a list $v_1, v_2, \ldots$ of vectors in V by induction. Let $v_1 \neq 0$ be a vector in V. Since V is infinite-dimensional, there must exist $v_2 \in V$ such that $v_2 \notin \text{span}(v_1)$. Similarly, if $v_1, \ldots, v_m$ is linearly independent, then there must exist v_{m+1} such that $v_{m+1} \notin \text{span}(v_1, \ldots, v_m)$. Since V is infinite-dimensional, we can always continue this process, and by the Linear Dependence Lemma (2.8), we deduce that $v_1, \ldots, v_m$ is linearly independent for every positive integer m. □

2.B Bases

Definition 2.13. Basis
A basis of V is a list of vectors in V that is linearly independent and spans V.

Proposition 2.14. Criterion for a basis
A list of $v_1, \ldots, v_n$ of vectors in V is a basis of V iff every $v \in V$ can be written uniquely in the form

$$v = a_1 v_1 + \cdots + a_n v_n \tag{2.15}$$

where $a_1, \ldots, a_n \in \mathbb{F}$.

Proof. First suppose $v_1, \ldots, v_n$ is a basis of V. Let $v \in V$. Because $v_1, \ldots, v_n$ spans V, then there exist $a_1, \ldots, a_n \in \mathbb{F}$ such that 2.15 holds. To show this representation is unique, suppose $c_1, \ldots, c_n$ exist so that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Then we may subtract this equation from 2.15 and use the unique representation of 0 to deduce that $a_1 = c_1, \ldots, a_n = c_n$.

Conversely, suppose that every $v \in V$ can be written in the form given by 2.15. Clearly, $v_1, \ldots, v_n$ spans V. Furthermore, $v_1, \ldots, v_n$ is linearly independent because the unique representation of each $v \in V$ in the form of 2.15 implies that there is a unique representation of 0, namely, $a_1 = \cdots = a_n = 0$ as desired. Thus $v_1, \ldots, v_n$ spans V and is linearly independent, and hence is a basis of V. □

Proposition 2.16. Spanning list contains a basis
Evert spanning list in a vector space can be reduced to a basis of the vector space (by discarding some, or possibly none, of the vectors in it).

Proof. Suppose the list B equal to $v_1, \ldots, v_n$ spans V. We want to remove some of the vectors in B to obtain a basis of V. We do so through the multi-step process described below.

Step 1: If $v_1 = 0$, then remove v_1. Else, leave B unchanged.

Step j: If $v_j \in \text{span}(v_1, \ldots, v_{j-1})$, remove v_j from B. Else, leave B unchanged.

After step n, the new list B will still span V because the original list spanned V and we did not remove any vector not in the span of the previous vectors. Further, the new list is linearly independent by the Linear Dependence Lemma (2.8). Thus B is a basis of V. □

Corollary 2.17.
Every finite-dimensional vector space has a basis.

Proof. By definition, a finite-dimensional vector space has a spanning list. The previous result tells us that each spanning list can be reduced to a basis. □

Proposition 2.18. Linearly independent list extends to a basis
Every linearly independent list of vectors in a finite-dimensional space can be extended to a basis of the vector space.

Proof. Suppose $u_1, \ldots, u_m$ is linearly independent and $w_1, \ldots, w_n$ is a basis of V. Then applying the procedure of the proof of 2.16, we may obtain a basis of V consisting of $u_1, \ldots, u_m$ and some of the w's (none of the u's are deleted because they are linearly independent). □

Proposition 2.19. Every subspace of V is a part of a direct sum equal to V
Suppose V is finite-dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.

Proof. By 2.11, suppose $u_1, \ldots, u_m$ is a basis of U. By 2.18, we may extend this list to a basis of $u_1, \ldots, u_m, w_1, \ldots, w_n$ of V. Now let $W = \text{span}(w_1, \ldots, w_n)$.

To prove that $V = U \oplus W$, we need to show that

$$V = U + W \quad \text{and} \quad U \cap W = \{0\}.$$

To prove the first equation, suppose $v \in V$. Because $u_1, \ldots, u_m, w_1, \ldots, w_n$ spans V, there exists $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathbb{F}$ such that

$$v = \underbrace{a_1 u_1 + \cdots + a_m u_m}_{u} + \underbrace{b_1 w_1 + \cdots + b_n w_n}_{w}.$$

Defining $u \in U$ and $w \in W$ as above, we have $v = u + w$ and thus $V = U + W$.

To show that $U \cap W = \{0\}$, consider $v \in U \cap W$. There exist scalars $a_1, \ldots, a_m, b_1, \ldots, b_n$ such that

$$v = a_1 u_1 + \cdots + a_m u_m = b_1 w_1 + \cdots + b_n w_n.$$

Thus,

$$0 = a_1 u_1 + \cdots + a_m u_m - b_1 w_1 - \cdots - b_n w_n.$$

Because $u_1, \ldots, u_m, w_1, \ldots, w_n$ are linearly independent, we have $a_1 = \cdots = a_m = b_1 = \cdots = b_m = 0$. Thus $v = 0$, so $U \cap W = \{0\}$ as desired. □

2.C Dimension

Proposition 2.20. Basis length does not depend on basis
Any two bases of a finite-dimensional vector space have the same length.

Proof. Consider two bases B_1, B_2 of V. Viewing B_1 as a linearly independent list and B_2 as a spanning list, 2.10 yields that the length of B_1 is at most the length of B_2. Interchanging the roles of B_1 and B_2, we see also that the length of B_2 is at most the length of B_1. Thus the length of B_1 equals the length of B_2, as desired. □

Definition 2.21. Dimension, $\dim V$
The dimension of a finite-dimensional vector space V is the length of any basis of V, and is denoted $\dim V$.

Proposition 2.22. Dimension of a subspace
If V is finite-dimensional and U is a subspace of V, then $\dim U \leq \dim V$.

Proof. Think of a basis of U as a linearly independent list of vectors in V, and think of a basis in V as a spanning list in V. By 2.10, $\dim U \leq \dim V$. □

Proposition 2.23. Linearly independent list of the right length is a basis
Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length $\dim V$ is a basis V.

Proof. Suppose $\dim V = n$ and $v_1, \ldots, v_n$ is a linearly independent list of vectors in V. By 2.18, we may extend $v_1, \ldots, v_n$ to a basis of V. But since every basis of V has length n, the extension is the trivial one, meaning no elements are adjoined. Thus $v_1, \ldots, v_n$ is a basis of V, as desired. □

Proposition 2.24.
If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Proof. Let $u_1, \ldots, u_m$ be a basis of $U_1 \cap U_2$; thus $\dim(U_1 \cap U_2) = m$. Because $u_1, \ldots, u_m$ is linearly independent in U_1, this list can be extended to a basis $u_1, \ldots, u_m, v_1, \ldots, v_j$ of U_1 (by 2.18). Thus $\dim U_1 = m + j$. Also extend $u_1, \ldots, u_m$ to a basis $u_1, \ldots, u_m, w_1, \ldots, w_k$ of U_2; thus $\dim(U_1 \cap U_2) = m + k$.

We will show that

$$u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$$

is a basis of $U_1 + U_2$. This will complete the proof, because then we will have

$$\begin{aligned}\dim(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)\end{aligned}$$

Clearly $\operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k)$ contains U_1 and U_2 and hence equals $U_1 + U_2$. So to show that this list is a basis of $U_1 + U_2$, we need only show that it is linearly independent. To prove this, suppose

$$a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_j v_j + c_1 w_1 + \cdots + c_k w_k = 0.$$

We need to prove that all the a's, b's, and c's equal 0. The equation can be rewritten as

$$c_1 w_1 + \cdots + c_k w_k = -(a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_j v_j)$$

We see $c_1 w_1 + \cdots + c_k w_k \in U_2$ because it is a linear combination of $w_1, \ldots, w_k \in U_2$, and also $c_1 w_1 + \cdots + c_k w_k \in U_1$ because it is a linear combination of $u_1, \ldots, u_m, v_1, \ldots, v_j$. Now because $u_1, \ldots, u_m$ is a basis of $U_1 \cap U_2$, we can write

$$c_1 w_1 + \cdots + c_k w_k = d_1 u_1 + \cdots + d_m u_m$$

for some scalars $d_1, \ldots, d_m$. But $u_1, \ldots, u_m, w_1, \ldots, w_k$ is linearly independent, hence all the c's (and d's) equal 0. Thus the original equation with a's, b's, and c's becomes

$$a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_j v_j = 0.$$

Because $a_1, \ldots, a_m, b_1, \ldots, b_j$ is linearly independent, this equation implies that all the a's and b's equal 0. Thus $u_1, \ldots, u_m, v_1, \ldots, v_j, w_1, \ldots, w_k$ are linearly independent, as desired. □

3 Linear Maps

Notation 3.1. V, W
Standing notation throughout this section:

- V and W denote vector spaces over $\mathbb{F}$.

3.A The vector space of linear maps

Definition 3.2. Linear map
A linear map from V to W is a function $T : V \to W$ with the following properties:

- Additivity: $T(u + v) = Tu + Tv$ for all $u, v \in V$
- Homogeneity: $T(\lambda v) = \lambda T(v)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$

For linear maps we use the notation Tv as well as the more standard functional notation $T(v)$.

Notation 3.3. $\mathcal{L}(V, W)$
The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Proposition 3.4. Linear maps and basis of domain
Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that

$$Tv_j = w_j$$

for each $j = 1, \ldots, n$.

Proof. First we show existence. Let

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n$$

where $c_1, \ldots, c_n$ are arbitrary elements of $\mathbb{F}$. It is easy to verify that T is a function from V to W as well that T satisfies additivity and homogeneity, and so T is a linear map as desired. For each j, letting $c_j = 1$ and all other c's equal to 0 yields that $Tv_j = w_j$.

Now we show uniqueness. Suppose $T \in \mathcal{L}(V, W)$ such that $Tv_j = w_j$. Then homogeneity and additivity imply that

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n.$$

Thus T is uniquely determined on $V = \operatorname{span}(v_1, \ldots, v_n)$ by the above equation. □

Definition 3.5. Addition and scalar multiplication on $\mathcal{L}(V, W)$
Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. Define the linear maps

$$(S + T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.
It is easy to verify that $\mathcal{L}(V, W)$ is a vector space with the operations of addition and scalar multiplication as defined above.

Definition 3.6. Product of linear maps
If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ for vector spaces U, V, W over $\mathbb{F}$, then define the linear map

$$(ST)(u) = S(Tu)$$

for $u \in U$.
See that ST is the usual composition $S \circ T$ of two functions, but most mathematicians write ST instead when both functions are linear.

Proposition 3.7. Algebraic properties of linear maps
It is easy to verify that the following hold:

- Associativity:
$$(T_1T_2)T_3 = T_1(T_2T_3)$$
whenever T_1, T_2, T_3 map into the appropriate domains.
- Identity:
$$TI = IT = T$$
where $T \in \mathcal{L}(V, W)$, the first I is the identity map on V, and the second I is the identity map on W.
- Distributivity:
$$(S_1 + S_2)T = S_1T + S_2T \quad \text{and} \quad S(T_1 + T_2) = ST_1 + ST_2$$
where $T, T_1, T_2 \in \mathcal{L}(U, V)$ and $S, S_1, S_2 \in \mathcal{L}(V, W)$.

Note that multiplication of linear maps is not commutative, i.e., it is not necessarily true that $ST = TS$ even if both sides of the equation make sense.

Proposition 3.8. Linear maps take 0 to 0
Suppose $T \in (V, W)$. Then $T(0) = 0$.

Proof. By additivity, we have

$$T(0) = T(0 + 0) = T(0) + T(0) \implies T(0) = 0.$$

□

Example 3.9. Exercise 3.A.3
Suppose $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. Show that there exist scalars $A_{j,k} \in \mathbb{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$ such that

$$T(x_1, \dots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \dots, A_{m,1}x_1 + \cdots + A_{m,n}x_n)$$

for every $(x_1, \dots, x_n) \in \mathbb{F}^n$.

Proof. Denote

$$\begin{aligned} T(1, 0, \dots, 0) =& (A_{1,1}, \dots, A_{1,m}) \\ &\vdots \\ T(0, \dots, 0, 1) =& (A_{1,n}, \dots, A_{m,n}). \end{aligned}$$

Noting that $(1, 0 \dots, 0), \dots, (0, \dots, 0, 1)$ is a basis of $\mathbb{F}^n$, then by the proof of 3.4 we have a linear map T such that

$$T(x_1, \dots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \dots, A_{m,1}x_1 + \cdots + A_{m,n}x_n).$$

□

Example 3.10. Exercise 3.A.4
Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_m$ is a list of vectors in V such that $Tv_1, \dots, Tv_m$ is a linearly independent list in W. Prove that $v_1, \dots, v_m$ is linearly independent.

Proof. Suppose $c_1v_1 + \cdots + c_mv_m = 0$. Then $T(c_1v_1 + \cdots + c_mv_m) = c_1Tv_1 + \cdots + c_mTv_m = T(0) = 0$. Because $Tv_1, \dots, Tv_m$ are linearly independent, then $c_1 = \cdots = c_m = 0$, as desired. □

3.B Null spaces and ranges

Null space and injectivity

Definition 3.11. Null space, $\operatorname{null} T$
For $T \in \mathcal{L}(V, W)$, the null space of T, denoted $\operatorname{null} T$, is the subset of V consisting of those vectors that T maps to 0:

$$\operatorname{null} T = \{v \in V : T(v) = 0\}.$$

The null space is also called the kernel.

Proposition 3.12. The null space is a subspace
Suppose $T \in \mathcal{L}(V, W)$. Then $\operatorname{null} T$ is a subspace of V.

Proof. First, we show the additive identity is in $\operatorname{null} T$. By 3.8, we have that $T(0) = 0$ and thus $0 \in \operatorname{null} T$. Second, we show closure under addition. Suppose $u, v \in \operatorname{null} T$. Then by additivity, $T(u + v) = T(u) + T(v) = 0 + 0 = 0$ and thus $u + v \in \operatorname{null} T$. Finally, we show closure under scalar multiplication. Suppose $v \in \operatorname{null} T$ and $\lambda \in \mathbb{F}$. Then by homogeneity, $T(\lambda v) = \lambda T(v) = \lambda 0 = 0$ and thus $\lambda v \in \operatorname{null} T$. Thus by 1.5, $\operatorname{null} T$ is a subspace of V. □

Definition 3.13. Injective
A function $T : V \to W$ is called injective if $Tu = Tv$ implies $u = v$.
Equivalently, T is injective if $u \neq v$ implies $Tu \neq Tv$. In other words, T is injective if it maps distinct inputs to distinct outputs. Injectivity is also termed one-to-one.

Proposition 3.14. Injectivity is equivalent to null space equals $\{0\}$
Let $T \in \mathcal{L}(V, W)$. Then T is injective iff $\operatorname{null} T = \{0\}$.

Proof. First suppose T is injective. By 3.8, we know that $\{0\} \subseteq \operatorname{null} T$. Then $T(v) = 0 = T(0)$ implies that $v = 0$ because T is injective, hence $\operatorname{null} T = \{0\}$, as desired.

Conversely, suppose $\operatorname{null} T = \{0\}$. Then $T(u) = T(v)$ implies $T(u) - T(v) = T(u - v) = 0$. Thus $u - v \in \operatorname{null} T$, so $u - v = 0$ and $u = v$. Hence T is injective, as desired. □

Example 3.15. Exercise 3.B.20
Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is injective iff there exists $S \in \mathcal{L}(W, V)$ s.t. ST is the identity map on V.

Proof. First suppose T is injective. Denote $S' : \operatorname{range} T \to V$ the function s.t. $S'(Tv) = v$. It is easily verified that S' is linear and it is possible to extend S' to a map on W because $\operatorname{range} T$ is a subspace of W (see Exercise 3.A.11). Clearly for $v \in V$, we have that $S'T$ is the identity map.

Now suppose that $S \in \mathcal{L}(W, V)$ exists s.t. ST is the identity on V. Consider $u, v \in V$ s.t. $Tu = Tv$, then

$$u = ST(u) = S(Tu) = S(Tv) = v,$$

so T is injective as desired. □

Range and surjectivity

Definition 3.16. Range
For T a function from V to W, the range of T is the subset of W consisting of the vectors of form Tv for some $v \in V$:

$$\operatorname{range} T = \{Tv : v \in V\}.$$

The range is also called the image.

Proposition 3.17. The range is a subspace
If $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T$ is a subspace of W.

Proof. Routine proof using 1.5. □

Definition 3.18. Surjective
A function $T : V \to W$ is surjective if its range equals W.
Note that whether a linear map is surjective depends on what we are thinking of as the vector space into which it maps. For example, the differentiation map $D \in \mathcal{L}\big(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R})\big)$ is surjective, while the differentiation map $S \in \mathcal{L}(\mathcal{P}_5(\mathbb{R}), \mathcal{P}_5(\mathbb{R}))$ is not surjective because $\operatorname{range} S = \mathcal{P}_4(\mathbb{R})$.

Fundamental theorem of linear maps

Theorem 3.19. Fundamental theorem of linear maps
Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\operatorname{range} T$ is finite-dimensional and

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof. Let $\dim \operatorname{null} T = m$ and $u_1, \ldots, u_m$ be a basis of $\operatorname{null} T$. Then this linearly independent list may be extended to a basis of V, say

$$u_1, \ldots, u_m, v_1, \ldots, v_n.$$

Now we show $\dim \operatorname{range} T = n$ by proving that $Tv_1, \ldots, Tv_n$ is a basis of $\operatorname{range} T$. Let $v \in V$. Then

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n$$

for a's and b's in $\mathbb{F}$. Applying T to both sides of the equation, we have

$$\begin{aligned} Tv &= T(a_1 u_1 + \cdots + a_m u_m) + T(b_1 v_1 + \cdots + b_n v_n) \\ &= b_1 T v_1 + \cdots + b_n T v_n. \end{aligned}$$

Hence $Tv \in \operatorname{span}(Tv_1, \ldots, Tv_n)$. To show $Tv_1, \ldots, Tv_n$ is linearly independent, suppose

$$c_1 T v_1 + \cdots + c_n T v_n = 0$$

for some $c_1, \ldots, c_n \in \mathbb{F}$. Then

$$T(c_1 v_1 + \cdots + c_n v_n) = 0.$$

Thus $c_1 v_1 + \cdots + c_n v_n \in \operatorname{null} T$ and so

$$c_1 v_1 + \cdots + c_n v_n = d_1 u_1 + \cdots + d_m u_m$$

for some $d_1, \ldots, d_m \in \mathbb{F}$. But because $u_1, \ldots, u_m, v_1, \ldots, v_n$ are linearly independent, then all the c's (and d's) equal 0. Thus $Tv_1, \ldots, Tv_n$ is both linearly independent and a spanning list of $\operatorname{range} T$, as desired. □

Proposition 3.20. A map to a smaller dimensional space is not injective
Suppose V and W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof. Let $T \in \mathcal{L}(V, W)$. By 3.14, we need to prove that $\operatorname{null} T$ contains vectors other than 0. By the fundamental theorem of linear maps (3.19), we have

$$\begin{aligned} \dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &\geq \dim V - \dim W \\ &> 0, \end{aligned}$$

hence $\operatorname{null} T$ contains vectors other than 0 and so T is not injective. □

Proposition 3.21. A map to a larger dimensional space is not surjective
Suppose V and W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map form V to W is surjective.

Proof. Similarly to the above proof, we use the fundamental theorem of linear maps (3.19) to show that $\dim \operatorname{range} T < \dim W$ and hence $\operatorname{range} T$ cannot equal W. □

Proposition 3.22. Homogenous system of linear equations
A homogeneous system of linear equations (where the RHS of each equation is 0) with more variables than equations has nonzero solutions.

Proof. Represent the system using a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ where n is the number of variables and m is the number of equations. That is, solutions to the system correspond to solutions to $T(x_1, \ldots, x_n) = 0$. From 3.20 we see that T is not injective if $n > m$, so there $\operatorname{null} T$ contains vectors other than 0. □

Proposition 3.23. Inhomogenous system of linear equations
An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.

Proof. Again represent the system as a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$. From 3.21 we see that T is not surjective if $n < m$. □

3.C Matrices

Representing a linear map by a matrix

Definition 3.24. Matrix of a linear map
Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The matrix of T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases are not clear from context, then the notation $\mathcal{M}\big(T, (v_1, \ldots, v_n), (w_1 \ldots, w_m)\big)$ is used.

See this corresponding diagram:

$$\begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{cccccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \begin{pmatrix} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{pmatrix} \end{array}$$

Addition and scalar multiplication of matrices

It is easily verified that the intuitive definitions of matrix addition and scalar multiplication have the right properties, i.e., $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$ and $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$ for $S, T \in \mathcal{L}(V, W)$.

Notation 3.25. $\mathbb{F}^{m,n}$
For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{m,n}$.

Proposition 3.26. $\dim \mathbb{F}^{m,n} = mn$
With the intuitive matrix addition and scalar multiplication operations, $\mathbb{F}^{m,n}$ is a vector space with dimension mn.

Proof. Showing $\mathbb{F}^{m,n}$ is a vector space is routine. It is easily verified that the list of m-by-n matrices that have 0 in all entries except for 1 in one entry is a basis of $\mathbb{F}^{m,n}$, hence $\dim \mathbb{F}^{m,n} = mn$. □

Matrix multiplication

Definition 3.27. Matrix multiplication
Matrix multiplication is purposely defined in a way such that $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$. Matrix multiplication is distributive and associative, but not commutative.

Notation 3.28. $A_{j,\cdot}, A_{\cdot,k}$
Suppose A is an m-by-n matrix. Denote by $A_{j,\cdot}$ the 1-by-n matrix consisting of row j of A, and denote by $A_{\cdot,k}$ the m-by-1 matrix consisting of column k of A. Also note that we frequently identify a 1-by-1 matrix with its entry.

Remark 3.29.
Suppose $A \in \mathbb{F}^{m,n}$ and $C \in \mathbb{F}^{n,p}$. We have many ways to think of matrix multiplication:

- Each entry of AC is the product of a row and column:

$$(AC)_{j,k} = Aj, \cdot C_{\cdot,k}$$

- Each column of AC is the product of A and a column:

$$(AC)_{\cdot,k} = AC_{\cdot,k}$$

  And similarly for rows.

3.D Invertibility and isomorphic vector spaces

Invertible linear maps

Definition 3.30. Invertible, inverse
A linear map $T \in \mathcal{L}(V, W)$ is called invertible if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equals the identity map on W. Such a linear map S is called an inverse of T and (because it is unique) denoted T^{-1}.

Proposition 3.31. Inverse is unique
An invertible linear map has a unique inverse.

Proof. Suppose $T \in \mathcal{L}(V, W)$ and S_1, S_2 are inverses of T. Then

$$S_1 = (S_2T)S_1 = S_2(TS_1) = S_2.$$

□

Proposition 3.32. Invertibility is equivalent to injectivity and surjectivity
A linear map is invertible iff it is injective and surjective.

Proof. Suppose $T \in \mathcal{L}(V, W)$.

First suppose T is invertible. To show T is injective, suppose $u, v \in V$ with $Tu = Tv$. Then

$$u = T^{-1}(Tu) = T^{-1}(Tv) = v,$$

as desired. To show T is surjective, suppose $w \in W$. Then $w = T(T^{-1}w)$, so $\operatorname{range} T = W$, as desired.

Now suppose T is injective and surjective, so we need to prove that T is invertible. For each $w \in W$, denote by Sw the element of V such that $T(Sw) = w$. The existence and uniqueness of Sw follows from surjectivity and injectivity of T, respectively. Then $T \circ S$ is the identity on W (though we do not yet know S is linear). To show $S \circ T$ is the identity on V, let $v \in V$. Then

$$Tv = T \circ S(Tv) = T\big((S \circ T)v\big).$$

Because T is injective, we have $v = (S \circ T)v$ and so $S \circ T$ is the identity on V.

To finish the proof, we need to show that S is linear by demonstrating additivity and homogeneity. To show additivity, suppose $w_1, w_2 \in W$. Then

$$T(Sw_1 + Sw_2) = T(Sw_1) + T(Sw_2) = w_1 + w_2.$$

Thus $Sw_1 + Sw_2$ is the unique element of V that T maps to $w_1 + w_2$. By the definition of S, we thus have $S(w_1 + w_2) = Sw_1 + Sw_2$, as desired. Homogeneity is proved similarly, and thus S is linear, as desired. □

Isomorphic vector spaces

Definition 3.33. Isomorphism, isomorphic
An isomorphism is an invertible linear map. Two vector spaces are called isomorphic if there is an isomorphism from one vector space onto the other one.

Proposition 3.34. Dimension shows whether vector spaces are isomorphic
Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic iff they have the same dimension.

Proof. First suppose V, W are isomorphic with an isomorphism T between V and W. Thus $\operatorname{range} T = W$ and $\operatorname{null} T = \{0\}$. Then by the fundamental theorem of linear maps 3.19, we have $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim W$, as desired.

Now suppose V, W have the same dimension with bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$, respectively. Define the linear map

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n,$$

which is a well-defined linear map by the proof of 3.4. T is surjective because we have $\operatorname{range} T = \operatorname{span}(w_1, \ldots, w_n) = W$. Also, $w_1, \ldots, w_n$ are linearly independent, so $\operatorname{null} T = \{0\}$ and thus T is injective, as desired. Thus there exists an isomorphism between V and W (by 3.32). □

Proposition 3.35. $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$ are isomorphic
Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then M is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.

Proof. First we show surjectivity. For any matrix $A \in \mathbb{F}^{m,n}$, we simply define T by

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

Then $\mathcal{M}(T) = A$, so $\operatorname{range} \mathcal{M} = \mathbb{F}^{m,n}$.

Now we show injectivity, i.e., $\operatorname{null} \mathcal{M} = \{0\}$. Suppose $\mathcal{M}(T) = 0$. Then $Tv_k = 0$ for all basis vectors v_k, hence $T = 0$, as desired. □

Proposition 3.36. $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$
By 3.35 and 3.34, $\dim \mathcal{L}(V, W) = \dim \mathbb{F}^{m,n}$. By 3.26, $\dim \mathcal{L}(V, W) = mn = (\dim V)(\dim W)$, as desired.

Linear maps thought of as matrix multiplication

Definition 3.37. Matrix of a vector, $\mathcal{M}(v)$
Suppose $v \in V$ and $v_1, \ldots, v_n$ is a basis of V. The matrix of v w.r.t. this basis is the n-by-1 matrix

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \ldots, c_n$ are the scalars s.t. $v = c_1 v_1 + \cdots + c_n v_n$. Note that the function $\mathcal{M}$ here is an isomorphism of V onto $\mathbb{F}^{n,1}$ (once the basis $v_1, \ldots, v_n$ is chosen), and because this isomorphism is so natural, we may think of it as simply relabeling vectors v as n-by-1 matrices.

Proposition 3.38.
Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

Proof. We have $Tv = c_1 Tv_1 + \cdots + c_n Tv_n$ for some $c_1, \ldots, c_n \in \mathbb{F}$. Then

$$\begin{aligned} \mathcal{M}(Tv) &= c_1 \mathcal{M}(Tv_1) + \cdots + c_n \mathcal{M}(Tv_n) \\ &= c_1 \mathcal{M}(T)_{\cdot,1} + \cdots + c_1 \mathcal{M}(T)_{\cdot,n} \\ &= \mathcal{M}(T)\mathcal{M}(v), \end{aligned}$$

as desired. Note that the final equality, which demonstrates how multiplying a matrix and vector gives a linear combination of the matrix's columns. □

Operators

Definition 3.39. Operator, $\mathcal{L}(V)$
A linear map from a vector space to itself is called an operator. The notation $\mathcal{L}(V)$ denotes the set of all operators on V, i.e., $\mathcal{L}(V) = \mathcal{L}(V, V)$.

Proposition 3.40. Injectivity of operators is equivalent to surjectivity in finite dimensions
Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is invertible;

(b) T is injective;

(c) T is surjective.

Note this conclusion does not apply for infinite dimensions, e.g., multiplication by x^2 on $\mathcal{P}(\mathbb{R})$ is injective but not surjective because 1 is not in the range.

Proof. To prove equivalence, we prove that (a) implies (b), (b) implies (c), and (c) implies (a).

First suppose (a). Then (b) follows.

Now suppose (b), that the operator T is injective. Then $\dim \operatorname{null} T = 0$, hence

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = \dim \operatorname{range} T.$$

And because $\operatorname{range} T$ is a subspace of V with the same dimension, then $\operatorname{range} T = V$, as desired.

Now suppose (c), that the operator T is surjective. Then $\dim \operatorname{range} T = \dim V$, hence

$$\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T = 0.$$

Thus T is both surjective and injective, and thus invertible, as desired. □

3.E Products and quotients of vector spaces

Definition 3.41. Product of vector spaces
Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbb{F}$. Define their product

$$V_1 \times \cdots \times V_m = \{(v_1, \ldots, v_m) : v_1 \in V_1, \ldots, v_m \in V_m\}.$$

Addition and scalar multiplication are defined intuitively, and they make $V_1 \times \cdots \times V_m$ a vector space itself.

Proposition 3.42. Dimension of a product is the sum of dimensions
Suppose $V_1, \ldots, V_m$ are finite-dimensional vector spaces over $\mathbb{F}$. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m.$$

Proof. Choose a basis of each V_j. For each basis vector of each V_j, consider the element of $V_1 \times \cdots \times V_m$ that equals the basis vector in the j^{th} slot and 0 in the other slots. The list of all such vectors is linearly independent and spans $V_1 \times \cdots \times V_m$. Thus it is a basis of length of $\dim V_1 + \cdots + \dim V_m$, as desired. □

Proposition 3.43. Products and direct sums
Suppose that $U_1, \ldots, U_m$ are subspaces of V. Define a linear map $\Gamma = U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by

$$\Gamma(u_1, \ldots, u_m) = u_1 + \cdots + u_m.$$

Then $U_1 + \cdots + U_m$ is a direct sum iff Γ is injective. (Also note Γ is surjective by the definition of $U_1 + \cdots + U_m$.)

Proof. The linear map Γ is injective iff the only way to write 0 as a sum $u_1 + \cdots + u_m$ is by taking each u_j equal to 0, as desired. □

Proposition 3.44. A sum is a direct sum if and only if dimensions add up
Suppose V is finite-dimensional and $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum iff

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

Proof. By 3.43, $U_1 + \cdots + U_m$ is a direct sum iff the above map Γ is injective. Thus we show that the equation

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m$$

holds iff Γ is injective.

We have that Γ is surjective. Thus by the fundamental theorem of linear maps, Γ is injective iff

$$\dim(U_1 \times \cdots \times U_m) = \dim(U_1 + \cdots + U_m).$$

By 3.42, we have that $\dim(U_1 \times \cdots \times U_m) = \dim U_1 + \cdots + \dim U_m$, so Γ is injective iff

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m,$$

as desired. □

Quotients of vector spaces

Definition 3.45. $v + U$, affine subset, parallel
Suppose $v \in V$ and U is a subspace of V.Then $v + U$ is the subset of V defined by

$$v + U = \{v + u : u \in U\}.$$

An affine subset of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V. The affine subset $v + U$ is said to be parallel to U.

Note that v may equal 0. For example, all the lines in $\mathbb{R}^2$ with slope 2 are parallel to $U = \{(x, 2x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$, including the line through the origin. However, also note that with this definition of parallel, if U is a plane then no line in $\mathbb{R}^3$ is considered to be an affine subset that is parallel to U.

Definition 3.46. Quotient space
Suppose U is a subspace of V. Then the quotient space V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U = \{v + U : v \in V\}.$$

The following result will help us make V/U into a vector space.

Proposition 3.47. Two affine subsets parallel to U are equal or disjoint
Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent:

(a) $v - w \in U$;

(b) $v + U = w + U$;

(c) $(v + U) \cap (w + U) \neq \emptyset$

Proof. First suppose (a). To show (a) implies (b), suppose $u \in U$. then

$$v + U = w + (v - w) + U \in w + U.$$

Thus $v + U \subseteq w + U$, and similarly $w + U \subseteq v + U$. Hence $v + U = w + U$, as desired.

That (b) implies (c) is obvious.

Now suppose (c). To show (c) implies (a), observe that there must exist $u_1, u_2 \in U$ s.t.

$$v + u_1 = w + u_2.$$

Hence $v - w = u_2 - u_1 \in U$, as desired. □

Definition 3.48. Addition and scalar multiplication on V/U
Suppose U is a subspace of V. Then addition and scalar multiplication are defined on V/U by

$$(v + U) + (w + U) = (v + w) + U \qquad \lambda(v + U) = \lambda v + U$$

for $v, w \in V$ and $\lambda \in \mathbb{F}$.

Proposition 3.49. Quotient space is a vector space
Suppose U is a subspace of V. Then V/U is a vector space with the operations of addition and scalar multiplication as defined above.

Proof. The potential problem with the above definitions addition and scalar multiplication on V/U is that the representation of an affine subset parallel to U is not unique. Specifically, suppose $v, w \in V$. Suppose also that $\hat{v}, \hat{w} \in V$ are such that $v + U = \hat{v} + U$ and $w + U = \hat{w} + U$.

To show that the definition of addition on V/U given above makes sense, we must show that $(v + w) + U = (\hat{v} + \hat{w}) + U$. By 3.47, we have

$$v - \hat{v} \in U \qquad \text{and} \qquad w - \hat{w} \in U.$$

Because U is a subspace of V and thus is closed under addition, this implies that $(v-\hat{v})+(w-\hat{w}) \in U$. Thus $(v + w) - (\hat{v} + \hat{w}) \in U$. Using 3.47 again, we see that

$$(v + w) + U = (\hat{v} + \hat{w}) + U,$$

as desired. Thus the definition of addition on V/U makes sense.

Similarly, to show that the definition of scalar multiplication on V/U makes sense, suppose $\lambda \in \mathbb{F}$. Because U is a subspace of V and thus is closed under scalar multiplication, we have $\lambda(v - \hat{v}) = \lambda v - \lambda\hat{v} \in U$. Hence 3.47 implies that $(\lambda v) + U = (\lambda\hat{v}) + U$, as desired.

Now addition and scalar multiplication have been defined on V/U. Note that the additive identity of V/U is $0 + U = U$ and the additive inverse of $v + U$ is $-v + U$. From here, verifying V/U is a vector space is routine. □

Definition 3.50. Quotient map, π
Suppose U is a subspace of V. To help compute the dimension of V/U, define the quotient map π as the linear map $\pi : V \to V/U$ given by

$$\pi(v) = v + U.$$

for $v \in V$. It is easy to verify π is linear.

Proposition 3.51. Dimension of a quotient space
Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim V/U = \dim V - \dim U.$$

Proof. Let π be the quotient map from V to V/U. Then $\operatorname{null}\pi = U$ and $\operatorname{range}\pi = V/U$. Then the fundamental theorem of linear maps yields that

$$\dim V = \dim U + \dim V/U,$$

as desired. □

Definition 3.52. $\tilde{T}$
Suppose $T \in \mathcal{L}(V, W)$. Define $\tilde{T} : V/(\operatorname{null} T) \to W$ by

$$\tilde{T}(v + \operatorname{null} T) = Tv.$$

Each linear map T on V induces a linear map $\tilde{T}$ on $V/(\operatorname{null} T)$.

To show this definition of $\tilde{T}$ makes sense, note that we must verify (as in the proof of 3.49) that $Tu = Tv$ for any $u, v \in V$ such that $u + \operatorname{null} T = v + \operatorname{null} T$. This is easy: we see $u - v \in \operatorname{null} T$ by 3.47, implying $T(u - v) = 0$ and thus $Tu = Tv$.

Proposition 3.53. Null space and range of $\tilde{T}$
Suppose $T \in \mathcal{L}(V, W)$. Then the following hold:

(a) $\tilde{T}$ is a linear map from $V/(\operatorname{null} T)$ to W;

(b) $\tilde{T}$ is injective;

(c) $\operatorname{range}\tilde{T} = \operatorname{range} T$;

(d) $V/(\operatorname{null} T)$ is isomorphic to $\operatorname{range} T$

Proof.

(a) Linearity is verified routinely.

(b) Suppose $v \in V$ and $\tilde{T}(v + \operatorname{null} T) = 0$. Then $Tv = 0$, so $v \in \operatorname{null} T$. Hence $\operatorname{null}\tilde{T} = 0$ (recall the additive identity for $V/(\operatorname{null} T)$ is $\operatorname{null} T$), so $\tilde{T}$ is injective, as desired.

(c) From the definition of $\tilde{T}$.

(d) From (b) and (c), if we think of $\tilde{T}$ as a mapping into $\operatorname{range} T$, then $\tilde{T}$ is an isomorphism from $V/(\operatorname{null} T)$ onto $\operatorname{range} T$.

□

3.F Duality

Definition 3.54. Linear functional
A linear functional on V is a linear map from V to $\mathbb{F}$, i.e., an element of $\mathcal{L}(V, \mathbb{F})$.

Definition 3.55. Dual space, V'
The dual space of V, denoted V', is the vector space of all linear functionals on V. In other words, $V' = \mathcal{L}(V, \mathbb{F})$.

Proposition 3.56. $\dim V' = \dim V$
Suppose V is finite-dimensional. Then V' is also finite-dimensional and $\dim V' = \dim V$.

Proof. Follows from 3.36. □

Definition 3.57. Dual basis
If $v_1, \ldots, v_n$ is a basis of V, then the dual basis of $v_1, \ldots, v_n$ is the list $\varphi_1, \ldots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V s.t.

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Proposition 3.58. Dual basis is a basis of the dual space
Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'.

Proof. Suppose $v_1, \ldots, v_n$ is a basis of V and $\varphi_1, \ldots, \varphi_n$ denote that dual basis.

To show that $\varphi_1, \ldots, \varphi_n$ are linearly independent elements of V', suppose

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0.$$

Now $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$ for $j = 1, \ldots, n$. The equation above thus shows that $a_1 = \cdots = a_n = 0$.

Thus $\varphi_1 + \cdots + \varphi_n$ is a linearly independent list of elements of V' with length $\dim V'$, and hence a basis of V', as desired. □

Definition 3.59. Dual map, T'
If $T \in \mathcal{L}(V, W)$, then the dual map of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$.

Observe that $T'(\varphi)$ is indeed a map from V to $\mathbb{F}$, i.e. $T'(\varphi) \in V'$. Linearity is verified routinely.

Proposition 3.60. Algebraic properties of dual maps
The following hold:

- $(S+T)' = S' + T'$ for all $S, T \in \mathcal{L}(V, W)$.
- $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbb{F}$ and all $T \in \mathcal{L}(V, W)$.
- $(ST)' = T'S'$ for all $T \in \mathcal{L}(U, V)$ and all $S \in \mathcal{L}(V, W)$.

Proof. The first two points are easily verified. To prove the third, suppose $\varphi \in W'$. Then

$$(ST)'\varphi = \varphi \circ (ST) = (\varphi \circ S) \circ T = T'(\varphi \circ S) = T'\big(S'(\varphi)\big) = (T'S')(\varphi).$$

The second equality holds because the composition of functions is associative, and the fourth equality follows from the definition of composition. □

The null space and range of the dual of a linear map

In what follows, we aim to describe $\operatorname{null} T'$ and $\operatorname{range} T'$ in terms of $\operatorname{range} T$ and $\operatorname{null} T$. The following definition will help.

Definition 3.61. Annihilator, U^0
For $U \subseteq V$, the annihilator of U is defined by

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}.$$

It is easy to verify U^0 is a subspace of V'.

Proposition 3.62. Dimension of the annihilator
Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V.$$

Proof. Let $u_1, \ldots, u_m$ be a basis of U. Extend this to a basis $u_1, \ldots, u_m, \ldots, u_n$ of V, and let $\varphi_1, \ldots, \varphi_m, \ldots, \varphi_n$ be the dual basis of V'. We will show that $\varphi_{m+1}, \ldots, \varphi_n$ is a basis of U^0.

Clearly $\varphi_{m+1}, \ldots, \varphi_n$ is a linearly independent list of elements in U^0. To show $\varphi_{m+1}, \ldots, \varphi_n$ spans (and therefore is a basis of) U^0, suppose $\varphi \in U^0$ and $a_1, \ldots, a_n \in \mathbb{F}$ s.t. $\varphi = a_1\varphi_1 + \cdots + a_n\varphi_n$. For $j \in \{1, \ldots, m\}$, we have $\varphi(u_j) = a_j = 0$. Thus $U^0 = \operatorname{span}(\varphi_{m+1}, \ldots, \varphi_n)$, as desired.

Note: Axler presents a slicker proof in the text but recommends also constructing the above proof (Exercise 3.F.24), which follows a standard pattern. □

Proposition 3.63. The null space of T'
Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

(a) $T' = (\operatorname{range} T)^0$;

(b) $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$.

Proof.

(a) First we show $\operatorname{null} T' \subseteq (\operatorname{range} T)^0$. Suppose $\varphi \in \operatorname{null} T'$. Then $T'(\varphi) = \varphi \circ T = 0$, where 0 is the zero linear functional on V. Thus for all $v \in V$,

$$(\varphi T)v = \varphi(Tv) = 0,$$

i.e., $\varphi \in (\operatorname{range} T)^0$.

Now we show $(\operatorname{range} T)^0 \subseteq \operatorname{null} T'$. Suppose $\varphi \in (\operatorname{range} T)^0$. Then $\varphi(Tv) = 0$ for all $v \in V$ and hence $0 = \varphi \circ T = T'(\varphi)$, where 0 is the zero linear functional on V. Thus $\varphi \in \operatorname{null} T'$, as desired. Therefore $\operatorname{null} T' = (\operatorname{range} T)^0$, proving (a).

(b) We have

$$\begin{aligned}\dim \operatorname{null} T &= \dim(\operatorname{range} T)^0 \\ &= \dim W - \dim \operatorname{range} T \\ &= \dim W - (\dim V - \dim \operatorname{null} T) \\ &= \dim \operatorname{null} T + \dim W - \dim V,\end{aligned}$$

where the first equality comes from (a), the second from 3.62, and the third from the fundamental theorem of linear maps.

□

Proposition 3.64. T surjective is equivalent to T' injective
Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is surjective iff T' is injective.

Proof. The map T is surjective iff $\operatorname{range} T = W$. Taking the annihilator of both sides, $\operatorname{range} T = W$ happens iff $(\operatorname{range} T)^0 = \{0\}$. (On the RHS, if a linear functional takes all $w \in W$ to 0, then it is the zero linear functional over W.) This can also be seen from 3.62.

From part (a) of 3.63 (which says $\operatorname{null} T' = (\operatorname{range} T)^0$), we see $(\operatorname{range} T)^0 = \{0\}$ iff $\operatorname{null} T' = \{0\}$, which happens iff T' is injective. □

Proposition 3.65. The range of T'
Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

(a) $\dim \operatorname{range} T' = \dim \operatorname{range} T$

(b) $\operatorname{range} T = (\operatorname{null} T)^0$

Proof.

(a) We have

$$\begin{aligned}\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' \\ &= \dim W - \dim(\operatorname{range} T)^0 \\ &= \dim \operatorname{range} T,\end{aligned}$$

where the first equality comes from 3.19, the second comes from 3.56 and 3.63(a), and the third equality comes from 3.62.

(b) First suppose $\varphi \in \operatorname{range} T'$. Thus there exists $\psi \in W'$ such that $T'(\psi) = \varphi$. To show $\varphi \in (\operatorname{null} T)^0$, suppose $v \in \operatorname{null} T$. Then

$$\varphi(v) = \big(T'(\psi)\big)(v) = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0.$$

Hence $\varphi \in (\operatorname{null} T)^0$, so $\operatorname{range} T' \subseteq (\operatorname{null} T)^0$.

Instead of showing $(\operatorname{null} T)^0 \subseteq \operatorname{range} T'$ as usual, we complete the proof by showing $\operatorname{range} T'$ and $(\operatorname{null} T)^0$ have the same dimension. We have

$$\begin{aligned} \dim \operatorname{range} T' &= \dim \operatorname{range} T \\ &= \dim V - \dim \operatorname{null} T \\ &= \dim (\operatorname{null} T)^0, \end{aligned}$$

where the final equality comes from 3.62.

□

Proposition 3.66. T injective is equivalent to T' surjective
Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is injective iff T' is surjective.

Proof. The map T is injective iff $\operatorname{null} T = \{0\}$, which happens iff $(\operatorname{null} T)^0 = V'$ (taking the annihilator of both sides), which by 3.65 happens iff $\operatorname{range} T' = V'$, which happens iff T' is surjective.
□

The matrix of the dual of a linear map

Definition 3.67. Transpose, A^t
Denote the transpose of an m-by-n matrix A by the n-by-m matrix A^t given by

$$(A^t)_{k,j} = A_{j,k}.$$

Proposition 3.68. The transpose of the product of matrices
If A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^t = C^t A^t.$$

Proof. Simply calculate

$$\begin{aligned} \big((AC)^t\big)_{k,j} &= (AC)_{j,k} \\ &= \sum_{r=1}^{n} A_{j,r} C_{r,k} \\ &= \sum_{r=1}^{n} (C^t)_{k,r} (A^t)_{r,j} \\ &= (C^t A^t)_{k,j}, \end{aligned}$$

as desired. □

Proposition 3.69. The matrix of T' is the transpose of the matrix of T
Suppose $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^t$.

This is w.r.t. some basis $v_1, \ldots, v_n$ of V with its dual basis $\varphi_1, \ldots, \varphi_n$ of V', and some basis $w_1, \ldots, w_m$ of W with its dual basis $\psi_1, \ldots, \psi_m$ of W'.

Proof. Let $A = \mathcal{M}(T)$ and $C = \mathcal{M}(T')$. Suppose $1 \le j \le m$ and $1 \le k \le n$.

From the definition of $\mathcal{M}(T')$, we have

$$T'(\psi_j) = \sum_{r=1}^{n} C_{r,j}\varphi_r.$$

The LHS of the above equals $\psi \circ T$. Applying both sides of the above to v_k gives

$$(\psi_j \circ T)(v_k) = \sum_{r=1}^{n} C_{r,j}\varphi_r(v_k) = C_{k,j}.$$

But we also have

$$\begin{aligned}(\psi_j \circ T)(v_k) &= \psi_j(Tv_k) \\ &= \psi_j(\sum_{r=1}^{m} A_{r,k}w_r) \\ &= \sum_{r=1}^{m} A_{r,k}\psi_j(w_r) \\ &= A_{j,k}.\end{aligned}$$

Hence $C_{k,j} = A_{j,k}$, so $C = A^t$, as desired. □

The rank of a matrix

Definition 3.70. Row rank, column rank
Suppose $A \in \mathbb{F}^{m,n}$. The row rank of A is the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$, and similarly for column rank (though we will soon prove these to be equal).

Proposition 3.71. Dimension of $\operatorname{range} T$ equals column rank of $\mathcal{M}(T)$
Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\dim \operatorname{range} T$ equals the column rank of $\mathcal{M}(T)$. (And because $\operatorname{range} T$ does not depend on the choice of bases for $\mathcal{M}(T)$, then column rank is also the same for all bases.)

Proof. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The function that takes $w \in \operatorname{span}(Tv_1, \ldots, Tv_n)$ to $\mathcal{M}(w)$ is an isomorphism from $\operatorname{span}(Tv_1, \ldots, Tv_n) = \operatorname{range} T$ to $\operatorname{span}\big(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\big)$, and isomorphic vector spaces have the same dimension. Thus $\dim \operatorname{range} T = \dim \operatorname{span}\big(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\big)$, which equals the column rank of $\mathcal{M}(T)$. □

Proposition 3.72. Row rank equals column rank
Suppose $A \in \mathbb{F}^{m,n}$. Then the row rank of A equals the column rank of A.

Proof. Define $T : \mathbb{F}^{n,1} \to \mathbb{F}^{m,1}$ by $Tx = Ax$. Thus $\mathcal{M}(T) = A$, where $\mathcal{M}(T)$ is computed w.r.t. the standard bases of $\mathbb{F}^{n,1}$ and $\mathbb{F}^{m,1}$. Now

$$\begin{aligned}
\text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\
&= \dim \text{range } T \\
&= \dim \text{range } T' \\
&= \text{column rank of } \mathcal{M}(T') \\
&= \text{column rank of } A^t \\
&= \text{row rank of } A,
\end{aligned}$$

where the second and fourth equalities comes from the previous result and the third equality comes from 3.65. □

Definition 3.73. Rank
The rank of a matrix $A \in \mathbb{F}^{m,n}$ is the column rank of A (which equals the row rank).

4 Polynomials

Complex conjugate and absolute value

Proposition 4.1. Properties of complex numbers
Suppose $w, z \in \mathbb{C}$. Then

- Multiplicativity of complex conjugate: $\overline{wz} = \bar{w}\bar{z}$
- Triangle inequality: $|w + z| \leq |w| + |z|$

Proof. Verifying the first bullet is routine. To verify the triangle inequality, we have

$$\begin{aligned} |w+z|^2 &= (w+z)(\bar{w}+\bar{z}) \\ &= |w|^2 + |z|^2 + w\bar{z} + \overline{w\bar{z}} \\ &= |w|^2 + |z|^2 + 2\,\mathrm{Re}(w\bar{z}) \\ &\leq |w|^2 + |z|^2 + 2\,|w\bar{z}| \\ &\leq |w|^2 + |z|^2 + 2\,|w|\,|z| \\ &= (|w| + |z|)^2. \end{aligned}$$

Taking the square root of both sides gives the triangle inequality. □

Uniqueness of coefficients for polynomials

Proposition 4.2. If a polynomial is the zero function, then all coefficients are 0
Suppose $a_0, \ldots, a_m \in \mathbb{F}$. If

$$a_0 + a_1 z + \cdots + a_m z^m = 0$$

for every $z \in \mathbb{F}$, then $a_0 = \cdots = a_m = 0$. This result implies that the coefficients of a polynomial are uniquely determined (because if a polynomial had two different sets of coefficients, then subtracting the two representations would contradict this result).

Proof. We will prove the contrapositive, i.e., that if not all coefficients are 0, then a polynomial cannot be the zero function. If not all coefficients are 0, then by changing m we can assume $a_m \neq 0$. Now we will show that if $a_m \neq 0$ and $|z|$ is large enough, then $|a_m z^m| > \left|\sum_{j=0}^{m-1} a_j z^j\right|$ and therefore $-a_m z^m \neq \sum_{j=0}^{m-1} a_j z^j$.

To construct z with sufficiently big absolute value, we first assume $|z| \geq 1$ s.t. each $\left|z^j\right|$ is greater than the lower $\left|z^k\right|$. (In general in analysis, this is a common start to a "there exists a sufficiently large thing" proof: pick some initial threshold such that if the thing is bigger than that threshold, the situation as a whole becomes simpler.)

Under this assumption, we now apply the triangle inequality:

$$\left|\sum_{j=0}^{m-1} a_j z^j\right| \leq \sum_{j=0}^{m-1} \left|a_j z^j\right| \leq \left|z^{m-1}\right| \sum_{j=0}^{m-1} |a_j|\,.$$

Thus to show our desired inequality, it suffices to have $|a_m z_m| > \left|z^{m-1}\right| \sum_{j=0}^{m-1} |a_j|$, for which it suffices to have

$$|z| > \frac{1}{|a_m|} \sum_{j=0}^{m-1} |a_j| .$$

Thus if $z \geq 1$ and satisfies the above inequality, then $|a_m z^m| > \left|\sum_{j=0}^{m-1} a_j z^j\right|$, as desired. Such z may be constructed by

$$z = \frac{|1| + \cdots + \left|z^{m-1}\right|}{|a_m|} + 1.$$

For this z, we will have $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} \neq -a_m z^m$, and hence we conclude $a_0 + a_1 z + \cdots + a_m z^m \neq 0$.

Note: The Axler text provides this proof in cleaner style, first writing down the construction of z (without proper explanation) and then checking that it works. Here, the proof has been rewritten in poorer style to better show the proof might be developed from scratch. □

The division algorithm for polynomials

Axler provides an interesting linear algebra proof of the division algorithm.

Zeros of polynomials

Proposition 4.3. A polynomial has at most as many zeros as its degree
Suppose $p \in \mathcal{P}(\mathbb{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbb{F}$.

Proof. We use induction on m.

If $m = 0$, then $p(z) = a_0 \neq 0$ and so p has no zeros.

If $m = 1$, then $p(z) = a_0 + a_1 z$ with $a_1 \neq 0$, and thus p has exactly one zero, namely, $-a_0/a_1$.

Now suppose $m > 1$. By induction on m, we assume that every polynomial with degree $m - 1$ has at most $m - 1$ distinct zeros. If p has no zeros in $\mathbb{F}$, we are done. If p has a zero $\lambda \in \mathbb{F}$, then there is a polynomial q s.t.

$$p(z) = (z - \lambda) q(z)$$

for all $z \in \mathbb{F}$. Clearly $\deg q = m - 1$. The zeros of p consist of λ and the zeros of q, and by our inductive hypothesis, q has at most $m - 1$ distinct zeros in $\mathbb{F}$. Thus p has at most m distinct zeros in $\mathbb{F}$, as desired. □

Factorization of polynomials over $\mathbb{C}$

Previously, we simultaneously addressed polynomials with either complex coefficients or real coefficients (using our convention that $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$). Now these two following subsubsections will address some differences between the two cases.

Theorem 4.4. Fundamental theorem of algebra
Every nonconstant polynomial with complex coefficients has a zero.

Proof. The proof is excluded because it requires analysis. □

Proposition 4.5. Factorization of a polynomial over $\mathbb{C}$
If $p \in \mathcal{P}(\mathbb{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form

$$p(z) = c(z - \lambda_1) \dots (z - \lambda_m)$$

where $c, \lambda_1, \dots, \lambda_m \in \mathbb{C}$.

Proof. Let $m = \deg p$. We use induction on m. Clearly there exists a unique factorization if $m = 1$.

For $m > 1$, assume the desired factorization exists and is unique for all polynomials of degree $m - 1$. To show existence of the desired factorization of p, let λ be a zero of p by the fundamental theorem of algebra. Thus $p(z) = (z - \lambda)q(z)$ for all $z \in \mathbb{C}$. Because $\deg q = m - 1$, our induction hypothesis implies that q has the desired factorization, and thus so does p.

(Note from self: We have cannot argue that this factorization of p is unique from our inductive hypothesis that the factorization of q is unique. The factorization of q is indeed unique, but only after q is determined. The polynomial q is determined by the choice of λ, which is not a unique choice.)

Now we show uniqueness of this representation. Clearly c is uniquely determined by the coefficient of z^m in p. To show there is only one way to choose the zeros, suppose

$$(z - \lambda_1) \dots (z - \lambda_m) = (z - \tau_1) \dots (z - \tau_m)$$

for all $z \in \mathbb{C}$. Then when $z = \lambda_1$, the LHS equals 0, which implies that one of the τ's equals 0. We relabel to assume $\lambda_1 = \tau_1$. Now for $z \neq \lambda_1$, we may divide by $(z - \lambda_1$ to get

$$(z - \lambda_2) \dots (z - \lambda_m) = (z - \tau_2) \dots (z - \tau_m)$$

for all $z \in \mathbb{C}$ except possibly $z = \lambda_1$. But if the equation does not hold for $z = \lambda_1$, then subtracting the RHS from the LHS would give us a nonzero polynomial that has infinitely many zeros (a contradiction). Thus the equation holds for all $z \in \mathbb{C}$, implying the polynomials on the LHS and RHS are the same. Thus the factorization of both sides is unique by our inductive hypothesis (alternatively, we might repeat the relabeling process to show that each $\lambda_j = \tau_j$), completing the proof of uniqueness. □

Factorization of polynomials over $\mathbb{R}$

Proposition 4.6. Polynomials with real coefficients have zeros in pairs
Suppose $p \in \mathcal{P}(\mathbb{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbb{C}$ is a zero of p, then so is $\bar{\lambda}$.

Proof. Let

$$p(z) = a_0 + a_1 z \cdots + a_m z^m,$$

where $a_0, \dots, a_m \in \mathbb{R}$. Suppose $\lambda \in \mathbb{C}$ is a zero of p. Then

$$a_0 + a_1 \lambda \cdots + a_m \lambda^m = 0.$$

Taking the complex conjugate of both sides, we get

$$a_0 + a_1 \bar{\lambda} \cdots + a_m \bar{\lambda}^m,$$

showing that $\bar{\lambda}$. (Note that $(\bar{\lambda})^m = \overline{\lambda^m}$; this is obvious using Euler's formula.) □

5 Eigenvalues, Eigenvectors, and Invariant Subspaces

Notation 5.1. V
Standing notation throughout this section:

- V denotes a vector space over $\mathbb{F}$.

5.A Invariant subspaces

Eigenvalues and Eigenvectors

Definition 5.2. Invariant subspace
Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called invariant under T if $u \in U$ implies $Tu \in U$. In other words, U is invariant under T if $T|_U$ is an operator on U.

Now consider one-dimensional invariant subspaces. These are so special that we define the following.

Definition 5.3. Eigenvalue
Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an eigenvalue of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda\nu$.

Proposition 5.4. Equivalent conditions to be an eigenvalue
Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

1. λ is an eigenvalue of T;
2. $T - \lambda I$ is not injective;
3. $T - \lambda I$ is not surjective;
4. $T - \lambda I$ is not invertible.

Definition 5.5. Eigenvector
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an eigenvector of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Proposition 5.6. Linearly independent eigenvectors
Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding eigenvectors. Then $v_1, \ldots, v_m$ is linearly independent.

Proof. Suppose $v_1, \ldots, v_m$ is linearly dependent. Let k be the smallest positive integer such that

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$$

the existence of k with this property follows from the Linear Dependence Lemma). Thus there exist $a_1, \ldots, a_{k-1} \in \mathbf{F}$ such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}.$$

Apply T to both sides of this equation, getting

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}.$$

Multiply both sides of the first equation by λ_k and then subtract the second equation above, getting

$$0 = a_1 (\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1} (\lambda_k - \lambda_{k-1}) v_{k-1}.$$

By construction, $v_1, \ldots, v_{k-1}$ is linearly independent. Thus the equation above implies that all the a's are 0 (recall that λ_k is not equal to any of $\lambda_1, \ldots, \lambda_{k-1}$). However, this means that v_k equals 0, contradicting our hypothesis that v_k is an eigenvector. Therefore our assumption that $v_1, \ldots, v_m$ is linearly dependent was false. □

Proposition 5.7. Number of eigenvalues
Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Restriction and Quotient Operators

Definition 5.8. $T|_U$ and T/U
Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T. Then U determines two other operators $T|_U \in \mathcal{L}(U)$ and $T/U \in \mathcal{L}(V/U)$ in a natural way:

- The restriction operator $T|_U \in \mathcal{L}(U)$ is defined by

  $$T|_U (u) = Tu$$

  for $u \in U$.

- The quotient operator $T/U \in \mathcal{L}(V/U)$ is defined by

  $$(T/U)(v + U) = Tv + U$$

  for $v \in V$.

Note that sometimes $T|_U$ and T/U do not provide enough information about T. In the next example, both $T|_U$ and T/U are 0 even though T is not the 0 operator.

5.B Eigenvectors and Upper-Triangular Matrices

Definition 5.9. $p(T)$
Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for $z \in \mathbf{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

Note that for a fixed operator $T \in \mathcal{L}(V)$, the function from $\mathcal{P}(\mathbf{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear.

Proposition 5.10. Multiplicative properties
Suppose $p, q \in \mathcal{P}(\mathbf{F})$ and $T \in \mathcal{L}(V)$. Then

(a) $(pq)(T) = p(T)q(T)$;

(b) $p(T)q(T) = q(T)p(T)$.

Existence of Eigenvalues

Proposition 5.11. Operators on complex vector spaces have an eigenvalue
Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Proof. Suppose V is a complex vector space with dimension $n > 0$ and $T \in \mathcal{L}(V)$. Choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2 v, \ldots, T^n v$$

is not linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist complex numbers $a_0, \ldots, a_n$, not all 0 , such that

$$0 = a_0 v + a_1 Tv + \cdots + a_n T^n v.$$

Note that $a_1, \ldots, a_n$ cannot all be 0 , because otherwise the equation above would become $0 = a_0 v$, which would force a_0 also to be 0.

Make the a's the coefficients of a polynomial, which by the Fundamental Theorem of Algebra has a factorization

$$a_0 + a_1 z + \cdots + a_n z^n = c\,(z - \lambda_1) \cdots (z - \lambda_m),$$

where c is a nonzero complex number, each λ_j is in $\mathbf{C}$, and the equation holds for all $z \in \mathbf{C}$ (here m is not necessarily equal to n, because a_n may equal 0). We then have

$$\begin{aligned} 0 &= a_0 v + a_1 Tv + \cdots + a_n T^n v \\ &= \left(a_0 I + a_1 T + \cdots + a_n T^n\right) v \\ &= c\,(T - \lambda_1 I) \cdots (T - \lambda_m I)\, v. \end{aligned}$$

Thus $T - \lambda_j I$ is not injective for at least one j. In other words, T has an eigenvalue. □

Upper-Triangular Matrices

Proposition 5.12. Conditions for upper-triangular matrix
Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent: (a) the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular; (b) $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$; (c) $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$.

Proposition 5.13. Over C, every operator has an upper-triangular matrix
Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

Proof. We will use induction on the dimension of V. Clearly the desired result holds if $\dim V = 1$.

Suppose now that $\dim V = n > 1$ and the desired result holds for all complex vector spaces whose dimension is $n-1$. Let v_1 be any eigenvector of T (5.21 guarantees that T has an eigenvector). Let $U = \operatorname{span}(v_1)$. Then U is an invariant subspace of T and $\dim U = 1$.

Because $\dim V/U = n - 1$, we can apply our induction hypothesis to $T/U \in \mathcal{L}(V/U)$. Thus there is a basis $v_2 + U, \ldots, v_n + U$ of V/U such that T/U has an upper-triangular matrix with respect to this basis. Hence,

$$(T/U)(v_j + U) \in \operatorname{span}(v_2 + U, \ldots, v_j + U)$$

for each $j = 2, \ldots, n$. Unraveling the meaning of the inclusion above, we see that

$$Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$$

for each $j = 1, \ldots, n$. Thus by $5.26, T$ has an upper-triangular matrix with respect to the basis $v_1, \ldots, v_n$ of V, as desired (it is also easy to verify that $v_1, \ldots, v_n$ is a basis of V). □

Proposition 5.14. Determination of invertibility from upper-triangular matrix
Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Proposition 5.15. Determination of eigenvalues from upper-triangular matrix
Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

Proof. Suppose $v_1, \ldots, v_n$ is a basis of V with respect to which T has an upper-triangular matrix with values $\lambda_1, \ldots, \lambda_n$ along the diagonal.

Now suppose $\lambda \in \mathbf{F}$. Then $T - \lambda I$ is not invertible if and only if λ equals one of the numbers $\lambda_1, \ldots, \lambda_n$ by the previous result. Thus λ is an eigenvalue of T if and only if λ equals one of the numbers $\lambda_1, \ldots, \lambda_n$ □

5.C Eigenspaces and Diagonal Matrices

Definition 5.16. Eigenspace, $E(\lambda, T)$
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The eigenspace of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \text{null}(T - \lambda I).$$

In other words, $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

Proposition 5.17. Sum of eigenspaces is a direct sum
Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proposition 5.18. Conditions equivalent to diagonalizability
Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

(a) T is diagonalizable;

(b) V has a basis consisting of eigenvectors of T;

(c) there exist 1-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that

$$V = U_1 \oplus \cdots \oplus U_n$$

(d) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;

(e) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

6 Inner Product Spaces

Notation 6.1. $\mathbb{F}, V$
Standing notation throughout this chapter:

- $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$
- V denotes an inner product space over $\mathbb{F}$

6.A Inner product and norms

Definition 6.2. Inner product, inner product space
An inner product on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle v, w\rangle \in \mathbb{F}$ and has the following properties:

- Positivity: $\langle v, v\rangle \geq 0$;
- Definiteness: $\langle v, v\rangle = 0$ iff $v = 0$;
- Linearity in the first slot
- Conjugate symmetry: $\langle u, v\rangle = \overline{\langle v, u\rangle}$

This is simply symmetry for real vector spaces, and conjugate symmetry also implies conjugate linearity in the second slot.

Also define an inner product space as a vector space V along with an inner product on V.

Proposition 6.3. Basic properties of an inner product
The following hold:

(a) For each fixed $u \in V$, the function that takes v to $\langle v, u\rangle$ is a linear map from V to $\mathbb{F}$.

(b) $\langle 0, u\rangle = 0$ for every $u \in V$.

(c) $\langle u, 0\rangle = 0$ for every $u \in V$.

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

(e) $\langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle$ for all $\lambda \in \mathbb{F}$ and $u, v \in V$.

Definition 6.4. Norm
For $v \in V$, the norm $\|v\|$ of v is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Definition 6.5. Orthogonal
Two vectors $u, v \in V$ are called orthogonal if

$$\langle u, v \rangle = 0.$$

Theorem 6.6. Pythagorean Theorem
Suppose u and v are orthogonal vectors in V. Then

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

6.B Orthonormal bases

Proposition 6.7. Norm of an orthonormal linear combination
If $e_1, \ldots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \ldots, a_m \in \mathbb{F}$.

Proof. Follows from repeated application of Pythagorean theorem (6.6) and the fact that each e_j has norm 1. □

Proposition 6.8. Orthonormal list is linearly independent
Every orthonormal list of vectors is linearly independent. This implies that every orthonormal list of length $\dim V$ is a basis of V.

Proof. Supposing

$$a_1 e_1 + \cdots + a_m e_m = 0,$$

we have that the norm of either side is 0. Applying the previous result yields $a_1 = \cdots = a_m = 0$, as desired. □

Proposition 6.9. Writing a vector as linear combination of orthonormal basis
Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 .$$

Proof. After writing

$$v = a_1 e_1 + \ldots a_n e_n,$$

it is easy to see that $\langle v, e_j \rangle = a_j$. Then both equations follow easily. □

Proposition 6.10. Existence of orthonormal basis
Every finite-dimensional inner product space has an orthonormal basis.

We may simply choose a basis $v_1, \ldots, v_n$ of V and apply Gram-Schmidt, i.e., subtract from each vector its orthogonal projections onto all the vectors preceding it:

$$u_1 = v_1, u_i = v_i - \sum_{k=1}^{i-1} \text{proj}_{u_k} v_i.$$

Remark 6.11.
The guaranteed existence of an orthonormal basis implies that every n-dimensional inner product space (over $\mathbb{R}$) is isomorphic to $\mathbb{R}^n$ with the standard dot product, i.e., there is a linear isomorphism $V \to \mathbb{R}^n$ that carries the given inner product on V to the dot product on $\mathbb{R}^n$.

Theorem 6.12. Riesz Representation Theorem
Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

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

for every $v \in V$.

Proof. First we show there exists a vector $u \in V$ such that $\varphi(v) = \langle v, u \rangle$ for every $v \in V$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned} \varphi(v) &= \varphi\left(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n\right) \\ &= \langle v, e_1 \rangle \varphi(e_1) + \cdots + \langle v, e_n \rangle \varphi(e_n) \\ &= \left\langle v, \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \right\rangle \end{aligned}$$

for every $v \in V$. Thus setting

$$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n,$$

we have $\varphi(v) = \langle v, u \rangle$ for every $v \in V$, as desired. Now we prove that only one vector $u \in V$ has the desired behavior. Suppose $u_1, u_2 \in V$ are such that

$$\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$. In other words, $u_1 = u_2$, completing the proof of the uniqueness part of the result. □

6.C Orthogonal complements and minimization problems

Definition 6.13. Orthogonal complement, $U^\perp$
If U is a subset of V, then the orthogonal complement $U^\perp$ of U is the subspace of all vectors in V that are orthogonal to every vector in U:

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

Proposition 6.14. Direct sum of a subspace and its orthogonal complement
Suppose U is a finite-dimensional subspace of V. Then

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

This does not hold for bilinear forms $b \in B(V)$ in general, because we cannot guarantee $U \cap U^\perp = \{0\}$ since there may exist nonzero $v \in V$ such that $b(v, v) = 0$.

Proof. First we will show that

$$V = U + U^\perp.$$

To do this, suppose $v \in V$. Let $e_1, \ldots, e_m$ be an orthonormal basis of U. We may write the obvious equality

$$v = \underbrace{\langle v, e_1\rangle\, e_1 + \cdots + \langle v, e_m\rangle\, e_m}_{u} + \underbrace{v - \langle v, e_1\rangle\, e_1 - \cdots - \langle v, e_m\rangle\, e_m}_{w},$$

and let u and w be defined as above. Clearly $u \in U$. Because $e_1, \ldots, e_m$ is an orthonormal list, for each $j = 1, \ldots, m$ we have

$$\langle w, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0.$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$. In other words, $w \in U^\perp$. Thus we have written $v = u + w$, where $u \in U$ and $w \in U^\perp$, completing the proof that $V = U + U^\perp$.

To show uniqueness of the representation, observe that $U \cap U^\perp = \{0\}$, as desired. □

Proposition 6.15. Dimension of the orthogonal complement
Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^\perp = \dim V,$$

and this conclusion holds for nondegenerate bilinear forms in general. Nondegenerate bilinear forms are $b \in B(V)$ such that $\ker(b) = 0$ with

$$\ker(b) \coloneqq \{v \in V : b(v, w) = 0 \ \forall w \in V\}.$$

Proposition 6.16. The orthogonal complement of the orthogonal complement
Suppose U is a finite-dimensional subspace of V. Then

$$U = \left(U^\perp\right)^\perp.$$

Proof. First we will show that

$$U \subset \left(U^\perp\right)^\perp.$$

To do this, suppose $u \in U$. Then $\langle u, v\rangle = 0$ for every $v \in U^\perp$ (by the definition of $U^\perp$). Because u is orthogonal to every vector in $U^\perp$, we have $u \in \left(U^\perp\right)^\perp$, as desired.

To prove the inclusion in the other direction, suppose $v \in \left(U^\perp\right)^\perp$. By 6.14, we can write $v = u + w$, where $u \in U$ and $w \in U^\perp$. We have $v - u = w \in U^\perp$. Because $v \in \left(U^\perp\right)^\perp$ and $u \in \left(U^\perp\right)^\perp$, we have $v - u \in \left(U^\perp\right)^\perp$. Thus $v - u \in U^\perp \cap \left(U^\perp\right)^\perp$, which implies that $v - u$ is orthogonal to itself, which implies that $v - u = 0$, which implies that $v = u$, which implies that $v \in U$. Thus $\left(U^\perp\right)^\perp \subset U$, completing the proof. □

Definition 6.17. Definition orthogonal projection, P_U
Suppose U is a finite-dimensional subspace of V. The orthogonal projection of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $P_U v = u$.

Note that the direct sum decomposition $V = U \oplus U^\perp$ from above shows that each $v \in V$ can be uniquely written in the form $v = u + w$ with $u \in U$ and $w \in U^\perp$. Thus $P_U v$ is well defined.

Proposition 6.18. Properties of the orthogonal projection P_U
Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

(a) $P_U \in \mathcal{L}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) range $P_U = U$;

(e) null $P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$;

(g) $P_U^2 = P_U$;

(h) $\|P_U v\| \leq \|v\|$;

(i) for every orthonormal basis $e_1, \ldots, e_m$ of U,

$$P_U v = \langle v, e_1\rangle\, e_1 + \cdots + \langle v, e_m\rangle\, e_m.$$

Proposition 6.19. Minimizing the distance to a subspace
Suppose U is a finite-dimensional subspace of $V, v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \leq \|v - u\|.$$

Furthermore, the inequality above is an equality if and only if $u = P_U v$.

Proof. We have

$$\begin{aligned}\|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2\end{aligned}$$

where the first line above holds because $0 \leq \|P_U v - u\|^2$, the second line above comes from the Pythagorean Theorem [which applies because $v - P_U v \in U^\perp$ by property (f) from above, and $P_U v - u \in U$], and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

Our inequality above is an equality iff the first line is an equality, which happens if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$. □

A Proof Preliminaries

First, some common types of proof problems with examples.

A.1 Prove that either statement X holds or statement Y holds

Consider either casework or proof by contradiction.

Example A.1. Axler, page 17, exercise 1.B.2
Suppose $a \in \mathbb{F}, v \in V$, and $av = 0$. Prove that $a = 0$ or $v = 0$.

Proof. Let us perform casework on a with two cases, $a = 0$ or $a \neq 0$.
If $a = 0$, we are done.
If $a \neq 0$, then a has multiplicative inverse a^{-1}. Hence,

$$v = 1 \cdot v = (a^{-1}a)v = a^{-1} \cdot 0 = 0.$$

□

A.2 Prove there exists a unique x satisfying some condition Y

First prove existence, and then prove uniqueness.

Example A.2. Axler, page 17, exercise 1.B.3
Suppose $v, w \in V$. Explain why there exists a unique $x \in V$ s.t. $v + 3x = w$.

Proof. Let $x = \frac{1}{3}(u - v)$. Then,

$$v + 3x = v + (w - v) = v.$$

This shows existence. Now we show uniqueness. Suppose we have another vector x' such that $v + 3x' = w$. Similarly, $v + 3x = w$. Hence, subtracting we have

$$3x - 3x' = 3(x - x') = 0.$$

Thus, we must have $x - x' = 0$, i.e., $x = x'$. This shows uniqueness. □

A.3 Prove set X equals set Y

A standard proof is to show $X \subseteq Y$ and $Y \subseteq X$. Of course, there are many variations of this question statement and this proof.

In the case that X and Y are subspaces of the same vector space, one might show $X \subset Y$ and $\dim X = \dim Y$. For example, see 3.65.

Also consider this related question: Prove X is the smallest subspace containing $x_1, \ldots, x_m$. In this specific example, of course, first show that X is a subspace. Then to show X is the smallest one containing $x_1, \ldots, x_m$, show that (1) X contains $x_1, \ldots, x_m$ (which is usually easy), and then (2) every subspace containing $(x_1, \ldots, x_m)$ must contain X. For example, see 2.4.

A.4 Using induction

Examples of standard induction proofs with inductive hypotheses are 4.3 and 4.5. For an example of using induction to construct a list with some desired properties, see 2.12.

A.5 Proving statements are equivalent

To prove two statements are equivalent, a standard method is to prove the forward direction and then the converse. Another possibility is using a chain of iff statements, i.e., show that statement 1 happens iff x, which happens iff $y, \ldots$, which happens iff statement 2. For example, see 3.64.

To prove many statements are equivalent, an option is to show that statement (a) implies (b), (b) implies (c), and so on. For example, see 3.40.

A.6 Using the contrapositive

For example, see 4.2 (though this is closer to an analysis proof than linear algebra).

A.7 Analysis: There exists a sufficiently large thing

In analysis, a common start to this proof is to pick some initial threshold such that if the thing is bigger than that threshold, the situation as a whole becomes simpler. For example, see 4.2.