Linear Algebra

Patrick Oare

This will begin with an introductory section about linear algebra, which is the study of **vector spaces** and the maps between them (linear transformations). Vector spaces actually encode a relatively large amount of structure, and I'll be careful to point out what properties of vector spaces generalize to other objects. Vector spaces actually encode a large amount of structure, much more than simpler algebraic objects like groups, rings & fields, and modules; however, they are most people's entryway into the study of algebra because they are so important for calculations in the physical sciences. So, hopefully most of this chapter is a review (but it is OK if it is not!). This section will be written to emphasize properties of interest that vector spaces share with other algebraic objects: hopefully this will make the transition to studying these other objects much easier, and make it clear why we will make the definitions that we make in the future.

Definition 1: Vector Space

A vector space over $\mathbb{R}$ is a triple $(V, +, \cdot)$, where V is a set, often just called the vector space, $+ : V \times V \to V$ is a binary operation called addition, and $\cdot : \mathbb{R} \times V \to V$ is called scalar multiplication. An element of V is called a **vector**, and an element in $\mathbb{R}$ is called a **scalar**. Addition and scalar multiplication satisfy the following properties[a]:

1. (Commutativity of $+$) For all $u, v \in V$, $u + v = v + u$.
2. (Associativity of $+$) For all $u, v, w \in V$, $(u + v) + w = u + (v + w)$.
3. (Zero element) There is an element $0 \in V$ such that $0 + v = v$ for each $v \in V$.
4. (Additive inverse) For each $v \in V$, there is a $-v \in V$ such that $v + (-v) = 0$.
5. (Distributivity of $\cdot$ for vector addition) For $c \in \mathbb{R}$ and $u, v \in V$, $c(u+v) = cu+cv$.
6. (Distributivity of $\cdot$ for scalar addition) For $c, d \in \mathbb{R}$ and $v \in V$, $(c+d)v = cv+dv$.
7. (Associativity of $\cdot$) For all $c, d \in \mathbb{R}$ and $v \in V$, $(cd)v = c(dv)$.
8. (Identity multiplication) For each $v \in V$, $1v = v$.

[a] Note that many books often (especially those written for non-mathematicians) often say "a vector space is closed under addition and scalar multiplication". While this is important to note, it is implied by assuming the addition and scalar multiplication operations have codomain V. This is a pet peeve of mine that will show again when we consider groups; it is not a postulate or axiom that an object is closed under an operation.

These 8 axioms define how vector space multiplication and addition work, and from these you can derive all other properties that you expect these operations to satisfy. For example, $0 \in V$ is only defined via its action as $0 + v = v$, nothing more. However, there are a number of other properties of $0 \in V$ that we want to show and that we assume should hold!

- We expect $0 \in V$ is unique. Suppose it was not, and $0' \in V$ satisfies (3). Then $0 + 0' = 0$ from (3) applied to $v = 0'$, but also $0 + 0' = 0' + 0 = 0'$ from (3) applied to $v = 0$, hence we see $0 = 0'$. Thus $0 \in V$ is the unique element satisfying (3).

- For any $v \in V$, we expect the additive inverse $-v$ to be unique. Suppose it is not, and $-v'$ is another additive inverse. Then $-v = -v + 0 = -v + (v + (-v')) = (-v + v) + (-v') = 0 + (-v') = -v'$. A corollary of this is that if $v + w = v + u$, then we can cancel v from both sides of the equation, and it must be that $w = u$.

- We expect $0v = 0$ for any $v \in V$, and that $c0 = 0$ for any scalar $c \in \mathbb{R}$. This must be true because for any c,
$$cv = (c + 0)v = cv + 0v. \tag{1}$$
Adding $-(cv)$ to each side implies $0 = 0v$ as expected. The other side also goes through for the same reason, because $cv = c(v + 0) = cv + c0 \implies c0 = 0$. Note for both of these small proofs, we require 0 to be unique, so it's good we showed that first!

- We expect $c(-v) = -cv$. To see this, we show $c(-v)$ satisfies (4),
$$c(-v) + cv = c(-v + v) = c0 = 0. \tag{2}$$

- We expect $-1(v) = -v$. To show this, $v + (-1)v = 1v + (-1)v = (1 - 1)v = 0v = 0$, hence because $-v$ is unique, this implies $-1v = -v$.

You can likewise prove a large number of other properties if interested. Note how much all these properties build on one another; so many simple arithmetic tricks are reliant on the basic 8 axioms presented above, and so much of what we take for granted is constructed this way.

0.0.1 Subspaces and Bases

We next turn to subspaces.

Definition 2: Subspace

A **subspace** of a vector space is a subset of V which is, in itself, a vector space under the operations $(+, \cdot)$. We denote that $W \subseteq V$ is a subspace with $W \leq V$.

Note that since W inherits the addition and scalar multiplication operations from V, one doesn't need to prove that $(W, +, \cdot)$ satisfies the vector space axioms. Instead, we only need to show that W is closed under addition and scalar multiplication, and contains the zero vector. In fact, this is a sufficient condition to show that W is a subspace of V.

Theorem 1: Subspace criterion

Let $W \subseteq V$. W is a subspace of V with respect to $(+, \cdot)$ iff:

1. (Additive identity) $0 \in W$.
2. (W is closed under addition) For each $v, w \in W$, $v + w \in W$.
3. (W is closed under scalar multiplication) For each $c \in \mathbb{R}$ and $w \in W$, $vw \in W$.

Proof. The forward direction is trivial, and the only thing to check in the backwards direction is that for each $w \in W$, $-w \in W$ (everything else holds because $+$ and $\cdot$ satisfy the other axioms on V, and therefore on $W \subseteq V$). This is clear because W is closed under scalar multiplication, so $-1w = -w \in W$. $\square$

The easiest way to visualize a subspace is as a hyperplane embedded in V. Any hyperplane of any dimension containing the origin is a subspace, and vice versa; we will discuss what we mean by "dimension" next.

The fundamental building block of a vector space is a **basis** for a vector space. To define a basis, we need the notion of linear independence and the span of a set of vectors.

Definition 3: Linear combination, Linear (in)dependence

A **linear combination** of k vectors $\{v_1, ..., v_k\}$ is a sum of the form $c_1v_1 + ... + c_kv_k$. We say the set $\{v_1, ..., v_k\}$ is **linearly dependent** if there exists a non-trivial linear combination ($c_i \neq 0$ for at least one i) of $\{v_i\}$ equaling 0,

$$c_1v_1 + ... + c_kv_k = 0. \tag{3}$$

If $\{v_i\}$ is not linearly dependent, then it is linearly independent.

In other words, the only linear combination of linearly independent vectors that is 0 is the trivial combination, $c_1 = ... = c_k = 0$. Note that if the zero vector is included in a set, it is automatically linearly dependent; the zero vector is linearly dependent with every other vector. Linearly independent vectors all "point" in different directions of the vector space, which is formalized by defining the **span** of a set of vectors.

Definition 4: Span

The **span** of a set of vectors $\{v_1, ..., v_k\}$ is the space of all linear combinations of these vectors,

$$\text{span}\,\{v_1, ..., v_k\} = \left\{ \sum_{i=1}^{k} c_iv_i \in V : \forall i \in \{1, ..., k\}, c_i \in \mathbb{R} \right\} \tag{4}$$

The span of a set of vectors is the set of all linear combinations of these vectors. One can trivially show that the span of any set of vectors is a subspace of V:

Proposition 1

Let $\{v_1, ..., v_k\} \subset V$. Then $\text{span}\{v_1, ..., v_k\}$ is a subspace of V. Furthermore, any subspace $W \leq V$ containing these vectors, $\{v_1, ..., v_k\} \subseteq W$, also contains the span of these vectors, $\text{span}\{v_1, ..., v_k\} \leq W$.

The span operation allows us to generate subspaces of a vector space from individual sets of vectors. It is also **universal** in the sense that the span of a set of vectors is the *smallest* vector space containing these vectors; any other subspace containing these vectors must also contain their span.

Vector spaces are built from linear combinations, and the span operation and the notion of linear independence are the two most essential concepts to understand to formalize an intuitive picture of a vector space. They are deeply connected: the span of a set of vectors tells us about the maximal number of vectors in the set which are linearly independent.

Proposition 2

Suppose $\{v_1, ..., v_k\}$ is linearly independent, but $\{v_1, ..., v_k, v_{k+1}\}$ is linearly dependent. Then $v_{k+1} \in \text{span}\{v_1, ..., v_k\}$, and furthermore $\text{span}\{v_1, ..., v_k\} = \text{span}\{v_1, ..., v_k, v_{k+1}\}$.

Proof. We begin with the stated setup. In this case, there exist non-trivial c_i such that $\sum_{i=0}^{k+1} c_i v_i = 0$. Note that $c_{k+1} \neq 0$, because if instead $c_{k+1} = 0$, then this would imply that the vectors $\{v_1, ..., v_k\}$ were linearly independent, as there would be a non-trivial linear combination of them equalling zero. Thus, we can divide by c_{k+1},

$$v_{k+1} = \frac{1}{c_{k+1}} \sum_{i=1}^{k} c_i v_i = \sum_{i=1}^{k} \frac{c_i}{c_{k+1}} v_i \in \text{span}\{v_1, ..., v_k\}, \tag{5}$$

which proves the first claim. The second claim is also immediate, because for $w \in \text{span}\{v_1, ..., v_{k+1}\}$, we have:

$$w = \sum_{i=1}^{k+1} d_i v_i = \sum_{i=1}^{k} d_i v_i + d_{k+1} v_{k+1} = \sum_{i=1}^{k} \left(d_i + d_{k+1} \frac{c_i}{c_{k+1}} \right) v_i \in \text{span}\{v_1, ..., v_k\}. \tag{6}$$

□

That is, if a set of vectors $\{v_1, ..., v_k\}$ is linearly independent and adding another vector v_{k+1} to the set makes the set linearly dependent, then v_{k+1} is in the span of the other vectors, and adding it to the original set does not change the span. The only way we add additional information to the generated subspace (and enlarge it) is to add linearly independent vectors to the subspace.

Definition 5: Basis, Dimension

A **basis** for a vector space V is a set of vectors $\{v_1, ..., v_n\} \subset V$ that is linearly independent and spans V. If $\{v_1, ..., v_n\}$, we call $n = |\{v_1, ..., v_n\}|$ the **dimension** of V, denoted $n = \dim V$.

Note that adding any additional vector to a basis would make the set linearly dependent: suppose $\{v_1, ..., v_n\}$ is a basis for V, and $v \in V$ is non-zero (if $v = 0$ then it trivially makes the basis linearly dependent). Then because $\{v_i\}$ spans V, we can expand $v = \sum_{i=1}^n c_i v_i$, hence $v - \sum_{i=1}^n c_i v_i = 0$ is a non-trivial linear combination yielding 0. Thus, a basis is a maximal linearly independent set inside of V. In fact, a useful property of a basis is that one can *uniquely* expand any $v \in V$ as a linear combination of basis elements.

Theorem 2

Let $\{v_i\}_{i=1}^n$ be a basis for V. Then, for any $v \in V$ there exist unique scalars $c_i \in \mathbb{R}$ such that

$$v = \sum_{i=1}^n c_i v_i. \tag{7}$$

The scalars (c_i) are called the **coefficients** of v with respect to $\{v_i\}$, and are often written as a column vector.

Proof. The $v = 0$ case is immediate because the coefficients $c_i = 0$ are unique by the linear independence of the basis. For $v \in V \backslash \{0\}$, then if this was not the case, $v = \sum_i c_i v_i = \sum_i c_i' v_i'$ with at least some $c_i \neq c_i'$. Thus, the coefficients $c_i - c_i'$ are not all zero, yet $\sum_i (c_i - c_i') v_i = 0$, which contradicts the linear independence of the basis. □

There are a number of other properties that we will not prove, but may be fun to do on your own.

Proposition 3

Every spanning set for a vector space V can be reduced to a basis for V, and likewise every linearly independent set in V can be extended to a basis for V. Additionally, the length of any linearly independent set of vectors is less than or equal to the list of any spanning set of vectors.

Theorem 3

Every finite-dimensional vector space has a basis.

Note that Proposition 4 implies that any basis for a vector space has the same cardinality, as a basis is both a linearly independent set and a spanning set. This has the corollary that the dimension of a vector space is a well-defined concept! If $n = \dim V$, then the only n-dimensional subspace of V is V itself, and the only 0-dimensional subspace of V is the zero subspace $\{0\}$.

Interlude 1: Does every vector space have a basis?

You'll notice that in the previous theorem, Thm 3, we specifically mentioned that every **finite**-dimensional vector space has a basis. The qualifier "finite" here is extremely important: do all vector spaces have bases, including infinite-dimensional ones? The answer is not so easy... TODO finish

The final thing we will discuss this section is generation of subspaces. Instead of specifying a subspace with set builder notation, we will instead often just specify a basis for the subspace: this gives equivalent information in a much simpler way. Let's let $\{v_i\}_{i=1}^n$ be some set of linearly independent vectors in an ambient vector space V. Because $\{v_i\}$ is linearly independent, it **generates** a subspace W of V,

$$W = \text{span}\{v_1, ..., v_n\} \leq V. \tag{8}$$

This subspace W has dimension n because $\{v_i\}_{i=1}^n$ is a canonical basis for W.

We'll see the concept of generation a lot. By themselves, the n vectors $\{v_1, ..., v_n\}$ are not a vector space: they're just individual vectors with no additional structure. We *endow* these n vectors with a vector space structure by *generating* a subspace using these vectors: this subspace W is the **best** (smallest) subspace containing the vectors $\{v_i\}_{i=1}^n$.

How do we quantify that W is the best subspace containing the $\{v_i\}$? Suppose that W' is *any other subspace* of V that contains the n vectors $\{v_1, ..., v_n\}$. Then W' *must also contain all of* W,

$$W \leq W'. \tag{9}$$

This is a universal property of generating sets: the generated set can be embedded into any other subspace that contains the vectors $\{v_i\}$.

0.0.2 Linear Transformations

Let's discuss maps on vector spaces. In particular, let's discuss a particular type of map which preserves linear structure that I'm sure you're familiar with: the *linear transformation*.

Definition 6: Linear Transformation

Let V and W be vector spaces. A **linear transformation**, or linear map, between V and W is a map $T : V \to W$ such that

1. T respects addition: $T(v + v') = T(v) + T(v')$ for any $v, v' \in V$.
2. T respects scalar multiplication: $T(cv) = cT(v)$ for any $c \in \mathbb{R}$ and $v \in V$.

Proposition 4: Criterion for linear transformation

A map $T : V \to W$ is a linear transformation iff for all $c \in \mathbb{R}$ and $v, v' \in V$,

$$T(cv + v') = cT(v) + T(v'). \tag{10}$$

This proposition is just a slightly simpler way to check that a map is a linear transformation. Note that any linear transformation must fix 0, because $T(0) = T(0 \cdot v) = 0T(v) = 0$, where $v \in V$ is arbitrary.

Linear transformations are important because they preserve the linear structure of vector spaces. This will be a general theme that will come up again and again throughout these notes; we will see that if a map preserves specific properties, it is called a *morphism*. When we have an algebraic structure on a set, as we do in this case (the vector space structure), we will rarely use regular maps. Instead, we will impose additional constraints on the maps that in turn, preserves the structure of the spaces of interest. We do this because the purpose of algebra is to study spaces with specific algebraic properties, so if we want to study functions on these spaces, these functions had better well preserve the algebraic properties we set out to study!

Given a linear transformation $T : V \to W$, there are two subspaces of interest. These two subspaces will appear more generally throughout these notes, when the maps of interest are different types of morphisms, be it a group homomorphism, a continuous map, a diffeomorphism, and many other types of map (we will define all of these types of maps later).

Definition 7: Kernel, Image

Let $T : V \to W$ be a linear transformation. The **kernel** of T is the set of vectors mapped to zero under T,

$$\ker(T) := \{v \in V : T(v) = 0\}. \tag{11}$$

The **image** of T is the set of vectors that T can reach in W,

$$\mathrm{im}(T) := \{w \in W : \exists v \in V \text{ s.t. } T(v) = w\}. \tag{12}$$

Note the kernel of an operator is a subset of its domain, and the image of an operator is a subset of its codomain. The image of a map was defined in Section 1, Eq. (**??**), and the same definition applies here, just T has more structure than the arbitrary map f in Eq. (**??**) had. The kernel and image of a morphism will have the same definition as above, just defined slightly differently when the map has different algebraic properties; we will encounter more on this soon.

Now, although the kernel and image of a linear transformation were just defined as sets, they are actually subspaces.

Proposition 5

Let $T : V \to W$ be linear. The kernel of T is a subspace of V, and the image of T is a subspace of W.

Proof. We will prove this for the kernel, and you should prove it for the image. Let $c \in \mathbb{R}$, and suppose $v, v' \in \ker(T)$. Then $T(cv + v') = cT(v) + T(v') = c0 + 0 = 0$, because T annihilates both v and v'. □

The kernel and image give us information about the injectivity and surjectivity of the linear transformation.

Proposition 6

Let $T : V \to W$ be linear. Then, T is injective (an embedding) iff $\ker(T) = \{0\}$, and T is surjective (onto) iff $\mathrm{im}(T) = W$.

Proof. That surjectivity implies $\mathrm{im}(T) = W$ is immediate. For the first claim, we prove $\implies$ first. If T is injective, then only one element can map to $0 \in W$; this must be $0 \in V$, as we already showed that $T(0) = 0$ always. Hence, $\ker(T) = \{0\}$. For $\impliedby$, assume $\ker(T) = \{0\}$, and suppose $T(v) = T(v')$. Then $T(v - v') = 0$, which can occur iff $v - v' \in \ker(T) = \{0\}$. Hence, $v - v' = 0$, so $v = v'$ and the map is injective. □

This now allows us to state the **rank-nullity theorem**, which essentially tells us that there is only so much "space" to be taken up by a linear map. The proof of this theorem is instructive but it is a bit too complicated for these notes.

Theorem 4: Rank-nullity

Let $T : V \to W$ be linear. Then,

$$\dim V = \dim \ker(T) + \dim \mathrm{im}(T). \tag{13}$$

This should be reasonably intuitive. If we start by mapping V somewhere, then we can't lose information; any vectors mapped to zero go into the kernel, and otherwise they go into the image. We'll see later how this connects to other algebraic theorems, in particular the so-called "first isomorphism theorem" that relates the image of a group homomorphism to the original group modded out by the kernel of the transformation.

0.0.3 Product and Quotient spaces

Rank-nullity is often the first introduction to quotient spaces, but typically it's taught as a separate theorem and often quotient spaces are not brought up at all in a first linear algebra course (which is very strange). These are two constructions that allow you to form new vector spaces from old ones.

We'll start with product spaces, which we've already implicitly defined and worked with because n-dimensional Euclidean space $\mathbb{R}^n$ is a product space.

Definition 8: Product Vector Space

Let $V_1, \ldots, V_k$ be vector spaces. The Cartesian product $V_1 \times \ldots \times V_k$ can be equipped with a vector space structure via component-wise addition and scalar multiplication,

$$(v_1, ..., v_k) + (v_1', ..., v_k') := (v_1 + v_1', ..., v_k + v_k') \qquad c(v_1, ..., v_k) := (cv_1, ..., cv_k) \tag{14}$$

and the space $(V_1 \times ... \times V_k, +, \cdot)$ is called the **product space**.

The product of n 1-dimensional spaces $\mathbb{R}$ is Euclidean space $\mathbb{R}^n := \mathbb{R} \times ... \times \mathbb{R}$. Note that the dimension of product spaces add,

$$\dim (V_1 \times ... V_k) = \sum_{i=1}^{k} \dim V_i. \tag{15}$$

We can see this by explicitly constructing a basis for this space, given bases $\{v_j^{(i)}\}_{j=1}^{\dim V_i}$ for each space V_i,

$$\left\{(v_j^{(1)}, 0, ..., 0)\right\}_{j=1}^{\dim V_1} \cup \left\{(0, v_j^{(2)}, ..., 0)\right\}_{j=1}^{\dim V_2} \cup ... \cup \left\{(0, 0, ..., v_j^{(k)})\right\}_{j=1}^{\dim V_k}. \tag{16}$$

You should convince yourself that this is in fact a basis of $V_1 \times ... \times V_k$.

The product space is equpped with canonical projection operators onto each component,

$$\pi_i : V_1 \times V_k \twoheadrightarrow V_i \qquad\qquad (v_1, ..., v_k) \mapsto v_i. \tag{17}$$

These projection operators are linear (show this!) and surjective onto V_i, with kernel isomorphic to $V_1 \times ... \times \hat{V_i} \times ... \times V_k$, where the notation $\hat{V_i}$ means we remove V_i from the product.

Let's restrict to the case of the product of two spaces $V \times W$, where we have the following projection maps,

$$\begin{tikzcd} & V \times W \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ V & & W \end{tikzcd} \tag{18}$$

We'll later see that objects in category theory are defined by a universal property: a universal property is a statement about the existence of maps which extend or factorize into existing maps. Suppose there exists another vector space Z equipped with linear transformations into V and W,

$$p_1 : Z \to V \qquad\qquad p_2 : Z \to W. \tag{19}$$

then these maps **factor through the product space**. The universal property of the product states that then there exists a unique map,

$$f : Z \to V \times W, \tag{20}$$

such that p_i factorize through f and the projection maps, given by the commutation of this diagram,

$$\begin{tikzcd} & Z \arrow[dl, "p_1"'] \arrow[d, dashed, "f"] \arrow[dr, "p_2"] & \\ V & V \times W \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & W \end{tikzcd} . \tag{21}$$

The dashed line by f means that such an f exists when the other maps in this diagram exists. The reason this statement is non-trivial is because these maps do not need to have inverses, since the projection maps are not injective. This means we can't explicitly construct f by composing p_i with π_i^{-1} as we would like: instead, the universal property of the product states that such a map exists, but does not give a particular construction for this map. The intuition for this property is that the product space is the "best" vector space $V \times W$ that is equipped with products into each of its factors V and W. Any time we have a different space Z equipped with its own "projection maps", there is a way to map Z into the product $V \times W$ and reinterpret the "projection maps" p_1 and p_2 as the projections π_1 and π_2.

The other construction we will consider is the **quotient space**. This takes the notion of equivalence classes that we considered previously (Section **??**) and equips the set of equivalence classes with a vector space structure.

Let $W \leq V$ be a subspace of V. For $v \in V$, we can translate the subspace W by v,

$$v + W := \{v + w : w \in W\}. \tag{22}$$

This is the **translate** of v by W: note that $v + W$ is not a vector space unless $v \in W$, in which case the space equals W (in general, $v + W$ is an **affine space**, which is a shifted linear space). In fact, $v + W$ has the interpretation as the *coset* of an equivalence relation.

Proposition 7

Let $W \leq V$. The binary relation $\sim$ on V defined by

$$v \sim v' \text{ iff } v - v' \in W \tag{23}$$

is an equivalence relation. The coset (equivalence class) of $v \in V$ under $\sim$ is exactly the translate of v by W,

$$[v] = v + W. \tag{24}$$

Proof. 1. (Reflexivity) Clearly $v \sim v$ because $v - v = 0 \in W$, as W is a subspace of V.

2. (Symmetry) If $v - v' \in W$, then $v' - v = -1(v - v') \in W$, hence $v \sim v'$ implies $v' \sim v$.

3. (Transitivity) Let $v \sim v'$ and $v' \sim v''$. Then,

$$v - v'' = (v - v') + (v' - v'') \in W \tag{25}$$

because $v - v' \in W$ and $v' - v'' \in W$, hence $v \sim v''$.

Now we show $[v] = v + W$. For $\subseteq$, let $v' \in [v]$, so $v' \sim v$. Then $v' - v \in W$, hence $v' \in v + W$. For $\supseteq$, let $v' \in v + W$, so $v' = v + w$ for some $w \in W$. Then $v' - v = w \in W$, hence $v' \sim v$ and $v' \in [v]$, which proves the claim. $\square$

Based on what we talked about with equivalence classes, the set of cosets $\{v+W : v \in V\}$ form a partition of V. We can put a vector space structure on these cosets: this is called the **quotient space**.

Definition 9: Quotient space

Let V/W denote the set of cosets $\{v + W : v \in V\}$. Then V/W is a vector space, called the **quotient space**, under the following operations:

$$(v + W) + (v' + W) := (v + v') + W \qquad c(v + W) = cv + W. \tag{26}$$

The quotient V/W forms another vector space by identifying W with the zero vector. To get a feel for this geometrically, the idea is that you foliate the space V with translates of W: that is, you partition V as set set of disjoint translates $\{v + W : v \in V\}$. You then collapse each of these translates to a point, but consistent with the vector space structure that you had originally. It's the same idea we considered with equivalence classes previously, with a specific type of equivalence relation that produces a new vector space.

Example 1: Quotienting $\mathbb{R}^3$ by the yz plane

Let's work out the quotient of $V := \mathbb{R}^3$ by the plane $W := \{(0, y, z) : y, z \in \mathbb{R}\}$. The most intuitive way to do this is to parameterize the set of cosets $v + W$ in a disjoint manner. Each coset is related by translation by a vector: but as we said, if $v \in W$, then v does not translate W to anything interesting: it just gets absorbed into W. So, consider the linear subspace,

$$X := \{(x, 0, 0) : x \in \mathbb{R}\}. \tag{27}$$

This parameterizes the cosets of $v + W$ in a disjoint manner:

$$V/W = \bigcup_{x \in X} x + W \qquad \forall x_1, x_2 \in W,\ (x_1 + W) \cap (x_2 + W) = \emptyset. \tag{28}$$

The geometric picture here is to form the full space V, you start with W, and then translate W by each $x \in X$. This gives you a foliation of V in terms of affine spaces, which are each in turn translates of W. The quotient space is then formed by reducing each of these translates to a single point: it looks like a line and is isomorphic to $\mathbb{R}$,

$$V/W \cong \mathbb{R}, \tag{29}$$

(the map $\mathbb{R} \to V/W$, $x \mapsto (x, 0, 0) + W$ is an isomorphism). Note that although the *foliation* of V by W is unique, the parameterization of the foliation (X) is not unique. In particular, we could have done the same thing with $X := \{(x, x, 0) : x \in \mathbb{R}\}$. But, even with a different foliation, we still are left with the same quotient space: both of these parameterizations gives the same V/W in Eq. (28). See Figure 1 for a geometric interpretation of this idea.

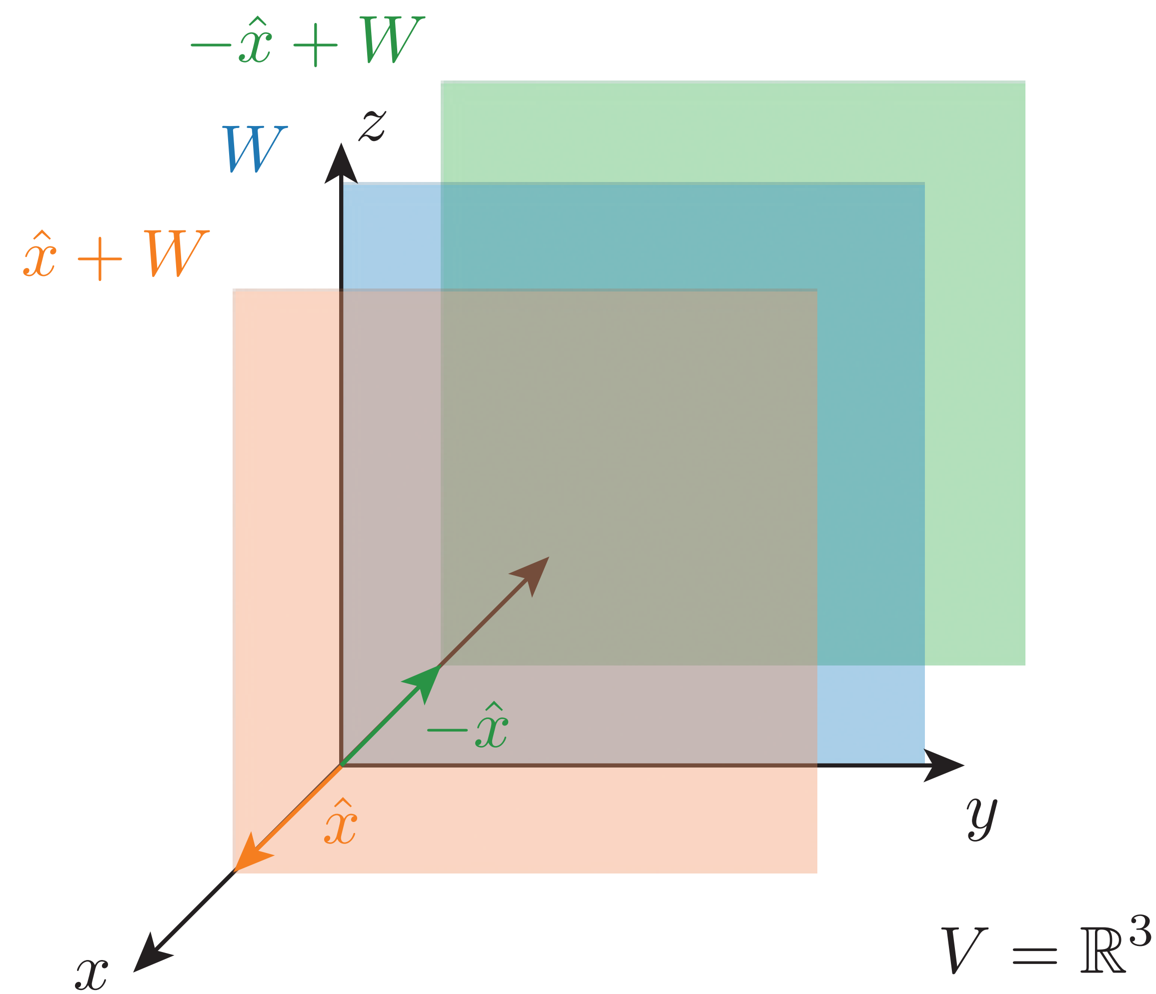

Figure 1: Example of what "foliating $\mathbb{R}^3$ by the yz plane" looks like as described in Example 1 for three different cosets: $\hat{x} + W$ (orange), W (blue), and $-\hat{x} + W$ (green). Each shifted yz plane is a different coset (translate) of W. Each coset is distinct ($v + W \cap v' + W = \emptyset$ if $v - v' \notin W$) and the union of all cosets, $\bigcup_{x\in\mathbb{R}}(x\hat{x} + W) = \mathbb{R}^3$, forms the volume $\mathbb{R}^3$. The quotient space V/W is the *space of cosets* and is isomorphic as a vector space to the x-axis.

What is the dimension of V/W, if $\dim V = n$ and $\dim W = m < n$? It's formed by essentially collapsing a dimension m subspace into a single point with dimension 0, so we expect the dimension to be $n - m$,

$$\dim(V/W) = \dim(V) - \dim(W). \tag{30}$$

How do we prove this? We'll use the rank-nullity theorem. There is a canonical projection map,

$$\pi : V \to V/W \qquad\qquad v \mapsto v + W, \tag{31}$$

which is a surjective linear transformation[1]. Why is it linear? Because

$$\pi(cv + v') = (cv + v') + W = c(v + W) + (v' + W). \tag{32}$$

The projection map in many ways *defines* the quotient space: given a vector $v \in V$, it projects it onto the appropriate coset. The kernel of π is simply W, since $0 + W$ is the zero vector in V/W and $v + W = W$ if and only if $v \in W$,

$$\ker \pi = W, \tag{33}$$

and the image is of course V/W because π is surjective. Applying the rank-nullity theorem to this yields,

$$\dim V = \dim \ker(\pi) + \dim \operatorname{im}(\pi) = \dim W + \dim V/W \tag{34}$$

which proves our claim.

The rank-nullity theorem is actually a specific generalization of a very important property of maps in algebraic categories: namely, there is a relationship between the kernel of a map, the image, and the domain of the map that we will explore in the next section (Theorem 6) after we discuss isomorphism.

0.0.4 Isomorphism: Equality of Vector Spaces

The final topic of discussion for this section is that of isomorphism.

Definition 10: Isomorphism

An **isomorphism** between V and W is a bijective linear transformation $T : V \to W$. If an isomorphism between V and W exists, we say that V and W are isomorphic, and we write $V \cong W$.

When we study maps that respect the linear vector space structure, they become linear transformations; when these maps are *bijections*, they become isomorphisms. An isomorphism between two vector spaces is the strongest statement you can make about these spaces, outside of saying they are equal. **Isomorphism essentially acts as equivalence from the algebraic perspective. If two spaces are isomorphic, they look and function the exact same way with respect to the vector space structure**[2]. Isomorphism is particularly strong in vector spaces (compared to other algebraic objects that we will study next) because the *dimension* of a finite-dimensional vector space determines its isomorphism class. **Essentially, this means that there is only a single way for a vector space of a given dimension to look: it looks like Euclidean space**[3] $\mathbb{R}^n$.

[1] Whenever we study quotients, we'll always have a projection map. It's extremely useful and in category theory actually defines what the quotient object is.

[2] I emphasize that these two spaces act the same way with respect to the vector space structure because isomorphic spaces are still different as sets, and can have *different* operations defined on them. For example, the spaces $\mathbb{R}^2$ and $\mathbb{C}$ are isomorphic as $\mathbb{R}$-vector spaces, and so function the same way in regards to linearity. However, $\mathbb{C}$ has an additional field structure that is induced by complex multiplication. You can define this on $\mathbb{R}^2$ if you want, but it is not as natural as it is in $\mathbb{C}$.

[3] Over a more general field K it will look like K^n.

Theorem 5

Two finite-dimensional vectors spaces V and W are isomorphic if and only if $\dim V = \dim W$.

Proof. Let $\{v_i\}_{i=1}^n$ and $\{w_j\}_{j=1}^m$ be bases for V and W, respectively. We will prove $\Longleftarrow$, and sketch the ideas of the $\Longrightarrow$ part of the proof for you to do. Suppose $n = m$. We define a map $T : V \to W$ by its action on the basis $\{v_i\}$. Note that if we define the action of a map on a basis of V, it canonically *extends to a unique linear map on all of* V via

$$T\left(\sum_{i=1}^n c_i v_i\right) = \sum_{i=1}^n c_i T(v_i). \tag{35}$$

You can verify that this map is both linear, and that it is the unique linear extension of a map sending $v_i \mapsto T(v_i)$. In the case $T : v_i \mapsto w_i$, we can explicitly show T is an isomorphism. It is already linear, hence we just need to show it is a bijection. It is clearly a surjection because $\{w_i\}$ is a basis for W, so for $w \in W$, one can expand $w = \sum_i c_i w_i$, hence $w = \sum_i c_i w_i = \sum_i c_i T(v_i) = T(\sum_i c_i v_i) \in \mathrm{im}(T)$. To show T is injective, let $v \in \ker(T)$, and expand $v = \sum_{i=1}^n c_i v_i$. Then $T(v) = \sum_i c_i w_i = 0$, but linear independence of $\{w_i\}$ implies that $c_i = 0$. This implies that $v = 0$, hence $\ker(T) = 0$ and T is injective. $\square$

Note that the map T that was constructed in the previous proof mapped $V \leftrightarrow W$ via sending basis elements of V to basis elements of W; it is not natural, because it depends on the choice of basis of V and W. An isomorphism between spaces that is defined *without reference to a specific basis* is called a **natural isomorphism**.

Exercise 1

Please prove the converse statement that we did not prove for Thm 5: that $V \cong W$ implies that $\dim V = \dim W$. The easiest way to show this is to take a basis $\{v_i\}$ for V, and show that $\{T(v_i)\}$ is basis for W if T is an isomorphism.

To summarize: a bijection between sets says that they are equal as sets; they have an equal cardinal, and they look the same from purely a counting perspective. Once an algebraic structure is put on a set (in our case the linear structure of a vector space), it is no longer enough for these sets to simply be in bijection: we also want them to have the same algebraic structure that is compatible with having the same structure at the set level. This is why isomorphisms are so important. They tell us that with regard to the algebraic structure of interest, the two spaces function in the exact same way. Two isomorphic vector spaces should be regarded as "the same" as one another, at least in terms of their linear structure.

Interlude 2: Embeddings

I think embeddings are very cool, and I certainly overuse them when I discuss algebra. Nonetheless, I'll try to inundate you into the propaganda. Let W and V be two vector spaces (not in the same ambient space, so they share no elements, which we can think of as $W \cap V = \emptyset$) with $\dim W \leq \dim V$. An **embedding of W into V** is an injective linear transformation $T: W \to V$, which we denote as

$$T: W \hookrightarrow V. \tag{36}$$

Note that equivalently the kernel of T is trivial, $\ker T = \{0_W\}$. What does an embedding do? It **embeds a copy of W into V**. Namely, the image of T, which is a subspace of V, is isomorphic to W, because the kernel of T is trivial,

$$W \cong T(W) \leq V. \tag{37}$$

Even though we started out with two separate vector spaces, if we have an embedding $W \hookrightarrow V$, we can view W as a subspace of V. This is particularly useful in group theory, because as we'll see in the next section, classificiation of groups is a much harder problem than that of vector spaces.

We will see this question asked frequently in other contexts: **can we classify all the algebraic objects of a given type up to isomorphism**? The "type" of an object is called a **category**, and it will be the focus of Sec **??** when we discuss Category Theory. Examples of categories are the category $\mathbb{R}-\mathrm{Vec}$ of vector spaces over $\mathbb{R}$, or the category Grp of groups that we will discuss in the next section. This "classificiation theorem" is very easy to do for finite-dimensional vector spaces (there is only one unique vector space of dimension n, $\mathbb{R}^n$), but other categories have more structure and make classification more difficult. We'll talk about this in the next section for groups.

The central idea behind classification is this: if I know "how big" the object is (dimension as a vector space, or cardinality as a group), how many different ways can it look up to isomorphism (since two isomorphic objects are "the same")? With vector spaces, the answer is pretty unambiguous. The "size" of a vector space is measured by its dimension. As a set, it has infinite cardinality, so it is not useful to discuss $|V|$ unless it is a vector space over a finite field[4]; thus the intuitive picture of how large a vector space is should be its dimension. It's the same way that you distinguish that 3D space $\mathbb{R}^3$ is "bigger" than the real line $\mathbb{R}$; the former has dimension 3, and the latter has dimension 1, despite the fact that they both have cardinality continuum, $\aleph_0$. For a vector space, knowledge of the dimension of the space immediately tells us what the vector space structure is; there is only one unique (non-isomorphic) vector space (over $\mathbb{R}$) of each dimension n, and that is $\mathbb{R}^n$. The dimension of a finite-dimensional vector space completely determines what it looks like. We'll see later that this is not true for other objects: I can write down two groups that have 8 elements, for instance!

Interlude 3: Space of linear transformations

Let V and W be vector spaces. The space of linear transformations from V to W is the set of all possible linear transformations between the spaces,

$$\mathrm{Hom}(V, W) := \left\{ T: V \to W \,\middle|\, T \text{ is linear.} \right\}. \tag{38}$$

[4] We will define what this means and discuss it in Section **??**.

The remarkable thing about this space is that it can be made into a vector space itself. We define addition and scalar multiplication on $\mathrm{Hom}(V, W)$ point-wise. For $S, T \in \mathcal{L}(V, W)$ and $c \in \mathbb{R}$, we define $S + T : V \to W$ and $cT : V \to W$ by their action on a vector $v \in V$,

$$(S+T)(v) := S(v) + T(v) \qquad\qquad (cT)(v) := cT(v). \tag{39}$$

This defines $S + T$ and cT as maps, but it does not necessarily show that they are elements of $\mathrm{End}(V, W)$. To do this, one must show they are linear. So, fix $\alpha \in \mathbb{R}$ and $v, w \in V$, and observe,

$$\begin{aligned}(S+T)(\alpha v + w) &= S(\alpha v + w) + T(\alpha v + w) = (\alpha S(v) + S(w)) + (\alpha T(v) + T(w)) \\ &= \alpha(S+T)(v) + (S+T)(w). \\ (cT)(\alpha v + w) &= cT(\alpha v + w) = c(\alpha T(v) + T(w)) = \alpha(cT)(v) + (cT)(w),\end{aligned} \tag{40}$$

which shows $S+T$ and cT are linear maps and indeed elements of $\mathrm{Hom}(V, W)$. I will not prove the 8 axioms hold here, but instead note that the zero element of $\mathrm{Hom}(V, W)$ is the zero map $0 \in \mathrm{Hom}(V, W)$, which sends every element $v \in V$ to $0 \in W$, i.e. $0(v) = 0$. It is easy to show the axioms hold.

The notion that we can construct a vector space from linear maps is more general than this case, as we will explore later. Anytime we have a morphism, we can consider the space of morphisms. If an operation is defined on this space of morphisms in a specific way, we can usually induce an algebraic structure on this space of morphisms, as we have seen above for $(\mathrm{Hom}(V, W), +, \cdot)$. We'll see this later when we define the *endomorphism ring* of a module (an endomorphism is a homomorphism from an object to itself).

The final topic we now discuss was alluded to in the previous section about quotients: it is a fundamental connection between the domain of a map, its kernel, and its image. A variant of this theorem shows up in all sorts of algebraic objects, and in the case of groups this theorem takes on the important name of "the first isomorphism theorem" (we will explore all of the isomorphism theorems when we discuss groups). The same exact intuition holds for vector spaces and linear transformations, and I find it simpler to gain intuition for this theorem when working with vector spaces.

Theorem 6: First isomorphism theorem for vector spaces

Let $T : V \to W$ be a linear map. Then we have the isomorphism,

$$\mathrm{im}(T) \cong V/\ker(T). \tag{41}$$

If $\ker T = \{0\}$ is trivial, then note that $\mathrm{im}\, T$ is isomorphic to V, i.e. T *embeds* V into the codomain W.

Proof. The proof of this theorem is illustrative, so I would encourage you to try to do it yourself. The kernel of T is a subspace, so we can quotient V by it: we must establish an isomorphism between $\operatorname{im}(T)$ and $V/\ker(T)$. Define,

$$\widetilde{T} : V/\ker(T) \longrightarrow \operatorname{im}(T) \qquad\qquad v + \ker(T) \mapsto T(v) \tag{42}$$

We are mapping the *coset* $v + \ker(T)$ in the quotient $V/\ker(T)$ onto the value that v takes on under T.

For notational ease, denote the kernel by $Z := \ker(T)$. Before we can show $\widetilde{T}$ is an isomorphism, we need to show it is well-defined! That is, if we take two representatives[5] $v, v' \in v + Z$, does $\widetilde{T}$ send $v + Z$ and $v' + Z$ to the same place? Because $v + Z = v' + Z$, we know $v = v' + w$ for some $w \in Z$, hence

$$\widetilde{T}(v + Z) = T(v) = T(v' + w) = T(v') + \cancel{T(w)} = \widetilde{T}(v' + Z), \tag{43}$$

since $w \in Z$ implies $T(w) = 0$. So, $\widetilde{T}$ is a well-defined map.

We now need to show that $\widetilde{T}$ is an isomorphism. $\widetilde{T}$ is linear because,

$$\begin{aligned} \widetilde{T}(c(v + Z) + (v' + Z)) &= \widetilde{T}((cv + v') + Z) = T(cv + v') = cT(v) + T(v') \\ &= c\widetilde{T}(v + Z) + \widetilde{T}(v' + Z). \end{aligned} \tag{44}$$

We can compute its kernel to show it is injective. We have:

$$\ker(\widetilde{T}) = \{v + Z \in V/Z : \widetilde{T}(v + Z) = 0\} = \{v + Z \in V/Z : T(v) = 0\} = \{Z\}. \tag{45}$$

Since the coset Z (equivalently $0 + Z$) is the zero vector in V/Z, this shows that $\ker(\widetilde{T})$ is trivial, hence $\widetilde{T}$ is injective. Furthermore, since we map $\widetilde{T}$ onto the image of T, it is clearly surjective: hence $\widetilde{T}$ is the desired isomorphism and we have proved the claim! $\square$

The maps in this theorem are best viewed with the following diagram,

$$\begin{tikzcd} V \arrow[r, two heads, "\pi"] \arrow[dr, "T"'] & V/\ker(T) \arrow[d, dashed, "\widetilde{T}", "\cong"'] \\ & \operatorname{im}(T) \subseteq W \end{tikzcd} \tag{46}$$

As we've seen in previous sections, this is a **factorization of the map T through the quotient** $V/\ker(T)$. The dashed line means that the map $\widetilde{T}$ is **induced** by the existence of the map T. This diagram defines an **universal property** of the quotient, which defines $V/\ker(T)$ up to isomorphism (more on this later).

The intuition behind the first isomorphism theorem is as follows. You can essentially "quotient out" the bad stuff from any given map to make it surjective. What this means mathematically is you quotient out the kernel of the map, and then the corresponding map that you induce on the quotient space ($\widetilde{T}$) will be injective. By restricting the codomain of $\widetilde{T}$ to be the image of T, the map is then also surjective and hence an isomorphism. Quotienting

[5] Recall this means $v+Z = v'+Z$, each one represents the same coset, and equivalently $v = v' \mod Z \iff v - v' \in Z$.

can be viewed intuitively as "collapsing a subspace and all of its cosets to a single point". This isomorphism theorem then means that the non-injectiveness of a map T is given by the non-triviality of the kernel $\ker(T)$ and all of its cosets: by collapsing each coset $v + \ker(T)$ to a single element of the quotient space, you get rid of the non-injectiveness of the map and induce an injective linear transformation $\widetilde{T}$.

This isomorphism thoerem can be used to derive the rank-nullity theorem– simply take the dimension of each term in Eq. (41)– although the rank-nullity theorem can also be derived independently. Many first courses in linear algebra discuss the rank-nullity theorem in depth but leave out this isomorphism theorem, which in my opinion is much more important and illustrates a more complex structure underlying linear maps. Typically first courses in linear algebra steer away from discussing quotient spaces, because it tends to require a bit more mathematical maturity than you typically have when you learn linear algebra. Nonetheless, I find that it is easiest to gain intuition for quotient spaces in the case of linear algebra, since most people are much more familiar with vector spaces than they are groups, which is when quotienting is usually seen first.

0.0.5 Takeaways

We began this section with real-valued vector spaces because there are a number of themes that will translate to studying other **algebraic objects**. When I say "algebraic object", I mean a set (in this case V) equipped with one or more operations that satisfy specific axioms (in this case, the vector space axioms). The additional algebraic structure that these operations add to the set is the primary study of interest of algebra. Many of the things we looked at in this section have counterparts in pure set theory, but in this section we allowed these counterparts to inherit the algebraic structure of interest (that of a vector space) so we can do more with them. For example, instead of just considering subsets of a vector space, we considered subspaces of a vector space. Because we impose additional structure on subsets to make a subspace, we can do more with them! Here is a list of the ideas that we will see propagate through the rest of this chapter.

- Sub-objects. We consider subsets $W \subseteq V$ that inherit its vector space structure, which we call subspaces, $W \leq V$. Because W is a vector space in its own right, this additional structure allows us to better understand W; we can write out its own bases, it has its own dimension, and we can use any

- Generating sets. The span of a set of vectors is the smallest subspace that contains the set of vectors. We say that $\text{span}\{v_1, ..., v_k\}$ is **generated** by $\{v_1, ..., v_k\}$, and it is **universal** in the sense that any subspace containing $\{v_1, ..., v_k\}$ also contains $\text{span}\{v_1, ..., v_k\}$. We will see this notion a lot; one can generate an algebraic space containing any number of elements by applying whatever the algebraic operation of interest is in all possible ways to close the space. If one only takes the bare minimum of elements in order to create a closed algebraic space, then this will be a universal object in the sense that any other space containing the initial set will also contain this generated space.

- Maps. Instead of considering any map $V \to V'$, we are specifically considering *linear transformations*; that is, maps that preserve the vector space structure (linearity).

The additional requirement of the map to preserve linearity allows us to understand it via its kernel and image, which are subspaces of V and V' in their own right. We will see that maps that respect algebraic structures are found in every area of algebra, and are more generally called **homomorphisms**. For example, we could call a linear transformation a **vector space homomorphism**.

- Kernels and images. The kernel and image of a structure-preserving map (morphism) will generally be sub-objects of the parent objects that we are studying. Note that from their definitions, they are subsets of either the domain or codomain, defined *without* specifically having the same algebraic properties as their parent space. However, because the maps of interest preserve algebraic structure, their kernels and images will inherit this structure automatically; we will never need to specify the kernel of a morphism is a sub-object, it will just come out for free from the nature of the morphism.

- Bijections and equivalence of objects. We will see that two objects are generally "algebraically equal" if an isomorphism exists between them; an isomorphism will always be defined to be a bijective homomorphism. Isomorphisms, in essence, allow one to "relabel" the object of interest. If an isomorphism exists, it shows that the two objects it maps between have the same number of elements, and that the algebraic operation acts on both objects in the exact same way. This is more clear when studying groups; because each finite-dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$, classifying how different vector spaces can look is pretty boring, whereas it's a very interesting question for groups.