A rough guide to linear algbera

Dongryul Kim

Version of April 18, 2020

Contents

Preface

At a certain point, one adopts a mode of learning mathematics. Then at a later point, after progressing through various styles of teaching and writing, one develops one's own point of view for how mathematics is to be taught. This book is an arrogant attempt to reassemble linear algebra according to the author's pedagogy.

About this book

There are countless (seriously, too many) books on linear algebra, and it is necessary to ask why we need another. Most of these books are designed for a first course in linear algebra for college students, which possibly include an emphasis in proof writing. However, there is a definite discrepancy between the presentation of linear algebra in such textbooks and how mathematicians understand the theory. Almost every professional mathematician has a natural functorial picture of linear algebra, but this is rarely written down in textbooks. Hence in reality, one acquires the functorial understanding while learning more advanced mathematics, such as commutative algebra or vector bundles. The philosophy of this book is that the determined reader can benefit from being directly introduced to the abstract development of linear algebra.

Hence this book will be most suitable to a reader who is (1) planning to pursue a career in mathematics, and (2) is already comfortable with rigorous presentations of mathematics. I believe it can also be used for a second course in linear algebra. The audience I had in mind when writing the book are students who had experience in mathematical competitions wanting to learn college-level mathematics. There are minimal prerequisites for reading this book, aside from fluency in mathematical communications, i.e., reading and writing proofs.

I tried to minimize the material dealt in this book. My conception of the book is that it is the bare bones of linear algebra. Most mathematicians, working in all fields, will be familiar with all the material in this book. Linear algebra arguably is the main tool for studying mathematical objects. On the other hand, the book will serve as a solid background for learning other parts of mathematics. I have tried to explain some of the applications of linear algebra in the epilogue.

Organization

In Chapter 1, we introduce naïve set theory. Set theory provides the foundation of modern mathematics, and it is necessary that the reader is familiar with sets and the basic operations. We have also included a first taste of category theory, introducing the concept of commutative diagrams and universal properties.

In Chapter 2, we introduce the theory of vector spaces. Linear algebra is, and should be, the study of vector spaces and linear maps, not of matrices. Matrices are a good tool for computations, but abstraction is necessary for a solid conceptual picture. During the first half of the chapter, we carefully develop various constructions in the category of vector spaces, e.g., products, direct sum, subspaces, quotient spaces, kernel, cokernel, image, internal hom, etc. In the second half, we prove the theorem that every vector space has a basis, and discuss applications of this fact. We will not ignore infinite-dimensional spaces. There is also a section on applications of linear algebra in combinatorics, for students with a competition mathematics background.

In Chapter 3, we introduce tensor products, the symmetric algebra, and the exterior algebra. We discuss the universal properties of these vector spaces as a motivation. The determinant is defined using the exterior algebra, as it should be. Using this definition, we prove Cramer's rule and discuss Gaussian elimination.

In Chapter 4, we introduce modules over a commutative ring. Modules are important objects in commutative algebra, but our main goal in this chapter is study the different possible ways a linear map can act on a vector space. We prove the classification theorem of finitely generated modules over a principal ideal domain, and use this to discuss Frobenius normal form, Jordan normal form, eigenvectors, eigenspaces, and generalized eigenspaces.

In Chapter 5, we do linear algebra combined with analysis. We define inner product structures and finite-dimensional Hilbert spaces, and lead up to proving the spectral theorem on finite-dimensional spaces. There is a more algebraic proof in the finite-dimensional case, using the Jordan normal form for instance, but we use an analytic argument that applies also to compact operators on infinite-dimensional Hilbert spaces.

There are plenty of exercises. I have tried to meld the exercise with the text so that they appear in context. Some of the exercises will be used in later proofs, and some of them will be used implicitly throughout the book. Others will be unimportant from a theoretical perspective, but nonetheless helpful in properly understanding the material. Abstraction can be a double-edged sword, as the abstract loses meaning when disconnected from the concrete. By doing the exercises, the reader is expected to train moving between the two worlds.

Acknowledgements

This project started with a small series of lectures delivered to the Korean team for the 2017 International Mathematical Olympiad, a group of six high school students. The course was aborted due to lack of time, but the idea of writing

a set of linear algebra notes remained. I would like to thank everyone who had attended my ill-prepared lectures, as well as those who were beside me when I was working on the book.

The current text is a draft, and will be updated occasionally. I beg everyone to send me typos, grammatical errors, mathematical errors, suggestions, comments, criticisms, anything about the book. You can reach me at `dkim04@stanford.edu`. The latest version of this book can be obtained at my website `http://dongryulkim.wordpress.com`. Happy reading!

April, 2020

Dongryul Kim

Here, I keep a list of chores I need to do. The draft will be a final draft once the list is empty.

- Probably expand the chapter “Preface” once I indicate optional/required material
- Do we move where we discuss Gaussian elimination? Maybe matrices and algorithms deserve a separate section?
- Write about the LU decomposition
- Include somewhere a discussion Cayley–Hamilton
- Fill in the section “Duality in linear programming”
- Introduce functional calculus for self-adjoint operators
- Check which are optional, which are required
- Rank exercises by difficulty?
- Give hints for exercises?
- Add references at the end of each chapter?

Chapter 1

Sets

Most of mathematics is grounded on the notion of sets. In this chapter, we quickly review the basic set theory we use throughout the book, assuming that the reader is familiar with most of the material.

1.1 Sets and maps

In standard mathematics, that is, Zermelo–Fraenkel set theory, literally every mathematical object is a set. Each of the numbers $0, 1, 2, \ldots$ is actually a set:

$$0 = \emptyset = \{\}, \quad 1 = \{\emptyset\} = \{\{\}\}, \quad 2 = \{\emptyset, \{\emptyset\}\} = \{\{\}, \{\{\}\}\}, \ldots.$$

But because the notion of sets provides a foundation for mathematics, it does not make sense to mathematically define what a set is. Instead, people take the notion as granted and assume that sets satisfy certain axioms. But we are not going to worry to much about these issues and the following definition will be good for us.

Definition 1.1.1. A **set** is a well-defined collection of objects. If an object x is inside the set X, we say that x is an **element of** X and write $x \in X$.

Example 1.1.2. The set of natural numbers $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \ldots\}$ is a set. The set of integers $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ is a set. The set of even integers

$$2\mathbb{Z} = \{\ldots, -2, 0, 2, 4, \ldots\} = \{2x : x \in \mathbb{Z}\}$$

is a set.

An issue arises when one tries to look at the set of all sets. In fact, in Zermelo–Fraenkel set theory, it is possible to prove that there is no set that contains all sets. But again, we are going to ignore such issues.

Definition 1.1.3. Given two sets X and Y, if $x \in Y$ implies $x \in X$, we say that Y is a **subset** of X and write $Y \subseteq X$. Equivalently, we say that X is a **superset** of Y and write $X \supseteq Y$. For instance, $\{1, 3\} \subseteq \{1, 2, 3\}$.

Definition 1.1.4. Given two sets X and Y, we define the **difference** $X \setminus Y$ as the subset

$$\{x \in X : x \notin Y\} \subseteq X$$

of X.

A subset of a subset is a subset

Exercise 1.1.A. Let A, B, C be sets. Show that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

Equality is inclusion in both directions

Exercise 1.1.B. Let A and B be sets. If $A \subseteq B$ and $B \subseteq A$, show that $A = B$. (Here, $A = B$ means that they have the same elements, i.e., $x \in A$ if and only if $x \in B$.) In the future, this will be a useful way to prove that two sets are equal.

Definition 1.1.5. Given two sets X and Y, their **intersection** $X \cap Y$ is the set consisting of elements in both of the sets, and their **union** $X \cup Y$ is the set consisting of elements in either one of the sets. For instance, $\{1,3\} \cap \{2,3\} = \{3\}$ and $\{1,3\} \cup \{2,3\} = \{1,2,3\}$.

Exercise 1.1.C. Let A and B be sets. Show that $A \cap B$ and $A \setminus B$ are disjoint, i.e., their intersection is $\emptyset$, and that their union is A.

Intersection and union are associative

Exercise 1.1.D. For sets A, B, C, show that $(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$.

Intersection distributes over union

Exercise 1.1.E. For sets A, B, C, show that $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

We are now ready to definite maps between sets. The right way to think about a map is as a machine that takes in an element of one set and spits out an element of another set. But this is not mathematically rigorous, and to make it rigorous, we need to complicate it a little bit. But be sure keep the intuitive picture of a map while reading the formal defenition.

Definition 1.1.6. Given two sets X and Y, define their **Cartesian product**

$$X \times Y = \{(x, y) : x \in X, y \in Y\}$$

as the set of (ordered) pairs (x, y) where x and y are elements of X and Y respectively.

Definition 1.1.7. Let X and Y be sets. A **map** or **function** $f : X \to Y$ is a subset $f \subseteq X \times Y$ such that for each element $x \in X$, there exists a unique $y \in Y$ such that $(x, y) \in f$. In such a case, we write $f(x) = y$ or $f : x \mapsto y$. We call the set X the **domain** of f and the set Y the **target** or **codomain** of f.

Example 1.1.8. Consider the map $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. According to this definition, this is actually a subset

$$f = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}.$$

If you draw this subset out on the coordinate plane $\mathbb{R} \times \mathbb{R}$, you get the graph of the parabola $y = x^2$.

Exercise 1.1.F. For each set X, show that there is a unique map $\emptyset \to X$. (Note that this is true even for $X = \emptyset$.) Also show that there is a unique map $X \to \{0\}$. (This set $\{0\}$ can be replaced by any set with exactly one element.)

Given an arbitrary set X, there is a canonical map $X \to X$ we can define.

Definition 1.1.9. The **identity map on** X is defined as

$$\mathrm{id}_X : X \to X; \quad x \mapsto x.$$

Formally, it is the diagonal subset

$$\mathrm{id}_X = \Delta = \{(x, x) : x \in X\} \subseteq X \times X.$$

We can also compose to maps to get another map.

Definition 1.1.10. Let $f : X \to Y$ and $g : Y \to Z$ be two maps. We define their **composite** to be

$$g \circ f : X \to Z; \quad x \mapsto g(f(x)).$$

It is rather unfortunate that $g \circ f$ means applying f first and then applying g. But this notation is pretty set in mathematics, so just remember to switching the order every time you compose maps. From now on, I will no longer use the "subset of $X \times Y$" interpretation of a map. Nobody seriously thinks of maps as sets, and instead tries to distinguish maps and sets as different types of objects. If we really want to discuss that particular subset, we will refer it to the **graph of** f.

Exercise 1.1.G. For $f : X \to Y$ a map, show that $f \circ \mathrm{id}_X = \mathrm{id}_Y \circ f = f$.

Composition is associative

Exercise 1.1.H. Given maps $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$, show that $(h \circ g) \circ f = h \circ (g \circ f)$. This shows that we can just write this composite as $h \circ g \circ f$ without ambiguity.

Definition 1.1.11. Let $f : X \to Y$ be a map. For a subset $T \subseteq Y$, we define its **inverse image** as

$$f^{-1}(T) = \{x \in X : f(x) \in T\}.$$

For a subset $S \subseteq X$, we define its **image** as

$$f(S) = \{f(x) : x \in S\}.$$

Inverse image distributes over unions and intersections

Exercise 1.1.I. Let $f : X \to Y$ be a map. If $A, B \subseteq Y$ are two subsets, show that $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$ and $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$. But the analogous statement does not hold for images. Find examples of maps f and $C, D \subseteq X$ such that $f(C \cap D) \neq f(C) \cap f(D)$.

Definition 1.1.12. A map $f : X \to Y$ is called **injective** if for every $y \in Y$ there exists at most one $x \in X$ such that $f(x) = y$. It is called **surjective** if for every $y \in Y$ there exists at least one $x \in X$ such that that $f(x) = y$. A map is called **bijective** if it is both injective and surjective.

When a map is injective, we will sometimes draw the arrow as $X \hookrightarrow Y$ to indicate injectivity. When it is surjective, we will draw it as $X \twoheadrightarrow Y$. When it is bijective, we will draw $X \xrightarrow{\sim} Y$ or sometimes $X \cong Y$.

Exercise 1.1.J. Show that a map $f : X \to Y$ is surjective if and only if $f(X) = Y$.

Injectivity and surjectivity are closed under composition

Exercise 1.1.K. Let $f : X \to Y$ and $g : Y \to Z$ be maps. Show that if f and g are both injective, then $g \circ f$ is injective. Likewise, show that if f and g are surjective, then $g \circ f$ is surjective.

Exercise 1.1.L. Let $f : X \to Y$ and $g : Y \to Z$ be maps. Show that if $g \circ f$ is injective, then f is injective. Dually, show that if $g \circ f$ is surjective, then g is surjective. Find counterexamples to the statement that the other map is injective/surjective.

Note that $\mathrm{id} : X \to X$ is always bijective. From this exercise, we see that if $g \circ f = \mathrm{id}_X$, then f is injective and g is surjective. Hence if $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$ then both f and g are bijective. In this case, we say that f is an **inverse map** of g and vice versa.

Bijective maps have inverses

Exercise 1.1.M. Let $f : X \to Y$ be a bijective map. Show that there exists a map $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$.

Definition 1.1.13. We say that a set X is **finite** if there is a bijection between X and $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{Z}_{\geq 0} = \{0, 1, \ldots\}$. In this case, we write $|X| = n$ or $\#X = n$, and call n the **cardinality** of **size** of the set X.

Let X and Y be finite sets. Then there exists an injection $X \hookrightarrow Y$ if and only if $|X| \leq |Y|$. However, it is *not* true that there exists a surjection $Y \twoheadrightarrow X$ if and only if $|X| \leq |Y|$. Take $X = \emptyset$.

1.2 Products, coproducts, and sets of maps

We have seen how to take the product $X \times Y$ of two sets X and Y. We just take all possible pairs (x, y) for $x \in X$ and $y \in Y$. This can be easily generalized to products of more sets. If we have sets $X_1, \ldots, X_n$, we should be able to take

$$\prod_{i=1}^{n} X_i = X_1 \times X_2 \times \cdots \times X_n = \{(x_1, \ldots, x_n) : x_1 \in X_1, \ldots, x_n \in X_n\}.$$

Note that if $X_1, \ldots, X_n$ are finite sets, then $X_1 \times \cdots \times X_n$ is also a finite set and

$$|X_1 \times X_2 \times \cdots \times X_n| = |X_1||X_2| \cdots |X_n|.$$

This explains why we call this a product set.

Exercise 1.2.A. Show that the canonical map

$$X \times Y \times Z \to (X \times Y) \times Z; \quad (x, y, z) \mapsto ((x, y), z)$$

is a bijection. We can easily think of doing more complicated things by placing parentheses differently.

We can also start taking infinite products in a similar way. Let I be a set, which indexes sets. That is, for each element $i \in I$ there is going to be a set X_i, of which we will try to take a product. This indexing set I can be infinite like $\mathbb{Z}_{\geq 0} = \{0, 1, \ldots\}$ or very infinite like $\mathbb{R}$. Given this data, we can define their product.

Definition 1.2.1. Let I be an indexing set, and for each $i \in I$ let X_i be a set. The **product of** X_i is defined as

$$\prod_{i \in I} X_i = \{(x_i)_{i \in I} : x_i \in X_i\}$$

where $(x_i)_{i \in I}$ is an infinite tuple indexed by I. If you want to be super rigorous, you can think of the infinite tuple as a map $\varphi : I \to \bigcup_{i \in I} X_i$ such that $\varphi(i) \in X_i$ for all $i \in I$.

Exercise 1.2.B. If $I = \emptyset$, what is $\prod_{i \in I} X_i$? How many elements does it have?

The product of the empty set with other sets is empty

Exercise 1.2.C. Show that if there exists an $i_0 \in I$ such that $X_{i_0} = \emptyset$, then the product is $\prod_{i \in I} X_i = \emptyset$.

On the other hand, what if all the sets X_i are nonempty?

Products of nonempty sets are nonempty

Proposition 1.2.2 (Axiom of Choice)**.** *Let I be an indexing set, and for each $i \in I$ let $X_i \neq \emptyset$ be a nonempty set. Then their product $\prod_{i \in I} X_i$ is also nonempty.*

A product of sets comes with canonical projections maps defined as, for each $i_0 \in I$,

$$\pi_{i_0} : \prod_{i \in I} X_i \to X_{i_0}; \quad (x_i)_{i \in I} \mapsto x_{i_0}.$$

Exercise 1.2.D. Suppose X_i are all nonempty. Assuming the Axiom of Choice, show that each projection map π_{i_0} is surjective.

There is a dual notion of products, called coproducts. A funny habit of mathematicians is that they put the prefix "co-" in front of a word to make another word that describes the notion that is dual to the original one. It might not seem very clear at this point why coproducts is the dual notion of products, but we will see it in the next section. It is amazing how a lot of mathematics come in dual pairs.

Definition 1.2.3. Let I be an indexing set, and for each $i \in I$ let X_i be a set. The **coproduct of** X_i or the **disjoint union of** X_i is defined as

$$\coprod_{i\in I} X_i = \{(i,x) : i \in I,\, x \in X_i\} \subseteq I \times \bigcup_{i\in I} X_i.$$

The coproduct also comes with canonical maps, this time from the individual sets to the coproduct. For each $i_0 \in I$, define the inclusion map

$$\iota_{i_0} : X_{i_0} \to \coprod_{i\in I} X_i; \quad x \mapsto (i_0, x).$$

Exercise 1.2.E. Convince yourself if $X_1, \dots, X_n$ are finite sets, then $X_1 \amalg \cdots \amalg X_n$ is also a finite set and

$$|X_1 \amalg X_2 \amalg \cdots \amalg X_n| = |X_1| + |X_2| + \cdots + |X_n|.$$

Exercise 1.2.F. If $I = \emptyset$, what is $\coprod_{i\in I} X_i$?

Exercise 1.2.G. Show that each ι_{i_0} is always injective, without any assumptions.

Before moving on, let me introduce one more construction. Given two sets X and Y, the maps from X to Y themselves form a set.

Definition 1.2.4. Let X and Y be sets. The **set of maps from** X **to** Y is denoted by Y^X or $\mathrm{Mor}_{\mathsf{Set}}(X, Y)$.

The sans-serif Set indicates that we are working with sets, not some other mathematical structure, and Mor stands for morphisms, which are just maps when we work with sets. In algebra, it is important to always be aware of what "type" of mathematical object we are working with, and this will be a sort of reminder. For the rationale behind the other notation, do the following exercise.

Exercise 1.2.H. Let X and Y be finite sets. Convince yourself that Y^X is finite and $|Y^X| = |Y|^{|X|}$.

The next exercise can be confusing at first, but it is a crucial idea in mathematics that you can sometimes play around with maps like this. We will see the same idea over and over in different contexts.

Maps between sets allow currying

Exercise 1.2.I. Let X, Y, and Z be sets. Show that the map

$$\begin{aligned}\mathrm{Mor}_{\mathsf{Set}}(X, \mathrm{Mor}_{\mathsf{Set}}(Y, Z)) &\to \mathrm{Mor}_{\mathsf{Set}}(X \times Y, Z);\\ (x \mapsto (f_x : y \mapsto z)) &\mapsto ((x,y) \mapsto z = f_x(y))\end{aligned}$$

is a bijection. This map can be described alternatively as sending a map $f : X \to \mathrm{Mor}_{\mathsf{Set}}(Y, Z)$ to $(x, y) \mapsto (f(x))(y)$.

1.3 Fun with diagrams

So far we have not been looking at complicated situations with many functions and sets, but once we go into serious business, it will be hard to keep track of all the sets and maps. Diagrams will make this job considerably easier and more intuitive. For instance, if we have maps $f : A \to B$, $g : B \to D$ and $h : A \to C$, $k : C \to D$, we can draw this entire data as

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "h"] & B \arrow[d, "g"] \\
C \arrow[r, "k"] & D.
\end{tikzcd}$$

There are two ways to get a map $A \to D$ from this diagram: $g \circ f$ and $k \circ h$. Oftentimes, these two maps will be equal, and in this case, we are going to say that this diagram **commutes**. In general, a diagram can be much more complicated, and we are going to say that it commutes when all possible ways of composing maps give the same map as long as they have the same domain and target.

For example, consider the following diagram:

$$\begin{tikzcd}
A \arrow[r, "a"] \arrow[d, "b"] & B \arrow[r, "c"] \arrow[d, "d"] & C \arrow[d, "e"] \\
D \arrow[r, "f"] & E \arrow[r, "g"] & F.
\end{tikzcd}$$

This diagram commutes when $f \circ b = d \circ a$ and $g \circ d = e \circ c$ and $g \circ f \circ b = g \circ d \circ a = e \circ c \circ a$. But note that the last condition is unnecessary because $f \circ b = d \circ a$ already implies $g \circ f \circ b = g \circ d \circ a$ and $g \circ d = e \circ c$ implies $g \circ d \circ a = e \circ c \circ a$. Therefore the diagram commutes if and only if the two small squares commute.

Let me now prove an interesting proposition.

Maps into two sets is the same as a map into their product

Proposition 1.3.1 (Universal property for products)**.** *Let X and Y be sets, and $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ be the projection maps. Let W be an arbitrary set and $f : W \to X$ and $g : W \to Y$ be arbitrary maps. Then there exists a unique map $h : W \to X \times Y$ such that $f = \pi_1 \circ h$ and $g = \pi_2 \circ h$, i.e., the following diagram commutes:*

$$\begin{tikzcd}
 & W \arrow[ddl, "f"'] \arrow[d, dashed, "h"] \arrow[ddr, "g"] & \\
 & X \times Y \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
X & & Y.
\end{tikzcd}$$

Proof. We first prove existence. Consider the map

$$h : W \to X \times Y; \quad w \mapsto (f(w), g(w)).$$

Then $\pi_1(h(w)) = \pi_1((f(w), g(w))) = f(w)$ means that $f = \pi_1 \circ h$ and $\pi_2(h(w)) = \pi_2((f(w), g(w))) = g(w)$ means that $g = \pi_2 \circ h$. That is, this h satisfies the condition.

On the other hand, for h to satisfy this condition, h has to be precisely this map. The condition $\pi_1 \circ h = f$ means that the first component of $h(w)$ is $f(w)$, and the other condition $\pi_2 \circ h = g$ means that the second component of $h(w)$ is $g(w)$. Therefore if h satisfies the condition, it needs to send w to $h(w) = (f(w), g(w))$. This shows uniqueness. □

Such a property is called a **universal property**. It gives a universal characterization of this product $X \times Y$. Suppose there is another set $(X \times Y)'$ with maps $\pi_1' : (X \times Y)' \to X$ and $\pi_2' : (X \times Y)' \to Y$ satisfying the same property: for each $f : W \to X$ and $g : W \to Y$ there exists a unique $h : W \to (X \times Y)'$ such that $f = \pi_1' \circ h$ and $g = \pi_2' \circ h$. If we apply this property to $W = X \times Y$ and $f = \pi_1$, $g = \pi_2$, then we get a unique map $\psi : X \times Y \to (X \times Y)'$ such that $\pi_1 = \pi_1' \circ \psi$ and $\pi_2 = \pi_2' \circ \psi$. But if we apply Proposition 1.3.1 to $W = (X \times Y)'$ and $f = \pi_1'$ and $g = \pi_2'$ then we get a unique map $\varphi : (X \times Y)' \to X \times Y$ such that $\pi_1' = \pi_1 \circ \varphi$ and $\pi_2' = \pi_2 \circ \varphi$.

$$\begin{tikzcd}
 & X \times Y \arrow[ddl, "\pi_1"'] \arrow[d, "\psi"] \arrow[ddr, "\pi_2"] & \\
 & (X \times Y)' \arrow[dl, "\pi_1'"] \arrow[dr, "\pi_2'"'] & \\
X & & Y
\end{tikzcd}
\qquad
\begin{tikzcd}
 & (X \times Y)' \arrow[ddl, "\pi_1'"'] \arrow[d, "\varphi"] \arrow[ddr, "\pi_2'"] & \\
 & X \times Y \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
X & & Y
\end{tikzcd}$$

So we have canonically obtained maps φ and ψ between $X \times Y$ and $(X \times Y)'$. But I further claim that these maps are inverses to each other, and hence bijections. In view of the remark after Exercise 1.1.L, we only need to show that $\psi \circ \varphi = \mathrm{id}_{X \times Y}$ and $\varphi \circ \psi = \mathrm{id}_{(X \times Y)'}$. Here, we are going to use the uniqueness part of the property. Note that the two diagrams

$$\begin{tikzcd}
 & X \times Y \arrow[dddl, "\pi_1"'] \arrow[d, "\psi"] \arrow[dddr, "\pi_2"] & \\
 & (X \times Y)' \arrow[ddl, "\pi_1'"'] \arrow[d, "\varphi"] \arrow[ddr, "\pi_2'"] & \\
 & X \times Y \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
X & & Y
\end{tikzcd}
\qquad
\begin{tikzcd}
 & X \times Y \arrow[ddl, "\pi_1"'] \arrow[d, "\mathrm{id}_{X \times Y}"] \arrow[ddr, "\pi_2"] & \\
 & X \times Y \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
X & & Y
\end{tikzcd}$$

are commutative. By the uniqueness part of Proposition 1.3.1, applied to $W = X \times Y$ and $f = \pi_1$, $g = \pi_2$, we conclude that $\varphi \circ \psi = \mathrm{id}_{X \times Y}$. By a similar argument, but this time applying the property for $(X \times Y)'$, we also see that $\psi \circ \varphi = \mathrm{id}_{(X \times Y)'}$. So we have shown, by just playing around with diagrams, that the two sets $(X \times Y)'$ and $X \times Y$ have a unique bijection between them. Maybe it seems rather foolish to make this complicated argument, but again, we will revisit this idea as we proceed. The power of the argument shines when there is too much structure in the objects we are studying, and so constructing invertible maps between the objects is too complicated.

It is now time for you to do something similar. The coproduct, or disjoint union, also has a universal property. Surprisingly, you just reverse all the arrows!

Maps from two sets is the same as a map from their coproduct

Exercise 1.3.A (Universal property for disjoint unions)**.** Show the following universal property for the coproduct: for arbitrary maps $f : X \to W$ and $g : Y \to W$, there exists a unique map $h : X \amalg Y \to W$ such that $f = h \circ \iota_1$ and $g = h \circ \iota_2$.

$$\begin{tikzcd}
X \arrow[dr, "\iota_1"] \arrow[ddr, "f"'] & & Y \arrow[dl, "\iota_2"'] \arrow[ddl, "g"] \\
& X \amalg Y \arrow[d, dashed, "h"] & \\
& W &
\end{tikzcd}$$

Again, you can use the same argument to show that this property determines $X \amalg Y$ up to invertible maps. Products and coproducts can be generalized to limits and colimits, but we won't talk about this. Look them up if you're interested.

Exercise 1.3.B. Fix sets A, B, C and maps $\alpha : A \to C$ and $\beta : B \to C$. Construct a set P along with maps $\pi_1 : P \to A$ and $\pi_2 : P \to B$ that satisfy the following property: for arbitrary maps $f : W \to A$ and $g : W \to B$ satisfying $\alpha \circ f = \beta \circ g$, there exists a unique map $h : W \to P$ such that $f = \pi_1 \circ h$ and $g = \pi_2 \circ h$.

$$\begin{tikzcd}
W \arrow[dr, dashed, "h"] \arrow[drr, "g"] \arrow[ddr, "f"'] & & \\
& P \arrow[r, "\pi_2"] \arrow[d, "\pi_1"] & B \arrow[d, "\beta"] \\
& A \arrow[r, "\alpha"] & C
\end{tikzcd}$$

(This set P is called the fiber product of A and B over C.)

1.4 Equivalence classes

Equivalent classes provide us with a way of saying that two things are the "same". Consider the set $\{1, 2, 3, 4, 5\}$. If we care about these numbers as they are, we can just work with them. But if we only care about the parity of the numbers, whether they are even or odd, 1 and 3 makes no difference at all. Then we may as well say that 1 is the "same" as 3, in this perspective. We want this notion of "same" to satisfy these properties:

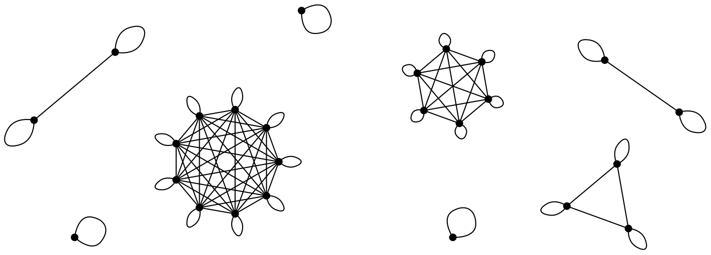

Figure 1.1: Visualizing an equivalence relation

(i) Always, x is the same as x.

(ii) If x is the same as y, then y is the same as x.

(iii) If x is the same as y and y is the same as z, then x is the same as z.

These are reasonable requirements for what we would like to call the same. This motivates the definition of equivalence relations.

Definition 1.4.1. Let X be a set. An **equivalence relation** R is a subset $R \subseteq X \times X$ satisfying the following properties: (we are going to denote $(x, y) \in R$ as $x \sim_R y$, because this is indicative of it being a relation rather than a subset)

(EQ1) For all $x \in X$, $x \sim_R x$.

(EQ2) For all $x, y \in X$, $x \sim_R y$ if and only if $y \sim_R x$.

(EQ3) For all $x, y, z \in X$, if $x \sim_R y$ and $y \sim_R z$ then $x \sim_R z$.

Imagine that there is a point corresponding to each element of X, and the two dots corresponding to $x, y \in X$ are connected by an edge if $x \sim_R y$, as in Figure 1.1. Then you could visualize the set of points forming groups, so that no two points in different groups are connected, while every two points in the same group are connected.

Definition 1.4.2. For $x \in X$, we define its **equivalence class** as

$$[x] = \{y \in X : x \sim_R y\} \subseteq X.$$

This is going to be the group which contains x, or the set of things that are the "same" as x.

Equivalence classes form a partition of the original set

Exercise 1.4.A. Show that if $x \sim_R y$, then $[x] = [y]$. (Hint: show $[x] \subseteq [y]$ and $[y] \subseteq [x]$.) Show that, on the other hand, if $x \not\sim_R y$, then $[x]$ and $[y]$ are disjoint, i.e., $[x] \cap [y] = \emptyset$.

We can now consider collapsing these sets like $[x]$ into a single point.

Definition 1.4.3. Let X be a subset with an equivalence relation R. Given a subset $S \subseteq X$, if there exists an element $x \in X$ such that $S = [x]$, then we say that S is an **equivalence class**. We define the **quotient** of X by R as

$$X/R = \{\text{equivalence classes of } X \text{ with respect to } R\}.$$

There is always a canonical projecion map

$$\pi : X \to X/R; \quad x \mapsto [x].$$

A map that sends equivalent elements to the same element is the saem as a map from the quotient

Exercise 1.4.B (Universal property for quotients)**.** Let X be a set with an equivalence relation R, and let $f : X \to Y$ be a function satisfying $f(x_1) = f(x_2)$ if $x_1 \sim_R x_2$. Show that f factors uniquely through π, i.e., there exists a unique function $g : X/R \to Y$ such that $f = g \circ \pi$.

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\pi"'] & Y \\
X/R \arrow[ur, dashed, "g"'] &
\end{tikzcd}$$

Sometimes, we would want to force some elements to be equal, but the elements we want to collapse might not be an equivalence relation. Consider the set $\{1, 3, 5\}$, and suppose that for some reason we want to identify $1 = 3$ and also $3 = 5$. The solution is easy. We collapse all $1, 3, 5$ into a single element.

Definition 1.4.4. Take an arbitrary subset $S \subseteq X$. The **equivalence relation generated by** S is the following equivalence relation: $x \sim_R y$ if and only if there exists an $n \geq 0$ and a sequence $x = x_0, x_1, \ldots, x_n = y$ of elements in X such that for each $0 \leq j \leq n-1$, either $(x_j, x_{j+1}) \in S$ or $(x_{j+1}, x_j) \in S$. (Here, if $n = 0$ we get $x \sim_R x$.)

Example 1.4.5. If $X = \{1, 2, 3, 4, 5\}$ and $S = \{(1, 3), (3, 5)\}$, then

$$R = \{(1, 1), \ldots, (5, 5), (1, 3), (3, 1), (1, 5), (5, 1), (3, 5), (5, 3)\}.$$

Exercise 1.4.C. Check that the R constructed as above is indeed an equivalence relation on X. (That is, verify the three axioms EQ1, EQ2, and EQ3.)

Exercise 1.4.D. If R' is another equivalence relation on X such that $S \subseteq R'$, show that $R \subseteq R'$. This means that R is the minimal (under inclusion) equivalence relation containing S.

Because we are lazy, for an arbitrary $S \subseteq X \times X$ we will write

$$X/S = X/R$$

where R is the equivalence relation generated by S. Because X/S is a quotient, it comes with a map $\pi : X \to X/R = X/S$.

Exercise 1.4.E. Let X be a set and let $S \subseteq X \times X$ be an arbitrary subset. Let $f : X \to Y$ be a function satisfying $f(x_1) = f(x_2)$ if $x_1 \sim_S x_2$. Show that f factors uniquely through $\pi : X \to X/S$, i.e., there exists a unique function $g : X/S \to Y$ such that $f = g \circ \pi$.

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\pi"'] & Y \\
X/S \arrow[ur, dashed, "g"'] &
\end{tikzcd}$$

Equivalence relations can be used to construct many different objects. Here is the dual version of Exercise 1.3.B.

Exercise 1.4.F. Fix sets A, B, C and maps $\alpha : C \to A$ and $\beta : C \to B$. Recall that there are inclusion maps $\iota_1 : A \to A \amalg B$ and $\iota_2 : B \to A \amalg B$. Define the set

$$P = A \amalg B/(\iota_1(\alpha(c)) \sim \iota_2(\beta(c)) \text{ for all } c \in C)$$

which will come with maps

$$j_1 : A \xrightarrow{\iota_1} A \amalg B \to P, \quad j_2 : B \xrightarrow{\iota_2} A \amalg B \to P.$$

Show that for arbitrary maps $f : A \to W$ and $g : B \to W$ satisfying $f \circ \alpha = g \circ \beta$, there exists a unique map $h : P \to W$ such that $f = h \circ j_1$ and $g = h \circ j_2$.

$$\begin{tikzcd}
C \arrow[r, "\beta"] \arrow[d, "\alpha"] & B \arrow[d, "j_2"] \arrow[ddr, "g"] & \\
A \arrow[r, "j_1"] \arrow[drr, "f"'] & P \arrow[dr, dashed, "h"] & \\
& & W
\end{tikzcd}$$

(This P is called the fiber coproduct of A and B over C.)

Exercise 1.4.G. Let's try to construct the integers $\mathbb{Z}$ from the natural numbers $\mathbb{Z}_{\geq 0}$. Consider the relation

$$(a, b) \sim (c, d) \quad \Leftrightarrow \quad a + d = b + c$$

on $\mathbb{Z}^2_{\geq 0}$. Check that this is an equivalence relation. There is a map $s : \mathbb{Z}^2_{\geq 0} \to \mathbb{Z}$ given by $(a, b) \mapsto a - b$, and check that $(a, b) \sim (c, d)$ implies $s(a, b) = s(c, d)$. This shows that the map s factors as

$$s : \mathbb{Z}^2_{\geq 0} \xrightarrow{\pi} (\mathbb{Z}^2_{\geq 0}/\sim) \xrightarrow{t} \mathbb{Z}.$$

Show that this map t is a bijection. This shows that we may construct the integers $\mathbb{Z}$ as taking the quotient of $\mathbb{Z}^2_{\geq 0}$ by this equivalence relation.

Exercise 1.4.H. We can similarly construct $\mathbb{Q}$ from $\mathbb{Z}$. Consider the relation

$$(a, b) \sim (c, d) \quad \Leftrightarrow \quad ad = bc$$

on the set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, and check that this is an equivalence relation. Show that we can identify $\mathbb{Q}$ with the quotient $(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\sim$.

Chapter 2

Vector spaces

Linear algebra is the study of linear structures. Historically, the motivation for the subject was solving systems of linear equations. Given an equation like

$$\begin{cases} 3x + y = 5, \\ x + 3y = 7, \end{cases}$$

how do we solve it? This question led mathematicians such as Leibniz, Cramer, and Gauss to study properties of systems of linear equations. Later, the notion of matrices was introduced to record and manipulate equations more easily. For examply, the above equation could be written as

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

using matrix notation. Then multiply the inverse matrix from the left on both sides to solve the equation.

But such a concrete point of view is not always helpful in developing a theory. When you know too much about an object, it is easy to get lost in the pile of information. Abstraction is supposed to deal with such problems by taking away some, but not too much, information so that only the essential ones stand out.

In the abstract approach to linear algbera, we study mathematical objects called vector spaces. A vector space is a set with a linear structure, as we shall define it in Section 2.2. We will try to understand these objects and the maps between them, look at ways to construct new vector spaces, and eventually classify all of them.

2.1 Fields

To definite a linear structure, we first need to choose a “base” for developing the notion of linearity . In Euclidean geometry, we work over the real numbers $\mathbb{R}$, roughly meaning that the notion of linearity works over $\mathbb{R}$. For instance, we can take a vector $v \in \mathbb{R}^2$ in the Euclidean plane and multiply it by any real number. A field is supposed be this scalar number system to which we can multiply a vector.

A field is a set with addition, subtraction, multiplication, and division

Definition 2.1.1. A **field** k is a set with the choice of two elements[1] $0, 1 \in k$ and two maps $+, \cdot : k \times k \to k$ satisfying the following conditions: (Here, we denote $+(a,b) = a + b$ and $\cdot(a,b) = a \cdot b$ because they act like addition and multiplication.)

(F0) $0 \neq 1$.

(F1) For all $a \in k$ we have $a + 0 = 0 + a = a$.

(F2) For all $a, b \in k$ we have $a + b = b + a$.

(F3) For all $a, b, c \in k$ we have $a + (b + c) = (a + b) + c$.

(F4) For all $a \in k$ there eaists an $(-a) \in k$ such that $a + (-a) = (-a) + a = 0$.

(F5) For all $a \in k$ we have $a \cdot 1 = 1 \cdot a = a$

(F6) For all $a, b \in k$ we have $a \cdot b = b \cdot a$.

(F7) For all $a, b, c \in k$ we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

(F8) For all $a \in k$ with $a \neq 0$ there eaists an $a^{-1} \in k$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

(F9) For all $a, b, c \in k$ we have $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

The first axiom (F0) is stating that the field is not degenerate, i.e., has at least two elements. The next axioms (F1) to (F4) tell us about how addition behaves. Addition should have an identity element called 0, be commutative, be associative, and have inverses. The axioms (F5) to (F8) state analogous properties for multiplication. One difference is that 0 need not have a multiplicative inverse. The last axiom (F9) states that multiplication distributes over addition.

Example 2.1.2. The set of rational numbers $\mathbb{Q}$ is a field under usual addition and multiplication. The set of real numbers $\mathbb{R}$ and the set of complex numbers $\mathbb{C}$ are also fields under normal addition and multiplication.

Example 2.1.3. Let p be a prime number. We consider the set

$$\mathbb{F}_p = \{0, 1, 2, \ldots, p - 1\},$$

and define addition and multiplication modulo p. Then $\mathbb{F}_p$ is a field.

Exercise 2.1.A. Let k be a field, and let $a, b, c \in k$. Show that $a + c = b + c$ implies $a = b$. Show that $a \cdot c = b \cdot c$ and $c \neq 0$ imply $a = b$. Show that $a \cdot b = 0$ implies $a = 0$ or $b = 0$.

Multiplication by zero makes everything to zero

[1] Here, 0 and 1 are not actually the integers 0 and 1. I could have used the symbols s and t to denote them if I wanted, but they play the role of 0 and 1. So it is helpful to indicate this fact by using the notations 0 and 1.

Exercise 2.1.B. Let k be a field. Show that for all $x \in k$, we have $x \cdot 0 = 0$. (Hint: multiply both sides of $0+0=0$ by x.) So 0 cannot have a multiplicative inverse, because of (F0).

Exercise 2.1.C. Consider the subset

$$\mathbb{Q}(i) = \{a + bi : a, b \in \mathbb{Q}\} \subseteq \mathbb{C}.$$

Check that $\mathbb{Q}(i)$ is a field under normal addition and multiplication. (Here, you also need to verify that $x, y \in \mathbb{Q}(i)$ implies $x + y, xy \in \mathbb{Q}(i)$, i.e., that addition and multiplication are well-defined.)

Sometimes, like in $\mathbb{F}_p$, adding 1 many times can get you back to 0.

Definition 2.1.4. Let k be a field. If the sequenece $1, 1+1, 1+1+1, \ldots$ never reaches 0, we say that k has **characteristic** zero, and write $\operatorname{char} k = 0$. If one of $1 + \cdots + 1$ is equal to 0, we define its **characteristic** $\operatorname{char} k$ as the least positive integer n such that

$$\overbrace{1 + \cdots + 1}^{n} = 0.$$

For instance, $\operatorname{char} \mathbb{Q} = 0$ and $\operatorname{char} \mathbb{C} = 0$ but $\operatorname{char} \mathbb{F}_p = p$.

Exercise 2.1.D. Show that the characteristic of a field is either 0 or a prime number.

Every characteristic zero field contains the rationals

Exercise 2.1.E. Show that if $\operatorname{char} k = 0$ for a field k, then there exists a map $i : \mathbb{Q} \to k$ satisfying $i(0) = 0$, $i(1) = 1$, $i(a+b) = i(a)+i(b)$, and $i(ab) = i(a)i(b)$ for all $a, b \in k$.

2.2 Vector spaces

We are now ready to define vector spaces. This notion is supposed to capture a linear structure, such as what a scaling means, or what a linear function on this space is.

A vector space is a set with addition and scalar multiplication

Definition 2.2.1. Fix a field k. A **vector space over** k, or a k**-vector space** is a set V with the choice of an element $0 \in V$ and two maps $+ : V \times V \to V$ and $\cdot : k \times V \to V$ satisfying the following conditions:

(VS1) For all $v \in V$ we have $v + 0 = 0 + v = v$.

(VS2) For all $v, w \in V$ we have $v + w = w + v$.

(VS3) For all $v, w, u \in V$ we have $(v + w) + u = v + (w + u)$.

(VS4) For all $v \in V$ there exists an $(-v) \in V$ such that $v + (-v) = (-v) + v = 0$.

(VS5) For all $v \in V$ we have $0 \cdot v = 0$ and $1 \cdot v = v$.

(VS6) For all $v \in V$ and $a, b \in k$ we have $(a \cdot b) \cdot v = a \cdot (b \cdot v)$.

(VS7) For all $v \in V$ and $a, b \in k$ we have $(a + b) \cdot v = a \cdot v + b \cdot v$.

(VS8) For all $v, w \in V$ and $a \in k$ we have $a \cdot (v + w) = a \cdot v + a \cdot w$.

An element of a vector space is called a **vector**. (The notion of a vector only makes sense when a vector space is given.)

This might look similar to the definition of a field, but an important difference is that multiplication is not defined between two vectors. We only have the notion of multiplication by an element of k, that is, by a scalar. Let us look at some examples.

Example 2.2.2. The space

$$k^3 = \{(a_1, a_2, a_3) : a_1, a_2, a_3 \in k\}$$

is a vector space over k. Here, addition and multiplication is defined as

$$(a_1, a_2, a_3)+(b_1, b_2, b_3) = (a_1+b_1, a_2+b_2, a_3+b_3), \quad a\cdot(b_1, b_2, b_3) = (ab_1, ab_2, ab_3),$$

and the zero vector is $0 = (0, 0, 0)$. You can easily check that this satisfies all the axioms. If $k = \mathbb{R}^3$, this is the usual 3-dimensional Euclidean space, where you can add vectors, or multiply vectors by real numbers. But here, we don't have dot products or cross products.

Example 2.2.3. More generally, let $n \geq 0$ be an integer. We define

$$k^n = \{(a_1, \dots, a_n) : a_i \in k\}$$

with addition and multiplication

$$(a_1, \dots, a_n)+(b_1, \dots, b_n) = (a_1+b_1, \dots, a_n+b_n), \quad a\cdot(b_1, \dots, b_n) = (ab_1, \dots, ab_n).$$

This is a vector space.

Example 2.2.4. What is k^0? Because a n-tuple is a map from $\{1, \dots, n\}$ to k, the vector space k^0 consists of maps $\emptyset \to k$, of which there is exactly one. So we can write

$$k^0 = 0 = \{0\}.$$

By abuse of notation, we shall call this vector space as 0.[2]

Example 2.2.5. What is k^1? This is going to be

$$k = k^1 = \{a : a \in k\},$$

as a vector space. When we regard k as a vector space, it forgets all about multiplication within itself.

Example 2.2.6. Consider $\mathbb{R}$, but now we are going to look at it as a vector space over $\mathbb{Q}$. That is, we consider the usual addition $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and usual multiplication $\cdot : \mathbb{Q} \times \mathbb{R} \to \mathbb{R}$. This defines $\mathbb{R}$ as a $\mathbb{Q}$-vector space.

[2] So far, we have three kinds of 0. The symbol can mean $0 \in k$, or $0 \in V$, or the 0 vector space. Make sure you do not confuse them.

Example 2.2.7. Here is a crazier example. Take

$$C([0,1]) = \{\text{continuous functions } [0,1] \to \mathbb{R}\},$$

and consider it as a $\mathbb{R}$-vector space. Addition and multiplication is defined as

$$(f+g)(x) = f(x) + g(x), \quad (a \cdot f)(x) = a \cdot f(x),$$

where this makes sense because the sum of two continuous functions and a continuous function times a constant are again continuous. This satisfies all the axioms, so $C([0,1])$ is a vector space over $\mathbb{R}$.

Exercise 2.2.A. Show that the axiom (VS4) is actually redundant. That is, it can be derived from the other axioms.

So we have defined this mathematical object called a vector space, which possesses a linear structure over k. Very often, when you define a mathematical object with structure, you want to also look at maps between the objects that behave well with respect to their structures.

A linear map is a map preserving addition and scalar multiplication

Definition 2.2.8. Let V and W be two k-vector spaces. We say that a map $T : V \to W$ is k-**linear** or a **linear transformation**, or simply a **homomorphism** if it satisfies:

(L0) $T(0) = 0$, where the first 0 is $0 \in V$ and the second 0 is $0 \in W$.

(L1) For all $v_1, v_2 \in V$ we have $T(v_1 + v_2) = T(v_1) + T(v_2)$.

(L2) For all $a \in k$ and $v \in V$ we have $T(av) = aT(v)$.

Exercise 2.2.B. Show that the identity map $\mathrm{id}_V : V \to V$ is always k-linear. Show also that the zero map $0 : V \to W$ given by $v \mapsto 0$ is also always k-linear.

Exercise 2.2.C. Show that the first axiom (L0) is redundant.

Linear maps are closed under composition

Exercise 2.2.D. Let $T : V \to W$ and $S : W \to U$ be k-linear maps. Show that the composite map $S \circ T : V \to U$ is also k-linear.

Exercise 2.2.E. Let V be a vector space over k. Show that there exists a unique k-linear map $V \to 0$ and also a unique k-linear map $0 \to V$. (Here 0 should mean a k-vector space!)

The Cauchy functional equation is a linearity condition over the rationals

Exercise 2.2.F. If you have every had to study functional equations in the context of mathematical olympiads, you must have encountered what is called that **Cauchy functional equation**. For a function $f : \mathbb{R} \to \mathbb{R}$, we say that the function satisfies the Cauchy equation if

$$f(x+y) = f(x) + f(y)$$

for all $x, y \in \mathbb{Q}$. Show that f satisfies the Cauchy equation if and only if f is $\mathbb{Q}$-linear, where $\mathbb{R}$ is regarded as a $\mathbb{Q}$-vector space.

Sometimes we can have two vector spaces that are set-theoretically different, but similar-looking so that we do not wish to distinguish them. For instance, suppose we have a vector space $\{(0,0,0)\}$ with one element. This looks just like the zero vector space $0 = \{0\}$, but technically the two are not equal.

Definition 2.2.9. Let V and W be k-vector spaces. A k-linear map $T : V \to W$ is called an **isomorphism** if there exists a map $S : V \to W$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. If there exists an isomorphism $V \to W$, we say that V and W are **isomorphic** and write $V \cong W$.

Exercise 2.2.G. If $T : V \to W$ and $S : W \to U$ are isomorphisms, show that $S \circ T : V \to U$ is also an isomorphism. Deduce that "isomorphic" is an equivalence relation.

An isomorphism is a bijective linear map

Exercise 2.2.H. Show that a k-linear map $V \to W$ is an isomorphism if and only if it is bijective.

Our initial goal in developing linear algebra will be to classify all vector spaces. This means that we want to produce a "list" of all vector spaces. But there are too many vector spaces if we interpret this goal literally. Instead, we are going to satisfied with producing a list such that every vector space is isomorphic to one of the vector spaces on our list. After all, we do not want to distinguish isomorphic vector spaces apart.

Goal. *Classify all vector spaces up to isomorphism.*

2.3 Matrices

Before initiating the grand project of classifying vector spaces, I would like to talk about matrices, which are generally presented to be the main objects of study in introductory linear algebra textbooks.

Definition 2.3.1. A $m \times n$ **matrix** with entries in k is a k-linear map $k^n \to k^m$, where k^n and k^m are regarded as k-vector spaces.

What does this mean? A matrix is supposed to be an array with numbers, not a linear map. But let us try to think about how we can encode a k-linear map $k^n \to k^m$.

As a warm-up, let us try to classify all k-linear maps $T : k \to k$, where both k are considered as k-vector spaces. If we demand that $T(1) = c$ for some $c \in k$, then we get

$$T(a) = T(a \cdot 1) = a \cdot T(1) = a \cdot c = ac$$

for all $a \in k$. That is, the value of $T(1)$ completely determines T. On the other hand, once we choose $T(1) = c$ arbitrarily, the map $T : a \mapsto ac$ is indeed a linear map, because

$$T(a+b) = (a+b)c = ac + bc = T(a) + T(b), \quad T(a \cdot b) = abc = a(bc) = aT(b).$$

This shows that there is a bijection

$$\{k\text{-linear maps } k \to k\} \quad \longleftrightarrow \quad k; \qquad T \mapsto T(1),$$

and the linear maps $k \to k$ are completely classified by k.

Linear maps from the field to a vector space is classified by the vector space

Exercise 2.3.A. Let V be an arbitrary vector space. Show that the linear maps $T : k \to V$ are classified by elements V. That is, exhibit a bijection

$$\{k\text{-linear maps } k \to V\} \quad \longleftrightarrow \quad V.$$

Let us now try to classify linear maps $k^n \to k^m$. Here, we are going to use the notation

$$e_1 = (1, 0, \dots, 0), \quad e_2 = (0, 1, 0, \dots, 0), \quad \dots .$$

So, for instance, we can write

$$(a_1, a_2, \dots, a_n) = a_1 e_1 + a_2 e_2 + \dots + a_n e_n \in k^n.$$

If $T : k^n \to k^m$ is a linear map, for each $1 \leq j \leq n$ we get an element $T(e_j) \in k^m$. Let us write

$$T(e_j) = (t_{1j}, t_{2j}, \dots, t_{mj}) = \sum_{i=1}^{m} t_{ij} e_i.$$

From a k-linear map $T : k^n \to k^m$, we have thus extracted mn numbers $t_{ij} \in k$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. These mn numbers completely determine T, because

$$T((a_1, \dots, a_n)) = T\biggl(\sum_{j=1}^{n} a_j e_j\biggr) = \sum_{j=1}^{n} a_j T(e_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} a_j t_{ij} e_i.$$

Conversely, for arbitrary numbers $t_{ij} \in k$, the map defined above is k-linear, because

$$\begin{aligned} T((a_1 + b_1, \dots, a_n + b_n)) &= \sum_{j=1}^{n} \sum_{i=1}^{m} (a_j + b_j) t_{ij} e_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} a_j t_{ij} e_i + \sum_{j=1}^{n} \sum_{i=1}^{m} b_j t_{ij} e_i \\ &= T((a_1, \dots, a_n)) + T((b_1, \dots, b_n)) \end{aligned}$$

and

$$T(a(b_1, \dots, b_n)) = \sum_{j=1}^{n} \sum_{i=1}^{m} (ab_j) t_{ij} e_i = a \sum_{j=1}^{n} \sum_{i=1}^{m} b_j t_{ij} e_i = aT((b_1, \dots, b_n)).$$

Hence the linear map $T : k^n \to k^m$ is uniquely determined by the numbers t_{ij} and we get a bijection

$$\{k\text{-linear maps } T : k^n \to k^m\} \quad \longleftrightarrow \quad \{mn \text{ numbers } t_{ij} \in k\}. \qquad (*)$$

To keep track of all the numbers, we are going to introduce a new notation. Let us write

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix}$$

when a linear map $T : k^n \to k^m$ on the left hand side of $(*)$ corresponds to $t_{ij} \in k$ on the right hand side. With this notation, we are going to write

$$T((a_1, \ldots, a_n)) = (b_1, \ldots, b_m)$$

as

$$\begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

These are all equivalent to saying that

$$b_i = \sum_{j=1}^{n} t_{ij} a_j,$$

but with the new notation, we can write out the equations explicitly in a single equation.

Remark 2.3.2. The notation seems to suggest that $(a_1, \ldots, a_n)$ is a $1 \times n$ matrix, rather than a vector in k^n. In a sense, this is true. Exercise 2.3.A shows that an element of a vector space can be canonically identified with a map from k to the vector space. So an element of k^n can be considered as a k-linear map $k \to k^n$, which is a $1 \times n$ matrix. If this is confusing, don't worry about it.

Exercise 2.3.B. Show that the identity map $\mathrm{id} : k^n \to k^n$ corresponds to the matrix

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Composition of linear maps corresponds to multiplication of matrices

Exercise 2.3.C. Let $S : k^n \to k^m$ and $T : k^m \to k^p$ be two matrices. (S will be $m \times n$ and T will be $p \times m$.) The composite $T \circ S = U : k^n \to k^p$ is going to be k-linear, and hence a matrix. Let t_{ij}, s_{jl}, u_{il} be the matrix entries corresponding to the linear maps T, S, U. Then show that

$$u_{il} = \sum_{j=1}^{m} t_{ij} s_{jl}.$$

If you have learned about matrix multiplications before, check that this agrees with the ordinary way of multiplying matrices.

So from the perspective of linear maps, multiplication of matrices is nothing other than composition of linear maps. If you have been bewildered by the fact that matrix multiplication is associative, that is,

$$A(BC) = (AB)C$$

for all matrices A, B, and C, now you know that this is because composition of maps is associative.

2.4 Products and direct sums

In this section, we are going to talk about constructing vector spaces. We have already seen the example of k^n for n a nonnegative integer. We can generalized this to when n is not a finite number, but infinite.

Definition 2.4.1. Let S be a set. We define

$$k^S = \{\text{set of maps } S \to k\}$$

with addition and scalar multiplication

$$(f+g)(s) = f(s) + g(s), \quad (a \cdot f)(s) = a \cdot f(s)$$

for $f, g \in k^S$, $s \in S$, and $a \in k$.

Exercise 2.4.A. Check that this indeed a k-vector space. What is the 0 vector?

Of course, if $S = \{1, 2, \ldots, n\}$ we retrieve the vector space k^n. But if S is much larger, this vector space can be pretty huge. An interesting feature of this construction is that it is functorial. That is, if I have two sets S_1, S_2 and a map $\alpha : S_1 \to S_2$ of sets, then I get a map

$$\alpha^* : k^{S_2} \to k^{S_1}; \quad (f : S_2 \to k) \mapsto (f \circ \alpha : S_1 \to k).$$

It is clear from the definition that if $\alpha : S_1 \to S_2$ and $\beta : S_2 \to S_3$ are maps between sets, then $(\beta \circ \alpha)^* = \alpha^* \circ \beta^*$.

Exercise 2.4.B. Show that α^* is always k-linear.

Although this construction is a natural generalization of k^n, there is a slightly different construction that is more important in studying vector spaces.

Definition 2.4.2. Let S be a set. We define the **free vector space on** S as

$$k^{\oplus S} = \{\text{set of maps } f : S \to k \text{ such that } f(s) \neq 0 \text{ only for finitely many } s\},$$

with addition and scalar multiplication

$$(f+g)(s) = f(s) + g(s), \quad (a \cdot f)(s) = a \cdot f(s)$$

for $f, g \in k^{\oplus S}$, $s \in S$, and $a \in k$.

Clearly, if S is finite, the two vector spaces k^S and $k^{\oplus S}$ agree, because anyhow there can be only finitely many s such that $f(s) \neq 0$. But if S is infinite, the vector space $k^{\oplus S}$ is strictly contained in k^S.

Exercise 2.4.C. Check that this is indeed a k-vector space. (Here you need to show that if f and g satisfy the "finiteness of nonzero values" condition, then $f + g$ and $a \cdot f$ satisfy the condition as well.)

This constructions deserves more explanation, as it might seem unmotivated to only take functions with finite nonzero values. For each $s \in S$, there is an function $S \to k$ that sends s to $1 \in k$ and all other elements to $0 \in k$. This satisfies the finiteness condition, and hence is an element of $k^{\oplus S}$. We shall denote it by $\underline{s} \in k^{\oplus S}$, so that

$$\underline{s}(s') = \begin{cases} 1 & s' = s \\ 0 & s' \neq s. \end{cases}$$

These vectors are supposed to be like $e_i \in k^n$. For pairwise distinct elements $s_1, \ldots, s_m \in S$ and nonzero scalars $a_1, \ldots, a_m \in k$, we can form the sum

$$v = a_1 \underline{s_1} + a_2 \underline{s_2} + \cdots + a_m \underline{s_m} \in k^{\oplus S}.$$

The vector v, regarded as a map $S \to k$, sends s_j to a_j and all other $s \in S$ to 0. In fact, every vector $v \in k^{\oplus S}$ can be written like this sum because v is nonzero only on a finite number of s. Moreover such a presentation of v as a linear combination of $\underline{s}$ is unique. So this is the "freest" vector space you can get by regarding elements of S as symbols. (You're never allowed to add infinitely many vectors at once, because addition is only defined on two elements.)

This vector space enjoys a universal property. For any set S, there exists a natural (set) map

$$\iota : S \to k^{\oplus S}; \quad s \mapsto \underline{s}.$$

(Because S is a set, it does not make sense to ask if this map is linear or not.) The map ι is always injective, because $s \neq s'$ implies $\underline{s} \neq \underline{s'}$. (To see this, note that $\underline{s}(s) = 1$ while $\underline{s'}(s) = 0$.)

A map from a set to a vector space is the same as a linear map from the free vector space

Proposition 2.4.3 (Universal property for free vector spaces)**.** *Let S be a set, and let V be an arbitrary k-vector space. If $f : S \to V$ is any (set) map, there exists a unique linear map $T : k^{\oplus S} \to V$ that extends f, that is, $f = T \circ \iota$.*

$$\begin{tikzcd} S \arrow[r, "f"] \arrow[d, hook, "\iota"'] & V \\ k^{\oplus S} \arrow[ur, dashed, "T"'] & \end{tikzcd}$$

The intuition is that once we choose where $\underline{s}$ are sent into V, it is also uniquely determined where their finite linear combinations $a_1\underline{s_1}+\cdots+a_n\underline{s_n}$ are sent to $a_1 f(s_1)+\cdots+a_n f(s_n)$.

Proof. We first prove uniqueness. Recall that an element of $k^{\oplus S}$ is formally a map $S \to k$ that is 0 except on a finite number of elements of S. Then we can write any element $\alpha \in k^{\oplus S}$ as

$$\alpha = \sum_{s\in S} \alpha(s)\underline{s},$$

where $\alpha(s) \neq 0$ for finitely many s ensures that the right hand side is a finite sum. Suppose that a linear map $T : k^{\oplus S} \to V$ satisfies the condition $f = T \circ \iota$. From this condition, we see that $f(s) = T(\iota(s)) = T(\underline{s})$. So by linearity of T, we get

$$T(\alpha) = T\left(\sum_{s\in S} \alpha(s)\underline{s}\right) = \sum_{s\in S} \alpha(s)T(\underline{s}) = \sum_{s\in S} \alpha(s)f(s).$$

That is, T is uniquely determined to be function.

To show existence, we only need to check that the T we defined above is indeed linear and satisfies the condition $f = T\circ\iota$. This can be easily checked. □

Another way to state this proposition is that there is a natural correspondence

$$\mathrm{Mor}_{\mathsf{Set}}(S, V) = \{\text{set maps } S \to V\} \quad \longleftrightarrow \quad \{\text{linear maps } k^{\oplus S} \to V\},$$

where one direction of the correspondence is given by composing with the map $\iota : S \to k^{\oplus S}$.

We have previously seen that a set map $\alpha : S_1 \to S_2$ induced a map

$$\alpha^* : k^{S_2} \to k^{S_1}; \quad (f : S_2 \mapsto k) \mapsto (f \circ \alpha : S_1 \to k)$$

in the other direction. Can this construction be done for $k^{\oplus S_2} \to k^{\oplus S_1}$ as well?

Exercise 2.4.D. Show that the map

$$\alpha^* : k^{\oplus S_2} \to k^{\oplus S_1}; \quad (f : S_2 \to k) \mapsto (f \circ \alpha : S_1 \to k)$$

is *not* a well-defined map. What is the problem here?

But we can do the following. Consider the map

$$c = \iota_{S_2} \circ \alpha : S_1 \to k^{\oplus S_2}; \quad s \mapsto \underline{\alpha(s)}.$$

By Proposition 2.4.3, this uniquely induces a linear map

A map of sets induces a linear map on the free vector spaces

$$\alpha_* : k^{\oplus S_1} \to k^{\oplus S_2}$$

such that $c = \alpha_* \circ \iota_{S_1}$, i.e., $\alpha_*(\underline{s}) = \underline{\alpha(s)}$. This can also be drawn as

$$\begin{tikzcd}
S_1 \arrow[r, "\alpha"] \arrow[d, "\iota_{S_1}"] \arrow[dr, "c"] & S_2 \arrow[d, "\iota_{S_2}"] \\
k^{\oplus S_1} \arrow[r, dashed, "\alpha_*"] & k^{\oplus S_2}.
\end{tikzcd}$$

Exercise 2.4.E. Check that if $\alpha : S_1 \to S_2$ and $\beta : S_2 \to S_3$ are (set) maps, then $(\beta \circ \alpha)_* = \beta_* \circ \alpha_*$.

Let us now consider two k-vector spaces V and W. We can look at their product

$$V \times W = \{(v, w) : v \in V, w \in W\}$$

as sets, but this set can naturally be given a structure of a k-vector space.

Definition 2.4.4. Given two vector spaces V and W, we define their **product** as $V \times W$ with addition and multiplication

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2), \quad c(v, w) = (cv, cw).$$

Exercise 2.4.F. Check that this satisfies all the axioms of a vector space. What is the 0 vector?

For instance, $k^2 \times k^3$ can naturally be identified with k^5 via the correspondence

$$((a_1, a_2), (b_1, b_2, b_3)) \leftrightarrow (a_1, a_2, b_1, b_2, b_3).$$

As with sets, we can also define infinite products in the exact same way.

Definition 2.4.5. Let $\{V_i\}_{i \in I}$ be a set of vector spaces, where I is the indexing set. We define their **product** as

$$\prod_{i \in I} V_i = \{(v_i)_{i \in I} : v_i \in V_i\}$$

with addition and multiplication

$$(v_i)_{i \in I} + (w_i)_{i \in I} = (v_i + w_i)_{i \in I}, \quad c(v_i)_{i \in I} = (cv_i)_{i \in I}.$$

Again, it is straightforward to check that this satisfies all the axioms of a vector space, and is thus a vector space.

Exercise 2.4.G. For S a set, show that k^S is naturally isomorphic to $\prod_{s \in S} k$. So this product is a generalization of the previous construction.

For each $i_0 \in I$, there are (set) maps

$$\pi_{i_0} : \prod_{i\in I} V_i \to V_{i_0}; \quad (v_i)_{i\in I} \mapsto v_{i_0},$$

and furthrmore they are also linear. This is because

$$\pi_{i_0}((v_i) + (w_i)) = \pi_{i_0}((v_i + w_i)) = v_{i_0} + w_{i_0} = \pi_{i_0}((v_i)) + \pi_{i_0}((w_i)),$$
$$\pi_{i_0}(c(v_i)) = \pi_{i_0}((cv_i)) = cv_{i_0} = c\pi_{i_0}((v_i)).$$

Linear maps to vector spaces is the same as a linear map to their product

Proposition 2.4.6 (Universal property for products)**.** *Let $\{V_i\}_{i\in I}$ be a set of k-vector spaces. For an arbitrary k-vector space W and linear maps $f_i : W \to V_i$, there exist a unique linear map $f : W \to \prod_{i\in I} V_i$ such that $f_{i_0} = \pi_{i_0} \circ f$ for each $i_0 \in I$.*

$$\begin{tikzcd}
W \arrow[r, dashed, "f"] \arrow[dr, "f_{i_0}"'] & \prod_{i\in I} V_i \arrow[d, "\pi_{i_0}"] \\
 & V_{i_0}
\end{tikzcd}$$

Proof. First we check that there exist a unique set map $f : W \to \prod_{i\in I} V_i$ such that $f_{i_0} = \pi_{i_0} \circ f$. This just a general version of Proposition 1.3.1, which an be verified similarly. Here, the map f is going to be defined as

$$f : w \mapsto (f_i(w))_{i\in I}.$$

Now it suffices to check that $w \mapsto (f_i(w))$ is indeed linear. This is also clear, because the maps f_i are all linear. □

Exercise 2.4.H. Fill in the gaps of the previous proof.

This really is the vector space version of Proposition 1.3.1. If $I = \{1, 2\}$, we can draw the diagram in the following way as well.

$$\begin{tikzcd}
 & W \arrow[d, dashed, "f"] \arrow[ddl, "f_1"'] \arrow[ddr, "f_2"] & \\
 & V_1 \times V_2 \arrow[dl, "\pi_1"] \arrow[dr, "\pi_2"'] & \\
V_1 & & V_2
\end{tikzcd}$$

We can also consider the subspace of the product space consisting only of $(v_i)_{i\in I}$ such that $v_i \neq 0$ only for finitely many I. This will be analogue of $k^{\oplus S}$.

Definition 2.4.7. Let $\{V_i\}_{i\in I}$ be a set of k-vector spaces. We define their **direct sum**

$$\bigoplus_{i\in I} V_i = \{(v_i)_{i\in I} : v_i \in V_i, v_i \neq 0 \text{ only for finitely many } i \in I\}$$

with addition and scalar multiplication

$$(v_i) + (w_i) = (v_i + w_i), \quad c(v_i) = (cv_i).$$

Exercise 2.4.I. Check again that this is a vector space.

Exercise 2.4.J. For S a set, show that $\bigoplus_{s\in S} k$ is naturally isomorphic to $k^{\oplus S}$.

For each $i_0 \in I$, there is a map

$$\iota_{i_0} : V_{i_0} \to \bigoplus_{i\in I} V_i; \quad v \mapsto (v_i)_{i\in I} \text{ where } v_i = \begin{cases} v \in V_{i_0} & i = i_0 \\ 0 \in V_i & i \neq i_0. \end{cases}$$

This map is clearly linear, because it is linear in each component.

Linear maps from vector spaces is the same as a linear map from their direct sum

Proposition 2.4.8 (Universal property for direct sums)**.** *Let $\{V_i\}_{i\in I}$ be a set of k-vector spaces. For an arbitrary k-vector space W and linear maps $f_i : V_i \to W$, there exists a unique linear map $f : \bigoplus_{i\in I} V_i \to W$ such that $f_{i_0} = f \circ \iota_{i_0}$ for each $i_0 \in I$.*

$$\begin{tikzcd}
V_{i_0} \arrow[d, "\iota_{i_0}"'] \arrow[dr, "f_{i_0}"] & \\
\bigoplus_{i\in I} V_i \arrow[r, dashed, "f"'] & W
\end{tikzcd}$$

Proof. For uniqueness, the condition $f_{i_0} = f \circ \iota_{i_0}$ forces

$$f((v_i)) = f\left(\sum_{i\in I} \iota_i(v_i)\right) = \sum_{i\in I} f(\iota_i(v_i)) = \sum_{i\in I} f_i(v_i).$$

Here, the sums are finite because $v_i \neq 0$ for only finitely many i. To show existence, it suffice to check that the above map is linear, but this is straightforward to check. □

Exercise 2.4.K. Check the details in the above proof.

Again, if $I = \{1, 2\}$ the diagram can be drawn as

$$\begin{tikzcd}
V_1 \arrow[dr, "\iota_1"] \arrow[ddr, "f_1"'] & & V_2 \arrow[dl, "\iota_2"'] \arrow[ddl, "f_2"] \\
& V_1 \oplus V_2 \arrow[d, dashed, "f"] & \\
& W. &
\end{tikzcd}$$

In view of Exercise 1.3.A, we see that direct sum $\oplus$ for vector spaces works precisely as disjoint union $\amalg$ for sets. Hence the direct sum is also be called the **coproduct** of vector spaces.

Coproducts commute with taking free vector spaces

Exercise 2.4.L. Let $\{S_i\}_{i\in I}$ be a set of sets. Show that there is a canonical isomorphism

$$\bigoplus_{i\in I} k^{\oplus S_i} \cong k^{\oplus(\amalg_{i\in I} S_i)}.$$

(Hint: Show that both satisfy the same universal property. In other words, using universal properties, construct maps in both directions.)

Exercise 2.4.M. Consider linear maps $T : k^n \to k^m$ and $S : k^n \to k^l$. From Proposition 2.4.6, they induce a linear map $M : k^n \to k^m \times k^l \cong k^{m+l}$. When all these maps are considered as matrices, show that

$$M = \begin{bmatrix} T \\ S \end{bmatrix}.$$

Exercise 2.4.N. Similarly, consider linear maps $T : k^n \to k^l$ and $S : k^m \to k^l$. According to Proposition 2.4.8, they induce a linear map $M : k^{n+m} \cong k^n \oplus k^m \to k^l$. When all these maps are considered as matrices, show that

$$M = \begin{bmatrix} T & S \end{bmatrix}.$$

2.5 Subspaces and quotients

When we defined $k^{\oplus S}$, we defined it as a subset of k^S, and then gave the exactly same structure. We can consider this procedure in a more general setting.

Exercise 2.5.A. Let V be a k-vector space. Suppose that a subset $W \subseteq V$ satisfies

(SS1) $0 \in W$,

(SS2) $v_1, v_2 \in W$ implies $v_1 + v_2 \in W$,

(SS3) $v \in W$ implies $cv \in W$ for all $c \in k$.

Then W inherits a k-vector space structure from V, with exactly the same addition and scalar multiplication.

Definition 2.5.1. In such a case, we say that W is a **subspace** of V.

For instance, $k^{\oplus S}$ is a subspace of k^S consisting of maps $f : S \to k$ such that $f(s) \neq 0$ for only finitely many s.

An injective linear map is an isomorphism onto a subspace

Exercise 2.5.B. Show that any injective linear map $T : W \hookrightarrow V$ gives an isomorphism between W and the image $T(W)$ as a subspace of V. Conversely, if W is a subspace of V, show that the natural inclusion map $i : W \to V$ is an injective linear map.

If $W, U \subseteq V$ are subspaces, we can define

$$W + U = \{w + u : w \in W, u \in U\} \subseteq V.$$

Exercise 2.5.C. Show that if $W, U \subseteq V$ are subspaces, then $W + U$ and $W \cap U$ are also subspaces.

Exercise 2.5.D. The subspace sum acts like a direct sum if the intersecion is trivial Let $W, U \subseteq V$ be subspaces. The two inclusion maps $W, U \hookrightarrow W + U$ induce a linear map $W \oplus U \to W + U$ be the universal property of the direct sum. Show that this map is always surjective. If $W \cap U = \{0\}$ then show that this map is an isomorphism. In such a case, we shall abuse notation to sometimes write $W \oplus U \subseteq V$ instead of $W + U$.

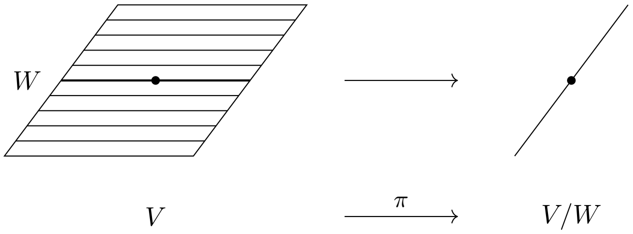

Figure 2.1: Quotienting V by W

Exercise 2.5.E. Show that it is false that if $W, U, X \subseteq V$ are subspaces, then $(W+U) \cap X = (W \cap X) + (U \cap X)$. Give a counterexample, and show that one always includes the other.

If $W \subseteq V$ is a subspace, we get the following equivalence relation on V:

$$v_1 \sim v_2 \quad \Leftrightarrow \quad v_1 - v_2 \in W.$$

Exercise 2.5.F. Check that this is an equivalence relation. (You have to check three things.)

Definition 2.5.2. For $W \subseteq V$ a subspace, we define the **quotient** of V by W as

$$V/W = V/(v_1 \sim v_2 \Leftrightarrow v_1 - v_2 \in W),$$

Addition and multiplication on a quotient space is defined by picking representatives

with addition and multiplication defined as

$$[v_1] + [v_2] = [v_1 + v_2], \quad c[v] = [cv].$$

This definition of addition and multiplication requires some checking. We might worry that maybe $[v_1] = [v_3]$ and $[v_2] = [v_4]$ but $[v_1 + v_2] \neq [v_3 + v_4]$. If this happens, addition is not well-defined and we have a problem. But $[v_1] = [v_3]$ means that $v_1 \sim v_3$, which is equivalent to $v_1 - v_3 \in W$. Likewise $[v_2] = [v_4]$ means $v_2 - v_4 \in W$. Then $(v_1 + v_2) - (v_3 + v_4) = (v_1 - v_3) + (v_2 - v_4) \in W$ and so $[v_1 + v_3] = [v_2 + v_4]$. Likewise, if $[v] = [v_0]$ then $v - v_0 \in W$ and so $cv - cv_0 = c(v - v_0) \in W$ implies $[cv] = [cv_0]$. This shows that addition and scalar multiplication are well-defined oprations on the quotient set V/W.

The quotient can be thought of as “killing off” or “ignoring” or “collapsing” the W part inside V. Anything that was in W becomes 0 in the quotient space V/W. There is a natural projection map

$$\pi : V \to V/W; \quad v \mapsto [v],$$

and $\pi(v) = 0$ means that $[v] = [0] \in V/W$, and this means $v \in W$.

Any surjective linear map is a projection map up to an isomorphism

Exercise 2.5.G. Show that the map $\pi : V \to V/W$ is always surjective. Conversely, show that if $f : V \twoheadrightarrow U$ any surjective linear map, then $U \cong V/f^{-1}(0)$ so that $f : V \to V/f^{-1}(0) \to U$ is a quotient map composed with an isomorphism.

Again, there is a universal property for quotients, which is the analogue of the universal property for quotienting out by equivalence relations.

A linear map that vanishes on a subspace is the same as a linear map from the quotient

Exercise 2.5.H (Universal property for quotients)**.** Let $W \subseteq V$ be a subspace of a k-vector space. For an arbitrary k-vector space U and a linear map $f : V \to U$ such that $f(w) = 0$ for all $w \in W$, prove that there exists a unique linear map $g : V/W \to U$ such that $f = g \circ \pi$.

$$\begin{tikzcd}
V \arrow[r, "f"] \arrow[d, "\pi"] & U \\
V/W \arrow[ru, dashed, "g"'] &
\end{tikzcd}$$

Exercise 2.5.I. For $W \subseteq V$ a subspace, show that $V/W \cong 0$ as vector spaces if and only if $W = V$.

Exercise 2.5.J (2nd isomorphism theorem)**.** Let $W, U \subseteq V$ be subspaces. Show that there is a natural isomorphism

$$(W + U)/U \to W/(W \cap U); \quad [w + u] \mapsto [w].$$

Be careful about what each $[-]$ means!

Quotients by the same vector space can be canceled out

Exercise 2.5.K (3rd isomorphism theorem)**.** Let $U \subseteq W \subseteq V$ be subspaces. Then W/U can be regarded as a subspace of V/U, via the injective linear map $W/U \hookrightarrow V/U$; $[w] \mapsto [w]$. Show that the natural map

$$V/W \to (V/U)/(W/U); \quad [v] \mapsto [[v]]$$

is well-defined and is an isomorphism. Again, be careful about $[-]$.

2.6 Vector spaces from linear maps

In this section, we are going to construct vector spaces from linear maps.

Definition 2.6.1. Let $f : V \to W$ be a linear map between k-vector spaces. We define its **kernel** $\ker f$ and **image** $\operatorname{im} f$ as

$$\begin{aligned}
\ker f &= f^{-1}(0_W) = \{v \in V : f(v) = 0\} \subseteq V, \\
\operatorname{im} f &= f(V) = \{f(v) : v \in V\} \subseteq W.
\end{aligned}$$

Exercise 2.6.A. Show that $\ker f$ is a subspace of V and that $\operatorname{im} f$ is a subspace of W.

A linear map canonically factors through the image

Exercise 2.6.B. Show that any linear map $f : V \to W$ factors as $f : V \to \operatorname{im} f \to W$, where $V \twoheadrightarrow \operatorname{im} f$ is surjective and $\operatorname{im} f \hookrightarrow W$ is injective.

Exercise 2.6.C. For a linear map $f : V \to W$, show that f is injective if and only if $\ker f = \{0\}$.

Quotient by the kernel is isomorphic to the image

Exercise 2.6.D (1st isomorphism theorem)**.** Let $f : V \to W$ be a linear map, and consider the natural map

$$V/\ker f \to \operatorname{im} f; \quad [v] \mapsto f(v).$$

Show that this map is well-defined and is an isomorphism of vector spaces.

A linear map that becomes zero after composition is the same as a linear map to the kernel

Exercise 2.6.E (Universal property for kernels)**.** Let $f : V \to W$ be a linear map, and denote by $i : \ker f \to V$ the inclusion map. For an arbitrary vector space U and a linear map $g : U \to V$ such that $f \circ g = 0$, show that there exists a unique map $h : U \to \ker f$ such that $g = i \circ h$.

$$\begin{tikzcd}
U \arrow[d, dashed, "h"'] \arrow[dr, "g"] \arrow[drr, "0"] & & \\
\ker f \arrow[r, "i"'] & V \arrow[r, "f"'] & W
\end{tikzcd}$$

We are now going to make a definition that might seem very unmotivated. It will also be hard to visualize what it is supposed to mean. But my advice is not to try too hard to find geometric meaning in this object. This definition is going to be useful in a more formal, symbolic context.

Definition 2.6.2. Let $f : V \to W$ be a linear map. We define its **cokernel** as

$$\operatorname{coker} f = W/\operatorname{im} f.$$

This makes sense because $\operatorname{im} f$ is a subspace of W.

This is supposed to be the dual notion of the kernel, and you can see this from the universal property. Note that there is a natural projection map $\pi : W \to W/\operatorname{im} f = \operatorname{coker} f$.

A linear map that becomes zero after composition is the same as a linear from the cokernel

Exercise 2.6.F (Universal property for cokernels)**.** Let $f : V \to W$ be a linear map, and denote by $\pi : W \to \operatorname{coker} f$ the projection map. For an arbitrary vector space U and a linear map $g : W \to U$ such that $g \circ f = 0$, show that there exists a unique map $h : \operatorname{coker} f \to U$ such that $g = h \circ \pi$.

$$\begin{tikzcd}
V \arrow[r, "f"] \arrow[drr, "0"'] & W \arrow[r, "\pi"] \arrow[dr, "g"] & \operatorname{coker} f \arrow[d, dashed, "h"] \\
& & U
\end{tikzcd}$$

Exercise 2.6.G. Let $f : V \to W$ be a linear map. Show that there are natural isomorphisms $\ker(W \to \operatorname{coker} f) \cong \operatorname{coker}(\ker f \to V) \cong \operatorname{im} f$.

All these constructions are "functorial", in the following sense. Suppose we have linear maps $f_i : V_i \to W_i$ and $\varphi_V : V_1 \to V_2$, $\varphi_W : W_1 \to W_2$ such that

$\varphi_W \circ f_1 = f_2 \circ \varphi_V$.

$$\begin{tikzcd}
V_1 \arrow[r, "f_1"] \arrow[d, "\varphi_V"] & W_1 \arrow[d, "\varphi_W"] \\
V_2 \arrow[r, "f_2"] & W_2
\end{tikzcd}$$

We can extend on the left and right by looking at the kernels and cokernels of f_i. But then, we note that

$$f_2 \circ (\varphi_V \circ \iota_1) = \varphi_W \circ f_1 \circ \iota_1 = \varphi_W \circ 0 = 0,$$

and therefore by the universal property of kernels, there exists a unique map $\ker f_1 \to \ker f_2$ such that makes the diagram commute. Likewise,

$$(\pi_2 \circ \varphi_W) \circ f_1 = \pi_2 \circ f_2 \circ \varphi_V = 0 \circ \varphi_V = 0,$$

and therefore there is a unique map $\operatorname{coker} f_1 \to \operatorname{coker} f_2$ that makes the diagram commute.

$$\begin{tikzcd}
\ker f_1 \arrow[r, "\iota_1"] \arrow[d, dashed] & V_1 \arrow[r, "f_1"] \arrow[d, "\varphi_V"] & W_1 \arrow[r, "\pi_1"] \arrow[d, "\varphi_W"] & \operatorname{coker} f_1 \arrow[d, dashed] \\
\ker f_2 \arrow[r, "\iota_2"] & V_2 \arrow[r, "f_2"] & W_2 \arrow[r, "\pi_2"] & \operatorname{coker} f_2
\end{tikzcd}$$

Let me also introduce the notion of exactness. This will probably not be helpful for us too much, but it is an extremely useful notion for keeping track of various data, once we start dealing with complicated situations. Also, wrapping your head around it will help you get familiar with all the notions we have been looking at so far.

Definition 2.6.3. A sequence of vector spaces and linear maps

$$V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3$$

is said to be **exact** if $\ker f_2 = \operatorname{im} f_1$. More generally, a sequence

$$V_0 \xrightarrow{f_0} V_1 \xrightarrow{f_1} \cdots \xrightarrow{f_{n-1}} V_n$$

is said to be **exact** if $\ker f_i = \operatorname{im} f_{i-1}$ for all $1 \le i \le n-1$.

Exercise 2.6.H. For $f_1 : V_1 \to V_2$ and $f_2 : V_2 \to V_3$, show that $\operatorname{im} f_1 \subseteq \ker f_2$ if and only if $f_2 \circ f_1 = 0$. As a consequence, if $V_1 \to V_2 \to V_3$ is exact, then the composite of the two maps is zero.

Exercise 2.6.I. Show that a map $f : V \to W$ is injective if and only if the sequence $0 \to V \to W$ is exact. Likewise, show that f is surjective if and only if the sequence $V \to W \to 0$ is exact.

Injectivity and surjectivity correspond to exactness conditions

Exercise 2.6.J. An exact sequence that looks like $0 \to A \to B \to C \to 0$ is called a **short exact sequence**. In this case, $A \to B$ is an injection, so A can be identified with a subspace of B. Show that C is naturally isomorphic to B/A.

Exercise 2.6.K. Consider the sequence

$$0 \longrightarrow k \xrightarrow{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]} k^2 \xrightarrow{[c\ d]} k \longrightarrow 0.$$

Find the condition that this sequence is exact, in terms of a, b, c, and d.

Exercise 2.6.L. Show that, for a linear map $f : V \to W$, the sequences

$$0 \to \ker f \to V \to \operatorname{im} f \to 0, \quad 0 \to \operatorname{im} f \to W \to \operatorname{coker} f \to 0$$

are exact.

Any linear map can be made into an exact sequence

Exercise 2.6.M. Show that, for a linear map $f : V \to W$, the sequence

$$0 \to \ker f \to V \to W \to \operatorname{coker} f \to 0$$

is always exact. (In general, if $\cdots \to A_{-1} \to A_0 \to B \to 0$ and $0 \to B \to A_1 \to A_2 \to \cdots$ are exact sequences, you can string them together to get an exact sequence $\cdots \to A_{-1} \to A_0 \to A_1 \to A_2 \to \cdots$.)

Exercise 2.6.N (the four lemma and the five lemma)**.** Do this exercise only if you feel super energetic. Consider a commutative diagram

$$\begin{tikzcd}
V_1 \arrow[r] \arrow[d, "\varphi_1"] & V_2 \arrow[r] \arrow[d, "\varphi_2"] & V_3 \arrow[r] \arrow[d, "\varphi_3"] & V_4 \arrow[d, "\varphi_4"] \\
W_1 \arrow[r] & W_2 \arrow[r] & W_3 \arrow[r] & W_4
\end{tikzcd}$$

such that the two rows are exact sequences.

(a) If φ_1 and φ_3 are surjective, and φ_4 is injective, show that φ_2 is surjective.

(b) If φ_2 and φ_4 are injective, and φ_1 is surjective, show that φ_3 is injective.

Conclude that if

$$\begin{tikzcd}
V_1 \arrow[r] \arrow[d, "\varphi_1"] & V_2 \arrow[r] \arrow[d, "\varphi_2"] & V_3 \arrow[r] \arrow[d, "\varphi_3"] & V_4 \arrow[r] \arrow[d, "\varphi_4"] & V_5 \arrow[d, "\varphi_5"] \\
W_1 \arrow[r] & W_2 \arrow[r] & W_3 \arrow[r] & W_4 \arrow[r] & W_5
\end{tikzcd}$$

commutes and rows are exact, and $\varphi_1, \varphi_2, \varphi_4, \varphi_5$ are isomorphisms, then φ_3 is an isomorphism as well.

Before concluding this section, let me definer another vector space.

Definition 2.6.4. Let V and W be two k-vector spaces. We define

$$\operatorname{Hom}_k(V, W) = \{k\text{-linear maps } f : V \to W\}.$$

This as a structure of a vector space, given by

$$(f + g)(v) = f(v) + g(v), \quad (cf)(v) = cf(v).$$

So $\mathrm{Hom}_k(V, W)$ is not just a set, but a k-vector space again. It can be checked that the addition and scalar multiplication satisfy all the axioms, with 0 being the zero map $0 : V \to 0 \to W$.

Exercise 2.6.O. For any k-vector space V, show that there is a natural isomorphism

$$\mathrm{Hom}_k(k, V) \cong V; \quad f \mapsto f(1).$$

Linear maps to vector spaces are the same as linear maps to the product

Exercise 2.6.P. Let $\{V_i\}_{i \in I}$ be a set of k-vector spaces. For any k-vector space W, show that the natural linear map

$$\mathrm{Hom}_k(W, \textstyle\prod_{i\in I} V_i) \to \displaystyle\prod_{i\in I} \mathrm{Hom}_k(W, V_i); \quad f \mapsto (\pi_i \circ f)_{i\in I}$$

is an isomorphism of vector spaces.

Linear maps from vector spaces are the same as linear maps from the direct sum

Exercise 2.6.Q. Let $\{V_i\}_{i \in I}$ be a set of k-vector spaces. For any k-vector space W, show that the natural linear map

$$\mathrm{Hom}_k(\textstyle\bigoplus_{i\in I} V_i, W) \to \displaystyle\prod_{i\in I} \mathrm{Hom}_k(V_i, W); \quad f \mapsto (f \circ \iota_i)_{i\in I}$$

is an isomorphism of vector spaces. As a consequence, there is a natural isomorphism $\mathrm{Hom}_k(k^{\oplus S}, V) \cong V^S$.

If $f : V_1 \to V_2$ is a linear map, and W is an arbitrary k-vector space, there are induced maps

$$f_* : \mathrm{Hom}_k(W, V_1) \to \mathrm{Hom}_k(W, V_2); \quad \alpha \mapsto f \circ \alpha,$$
$$f^* : \mathrm{Hom}_k(V_2, W) \to \mathrm{Hom}_k(V_1, W); \quad \alpha \mapsto \alpha \circ f.$$

(In case you're wondering what lower and upper stars mean, lower star usually means that the direction of the order did not change, e.g., $V_1 \to V_2$ inducing $\mathrm{Hom}_k(-, V_1) \to \mathrm{Hom}_k(-, V_2)$. Upper star means that the direction did change, e.g., $V_2 \to V_2$ inducing $\mathrm{Hom}_k(V_2, -) \to \mathrm{Hom}_k(V_1, -)$.)

Covariant Hom is left exact

Exercise 2.6.R. If $0 \to V_1 \to V_2 \to V_3$ is an exact sequence, show that the induced sequence

$$0 = \mathrm{Hom}_k(W, 0) \to \mathrm{Hom}_k(W, V_1) \to \mathrm{Hom}_k(W, V_2) \to \mathrm{Hom}_k(W, V_3)$$

is exact as well.

Contravariant Hom is left exact

Exercise 2.6.S. If $V_1 \to V_2 \to V_3 \to 0$ is an exact sequence, show that the induced sequence

$$0 = \mathrm{Hom}_k(0, W) \to \mathrm{Hom}_k(V_3, W) \to \mathrm{Hom}_k(V_2, W) \to \mathrm{Hom}_k(V_1, W)$$

is exact as well.

2.7 Bases and dimension

We are now more than ready to start classifying vector spaces. Recall that this was our goal for studying vector spaces: classifying vector spaces up to isomorphism.

Definition 2.7.1. Let V be a k-vector space and let $S \subseteq V$ be a subset (not a subspace) of V. The inclusion map $i : S \hookrightarrow V$ induces a linear map $f : k^{\oplus S} \to V$ by Proposition 2.4.3. We say that the set S is

- **linearly independent** if f is injective, **linearly dependent** if f is not injective,
- **generating** or **spanning** if f is surjective, and
- a **basis** if f is bijective.

Okay, this is some abstract definition. It would be helpful to know what they really mean in concrete terms.

Exercise 2.7.A. Show that a subset $S \subseteq V$ is linearly independent if and only if the following condition holds: for arbitrary distinct vectors $v_1, \dots, v_n \in S$ and scalars $a_1, \dots, a_n \in k$, the equation

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = 0$$

implies $a_1 = a_2 = \cdots = a_n = 0$.

Exercise 2.7.B. Show that a subset $S \subseteq V$ is generating if and only if the following condition holds: for each $v \in V$, there exist elements $v_1, \dots, v_n \in S$ and scalars $a_1, \dots, a_n \in k$ such that

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = v.$$

Intuitively speaking, linear independence means that you cannot combine the vectors in S in a nontrivial way to make 0. For instance, if you have the vectors $\vec{v}_1 = (1, 0)$ and $\vec{v}_2 = (2, 3)$ in $\mathbb{R}^2$, they point in different directions, so you cannot make some linear combination and get 0, unless you do something as meaningless as $0 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2 = \vec{0}$.

A subset $S \subseteq V$ spans the vector space V if every vector in V can be written as a linear combination of vectors in S. I think it is quite clear what this means. For instance, $\{(1, 0), (2, 3), (3, 5)\}$ spans $\mathbb{R}^2$, because we can, for instance, write $(x, y) = x(1, 0) - 3y(2, 3) + 2y(3, 5)$. But it is not linearly independent because, say, $(1, 0) - 5(2, 3) + 3(3, 5) = (0, 0)$.

Definition 2.7.2. Let $S \subseteq V$ be an arbitrary subset of a vector space. The inclusion map $i : S \hookrightarrow V$ induces a linear map $f : k^{\oplus S} \to V$. We define the **subspace spanned by** S as $\operatorname{span}(S) = \operatorname{im} f \subseteq V$.

It is clear that $\operatorname{span}(S)$ is the subspace consisting of vectors that can be expressed as a linear combination of elements of S. Also, S is spanning if and only if $\operatorname{span}(S) = V$. The two conditions, linear independence and spanning, when put together, means something very nice.

Exercise 2.7.C. Show that a subset $S \subseteq V$ is a basis if and only if the following condition holds: for each $v \in V$, there is a unique tuple $(a_s)_{s \in S}$ of elements in k such that $a_s \neq 0$ only for finitely many s and

$$\sum_{s \in S} a_s s = v.$$

One reason we care about bases is that a basis gives an isomorphism between $k^{\oplus S}$ and V, that is, an isomorphism between V and a free vector space. We know exactly what a free vector space looks like, so such an isomorphism is going to be useful in studying properties of the vector space V.

The surprising fact that actually every vector space has a basis!

Every vector space has a basis

Theorem 2.7.3 (under Axiom of Choice)**.** *Every vector space has a basis.*

Corollary 2.7.4. *Every vector space is isomorphic to $k^{\oplus S}$ for some set S.*

How would one prove such a theorem? We need to construct a basis, which is a set that is both linearly independent and generating. Let us start with $S_0 = \emptyset$. This is clearly linearly independent, because there is no vector to form a nontrivial linear relation. But also, unless $V = 0$, the set is not spanning. So take any nonzero vector $v_1 \in V$ and throw it into the set S_0. After this step, we have $S_1 = \{v_1\}$. As long as $v_1 \neq 0$, the set $\{v_1\}$ is linearly independent because $a_1 v_1 = 0$ implies $a_1 = 0$. If $V = \operatorname{span}(S_1) = \{a_1 v_1\}$ then this is spanning as well, and we are done. Otherwise, we need more vectors, so pick another vector $v_2 \notin \operatorname{span}(S_1)$ and throw it into S_1. Then we get a set $S_2 = \{v_1, v_2\}$ that is linearly independent, but not necessarily spanning. We are going to repeat this process, until we get a spanning set. One problem is that the process might continue on infinitely, and the Axiom of Choice is what will help us deal with this issue.

Exercise 2.7.D. Let us make the above process formal. Let $S \subseteq V$ be a linearly independent subset. Suppose that $\operatorname{span}(S) \subsetneq V$, and pick a vector $v \in V$ with $v \notin \operatorname{span}(S)$. Show that $S \cup \{v\}$ is linearly independent as well.

One general tool for dealing with infinite processes is Zorn's lemma. We will only state and not prove the lemma. It is known that Zorn's lemma is equivalent to the Axiom of Choice.

Lemma 2.7.5 (Zorn's lemma)**.** *Let X be a set, and consider a relation $R \subseteq X \times X$. (We're going to write $(x, y) \in R$ as $x \preceq y$. As the notation suggests, this relation is supposed to represent an ordering.) Suppose this relation satisfies the following:*

(1) $x \preceq x$ for all $x \in X$,

(2) if $x \preceq y$ and $y \preceq x$ then $x = y$,

(3) if $x \preceq y$ and $y \preceq z$ then $x \preceq z$,

(4) if $S \subseteq X$ is an subset such that either $x \preceq y$ or $y \preceq x$ for every $x, y \in S$, then there exists an $z \in X$ such that $x \preceq z$ for all $x \in S$.

Then there exists an element $m \in X$ such that $m \preceq x$ is true only for $x = m$.

The first three conditions capture a notion of ordering. A relation satisfying (1), (2), (3) is called a **partial order**, with "partial" meaning that not every two elements can be compared. The last condition (4) is the interesting one. It roughly says that every "chain" of comparable elements has an "upper bound". The conclusion is that there is an element that is maximal in the sense that nothing is strictly bigger. If you are interested in learning the proof, the idea is the same as what we sketched for finding a basis. Pick an arbitrary element, which is probably not going to be maximal. If it is not maximal, you can find an element that is strictly bigger than it. If this is not maximal, then you can continue finding an element strictly bigger that that and so on. The condition (4) allows you to run the process transfinitely.

Proof of Theorem 2.7.3. Let V be a vector space, and consider the collection

$$\mathcal{A} = \{S \subseteq V : S \text{ is linearly independent}\}$$

of linearly independent sets. Consider the inclusion order $S_1 \preceq S_2 \Leftrightarrow S_1 \subseteq S_2$ given on $\mathcal{A}$. By properties of sets, this automatically satisfies (1), (2), and (3) of Lemma 2.7.5.

We can also check the last condition (4). Suppose that $\mathcal{C} \subseteq \mathcal{A}$ is a collection of linearly independent sets, such that any two $S_1, S_2 \subseteq \mathcal{C}$ are ordered by inclusion. If we define

$$S_M = \bigcup_{S \in \mathcal{C}} S,$$

then it is clear that S_M is a set that contains all elements of $\mathcal{C}$. We also can check that $S_M \in \mathcal{A}$, i.e., S_M is linearly independent. If not, there exist distinct vectors $v_1, \ldots, v_n \in S_M$ and $a_1, \ldots, a_n \in k$, not all zero, such that $a_1v_1 + \cdots + a_nv_n = 0$. But each v_i is in the union of $S \in \mathcal{C}$, so there exist $S_i \in \mathcal{C}$ such that $v_i \in S_i$. If we look at the sets $S_1, \ldots, S_n \in \mathcal{C}$, every two of them are comparable by inclusion. This means that we can assume $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_n$ without loss of generality. Then $v_i \in S_i$ implies that $v_1, \ldots, v_n \in S_n$. This and $a_1v_1 + \cdots + a_nv_n = 0$ contradicts that $S_n \in \mathcal{A}$ is linearly independent. Therefore S_M has to be linearly independent, and is larger than all elements of $\mathcal{C}$. This verifies (4).

We now apply Lemma 2.7.5 to the collection $\mathcal{A}$ ordered by inclusion. There exists a maximal set $M \in \mathcal{A}$. (Maximality in this context means that $M \subseteq S$ and S linearly independent implies $M = S$.) Because M is already linearly independent, we are done if we can show that M spans V. Suppose not, that $\operatorname{span}(M) \subsetneq V$. We can pick a vector $v \in V \setminus \operatorname{span}(M)$, and then by Exercise 2.7.D, $M \cup \{v\}$ is a linearly independent set that is strictly larger than M. (The vector v can't already be in M, because then v would be in $\operatorname{span}(M)$.) This contradicts our assumption on M, and therefore $\operatorname{span}(M) = V$. That is, M is both linearly independent and spanning, hence a basis. □

This is a difficult proof, probably one of the hardest you will encounter in linear algebra. But the main idea is simple: if the set is not spanning, add in vectors while keeping it linearly independent. The rest is simply a standard technique of using Zorn's lemma, which you will get used to after seeing it three times.

Exercise 2.7.E. Let $B \subseteq V$ be a basis of a vector space V. For each vector space, show that the natural map

$$\operatorname{Hom}_k(V, W) \to \prod_{b \in B} W; \quad f \mapsto (f(b))_{b \in B}$$

is an isomorphism of vector spaces.

Every linearly independent set can be extended to a basis

Exercise 2.7.F. Let us prove something slightly stronger than just the existence of a basis. Let V be a k-vector space and consider an arbitrary linearly independent subset $S \subseteq V$. Prove that there exists a superset $S \subseteq T$ such that T is a basis of V. (Hint: we need to start the process from S instead of the empty set. Consider the family of linearly indepedent sets already containing S.)

There is a basis between any linearly independent set and any spanning set containing it

Exercise 2.7.G. We can even put an upper bound in finding the basis. Let V be a k-vector space and consider $S_1 \subseteq V$ a linearly indepedent set and $S_2 \subseteq V$ a spanning set. Suppose that $S_1 \subseteq S_2$. Prove that there exists a subset T with $S_1 \subseteq T \subseteq S_2$ such that T is a basis of V. (Hint: this time, we only add in elements of S_2 while running the process. Consider the lineraly independent subsets containing S_1 and contained in S_2.)

So we have proven that all vector spaces are isomorphic to some $k^{\oplus S}$. But this does not finish the classification of vector spaces, because there certainly are multiplicities within this classification. For instance, it is possible that $k^{\oplus S} \cong k^{\oplus T}$ for distinct sets $S \neq T$. For instance, if S and T have the same number of elements, then probably $k^{\oplus S} \cong k^{\oplus T}$.

Exercise 2.7.H. Suppose that $\alpha : S \to T$ is a map of sets. Show that the induced map $\alpha_* : k^{\oplus S} \to k^{\oplus T}$ is injective if and only if α is injective, and that α_* is surjective if and only if α is surjective. As a consequence, α_* is an isomorphism if α is bijective.

It is reasonable to guess that $k^{\oplus S} \cong k^{\oplus T}$ if and only if there is a bijection between S and T. This turns out to be a true statement, but we need some preparation before proving it.

Every short exact sequence splits

Proposition 2.7.6. *Let $f : V \twoheadrightarrow W$ be a surjective linear map. Then there exists a linear map $g : W \to V$ such that $f \circ g = \mathrm{id}_W$ is the identity map. (Such g necessarily has to be injective.)*

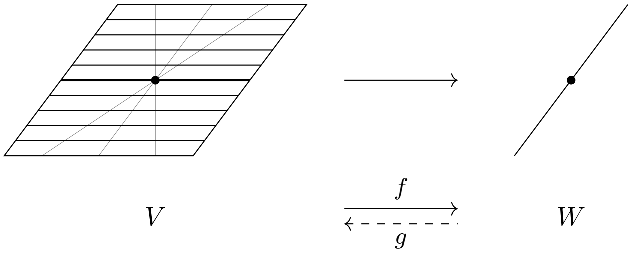

Figure 2.2: Different sections (gray) of a projection map $f : V \to W$

Proof. Pick a basis $B \subseteq W$. From Exercise 2.7.E, we see that a map $g : W \to V$ is uniquely determined by the values $g(b)$ for $b \in B$. Because we want $f(g(b)) = b$ for each $b \in B$, we pick $g(b)$ as an vector in the inverse image $f^{-1}(b)$. Here, note that $f^{-1}(b)$ is nonempty because f is surjective. But $f(g(b)) = b$ means that $f \circ g : W \to W$ and $\mathrm{id}_W : W \to W$ agree on all elements of B. Exercise 2.7.E immediately implies that $f \circ g = \mathrm{id}_W$ as linear maps. □

Example 2.7.7. Note that the function g is far from being unique. Let's try to see what we did in the previous proof with an example of $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = y$. We first pick a basis $B = \{1\}$ of $W = \mathbb{R}$. The proof tells us to pick $g(1)$ as any vector that is in $f^{-1}(1) = \{(x, 1)\}$. In general, let us pick $g_c(1) = (c, 1)$, so that $g_c(x) = (cx, x)$. It is clear that all these linear maps g_c satisfy $f \circ g_c = \mathrm{id}_W$. See Figure 2.2.

Exercise 2.7.I. Consider a short exact sequence

$$0 \longrightarrow V_1 \xrightarrow{f_1} V_2 \xrightarrow{f_2} V_3 \longrightarrow 0.$$

Then $f_2 : V_2 \to V_3$ is surjective, and hence there exists a linear map $g_2 : V_3 \to V_2$ such that $f_2 \circ g_2 = \mathrm{id}_{V_3}$. Show that every element $v_2 \in V_2$ can be uniquely presented as

$$v_2 = f_1(v_1) + g_2(v_3)$$

for $v_1 \in V_1$ and $v_3 \in V_3$. (Uniqueness means that if $v_2 = f_1(v_1') + g_2(v_3')$ then $v_1' = v_1$ and $v_3' = v_3$.) Also show that there exists a linear map $g_1 : V_2 \to V_1$ such that $g_1 \circ f_1 = \mathrm{id}_{V_1}$.

There is an injective map if and only if there is a surjective map in the other direction

Exercise 2.7.J. For two vector spaces V and W, show that there exists a surjective linear map $V \twoheadrightarrow W$ if and only if there exists an injective linear map $W \hookrightarrow V$. (Hint: use the previous exercise.)

We are now ready to prove that $k^{\oplus S} \cong k^{\oplus T}$ implies that there exists a bijection between the two sets S and T. In fact, we will prove a much more stronger statement. This is again a hard theorem, but try to at least grasp the main idea. If you are not familiar with set theory regarding cardinality, feel free to skip the case when T is infinite.

Theorem 2.7.8. *For arbitrary sets S and T, the following are equivalent:*

(i) There exists an injective linear map $k^{\oplus S} \hookrightarrow k^{\oplus T}$.

(ii) There exists a surjective linear map $k^{\oplus T} \twoheadrightarrow k^{\oplus S}$.

(iii) There exists an injective (set) map $S \hookrightarrow T$.

(iv) Either there exists a surjective (set) map $T \twoheadrightarrow S$ or $S = \emptyset$.

Proof. Exercise 2.7.J gives the equivalence between (i) and (ii). It follows from basic set theory that (iii) and (iv) are equivalent. (Actually (iv) $\Rightarrow$ (iii) requires the Axiom of Choice, but we're already assuming this.) So now we need to prove equivalence between (i) $\Leftrightarrow$ (ii) and (iii) $\Leftrightarrow$ (iv). One direction is immediate: (iii) $\Rightarrow$ (i) follows from Exercise 2.7.H. It suffices to show the other direction.

We are going to divide into two cases. First assume that T is infinite, and consider a surjective linear map $f : k^{\oplus T} \twoheadrightarrow k^{\oplus S}$. We are going to prove (ii) $\Rightarrow$ (iv). For each element $t_j \in T$, we look at $f(\underline{t_j}) \in k^{\oplus S}$. Write

$$f(\underline{t_j}) = a_{j,1}\underline{s_{j,1}} + a_{j,2}\underline{s_{j,2}} + \cdots + a_{j,n_j}\underline{s_{j,n_j}}.$$

Now we define a (set) map

$$\alpha : \{(t_j, m) \in T \times \mathbb{Z}_{\geq 0} : 1 \leq m \leq n_j\} \to S; \quad (t_j, m) \mapsto s_{j,m}.$$

The point is that this map is a surjective map. If some $s \in S$ doesn't appear in the image of α, this means that s doesn't appear in any of the $f(\underline{t_j})$. Then the image of f is contained in $k^{\oplus(S\setminus\{s\})}$, which contradicts that f is surjective. So we get a surjective map from a subset of $T \times \mathbb{Z}_{\geq 0}$ to S. Because T is infinite, there exists a bijection between T and $T \times \mathbb{Z}_{\geq 0}$ (this is another set-theoretic fact), and hence there exists a surjection from a subset of T to S. This gives (iv).

Now let us deal with the case when $T = \{t_1, \ldots, t_n\}$ is finite. We are going to use the next lemma (Lemma 2.7.9) here to prove (ii) $\Rightarrow$ (iii). Consider a surjective map $f : k^{\oplus T} \twoheadrightarrow k^{\oplus S}$. Then $A = \{f(\underline{t_1}), \ldots, f(\underline{t_n})\}$ is a spanning set for $k^{\oplus S}$, and on the other hand, $B = \{\underline{s} : s \in S\} = \{\underline{s_1}, \underline{s_2}, \ldots\}$ is a basis of $k^{\oplus S}$. By Lemma 2.7.9, we can remove one element in B and replace it by an element of A to get another basis. Let us remove $\underline{s_1}$ from B and replace it with some element of A to get a new basis

$$B_1 = \{\underline{t_{i_1}}, \underline{s_2}, \underline{s_3}, \ldots\}.$$

Then we can remove $\underline{s_2}$ and replace it by some element of A to get another basis

$$B_2 = \{\underline{t_{i_1}}, \underline{t_{i_2}}, \underline{s_3}, \ldots\}.$$

Here, we must have $i_1 \neq i_2$, because otherwise B_2 is strictly a subset of B_1. This means that this process should not be allowed to continue for more than n times, because A has only n elements and the same element cannot appear twice among $t_{i_1}, t_{i_2}, \ldots$. This shows that B cannot have more than n elements to start out with. Therefore S has at most n elements and there exists an injective (set) map $S \hookrightarrow T$. □

Given a basis minus one element, you can put in an element of a given spanning set to get another basis

Lemma 2.7.9 (basis exchange)**.** *Let V be a k-vector space. Let $B \subseteq V$ be a basis, and $A \subseteq V$ be a spanning set. Then for each $b \in B$, there exists an $a \in A$ such that $(B \setminus \{b\}) \cup \{a\}$ is again a basis.*

Proof. Because A is spanning, there exists an $a \in A$ such that $a \notin \operatorname{span}(B \setminus \{b\})$. Otherwise, $A \subseteq \operatorname{span}(B \setminus \{b\})$ and so $V = \operatorname{span} A \subseteq \operatorname{span}(B \setminus \{b\})$ gives a contradiction. By Exercise 2.7.D, the set $(B \setminus \{b\}) \cup \{a\}$ is linearly independent. Let us now show that it is spanning. Because B is a basis, there is a way to write

$$a = cb + c_1 b_1 + c_2 b_2 + \cdots + c_n b_n$$

where $c, c_1, \ldots, c_n \in k$ and $b_1, \ldots, b_n \in B \setminus \{b\}$ are distinct. Because $a \notin \operatorname{span}(B \setminus \{b\})$, we have $c \neq 0$. Then we can also write

$$b = c^{-1}a - c^{-1}c_1 b_1 - c^{-1}c_2 b_2 - \cdots - c^{-1}c_n b_n.$$

This shows that $b \in \operatorname{span}((B \setminus \{b\}) \cup \{a\})$ and so $V = \operatorname{span}(B) \subseteq \operatorname{span}((B \setminus \{b\}) \cup \{a\})$. Therefore $(B \setminus \{b\}) \cup \{a\}$ is a basis. □

Exercise 2.7.K. Translate Theorem 2.7.8 to the following statement. Let V and W be vector spaces, and let B_V and B_W be bases for V and W respectively. Then the following statements are equivalent:

(i) There exists an injective linear map $V \hookrightarrow W$.

(ii) There exists a surjective linear map $W \twoheadrightarrow V$.

(iii) There exists an injective map $B_V \hookrightarrow B_W$.

(iv) Either there exists a surjective map $B_W \twoheadrightarrow B_V$ or $B_V = \emptyset$.

The size of a basis is an invariant of the vector space

Corollary 2.7.10. *If B_1 and B_2 are two bases of a k-vector space V, then $|B_1| = |B_2|$, i.e., there exists a bijection between B_1 and B_2.*

Proof. Because $\mathrm{id}_V : V \to V$ is injective, there exists an injective map $B_1 \hookrightarrow B_2$. But $\mathrm{id}_V : V \to V$ is also surjective, so there exists an injective map $B_2 \hookrightarrow B_1$. By the Schröder–Bernstein theorem, which we will not prove, there exists a bijection between the two sets B_1 and B_2. □

So the number of elements of a basis of V is an invariant of V, not depending on the choice of the basis. This allows us to define the dimension of V.

Exercise 2.7.L. Let V be a k-vector space. Show that the following two statements are equivalent:

(i) There exists a basis $B \subseteq V$ such that B is finite.

(ii) Every basis $B \subseteq V$ is finite.

Definition 2.7.11. Let V be a k-vector space. We say that V is **finite-dimensional** if there exists a finite basis (and hence all bases are finite by the previous exercise). In this case, define the **dimension** of V as

$$\dim_k V = |B|$$

where B is a basis of V. This cardinality is independent of the choice of the basis B, so the dimension is well-defined.

Even if B is an infinite set, we can make sense of $\dim V$ by its "cardinality". We are going to say that $|S_1| = |S_2|$ if there exists a bijection between the two sets S_1 and S_2, and we are going to say that $|S_1| \leq |S_2|$ if there exists an injective map $S_1 \hookrightarrow S_2$. What Schröder–Berstein says is that $|S_1| \leq |S_2|$ and $|S_2| \leq |S_1|$ implies $|S_1| = |S_2|$. But in most cases, we are only going to worry about finite-dimensional vector spaces when talking about dimension.

Example 2.7.12. For any nonnegative integer n, we have $\dim_k k^n = n$ because $e_1, \dots, e_n$ forms a basis.

Corollary 2.7.13. *Let V and W be finite-dimensional k-vector spaces, and let B_V and B_W be bases for V and W respectively. Then the following statements are equivalent:*

(i) There exists an injective linear map $V \hookrightarrow W$.

(ii) There exists a surjective linear map $W \twoheadrightarrow V$.

(iii) $\dim_k V \leq \dim_k W$.

The dimension of the direct sum is the sum of the dimensions

Exercise 2.7.M. For finite-dimensional vector spaces $V_1, V_2, \dots, V_n$, show that

$$\dim_k(V_1 \oplus V_2 \oplus \cdots \oplus V_n) = \dim_k V_1 + \dim_k V_2 + \cdots + \dim_k V_n.$$

Exercise 2.7.N. Let $W \subseteq V$ be a subspace of a finite-dimensional vector space V. Show that W and V/W are also finite-dimensional and

$$\dim_k V = \dim_k W + \dim_k(V/W).$$

The alternating sum of dimensions of an exact sequence vanishes

Exercise 2.7.O. Consider an exact sequence

$$0 \to V_0 \to V_1 \to V_2 \to \cdots \to V_n \to 0,$$

where each V_i are finite-dimensional vector spaces. Then show that

$$\dim_k V_0 - \dim_k V_1 + \dim_k V_2 - \cdots + (-1)^n \dim_k V_n = 0.$$

It is thanks to this fact that the Euler characteristic behaves very well.

Exercise 2.7.P. Let V and W be finite-dimensional vector spaces with $\dim_k V = \dim_k W$. Show that any injective linear $V \hookrightarrow W$ is an isomorphism. Similarly, show that any surjective linear $V \twoheadrightarrow W$ is an isomorphism. Find counterexamples to both statements when V and W are allowed to be infinite. (Hint: try shifting sequences around in $\mathbb{R}^{\oplus \mathbb{Z}_{\geq 0}}$.)

A one-sided inverse of a matrix is automatically a two-sided inverse

Exercise 2.7.Q. Let A and B be $n \times n$ square matrices with entries in a field k. Show that $AB = I$ if and only if $BA = I$. (Here, I denotes the identity matrix, corresponding to $\mathrm{id} : k^n \to k^n$.)

Exercise 2.7.R. Let V and W be finite-dimensional vector spaces. We define the **rank** of a linear map $f : V \to W$ as $\operatorname{rank} f = \dim_k(\operatorname{im} f)$. Show that $\operatorname{rank} f \leq \min\{\dim_k V, \dim_k W\}$.

Let us also do some computations with finite-dimensional vector spaces.

Exercise 2.7.S. Consider the linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by the matrix

$$T = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

(a) Find a basis of $\ker T$. What is $\dim_{\mathbb{R}}(\ker T)$?

(b) Find a basis of $\operatorname{im} T$. What is $\dim_{\mathbb{R}}(\operatorname{im} T)$?

(c) Find a basis of $\operatorname{coker} T$. What is $\dim_{\mathbb{R}}(\operatorname{coker} T)$?

(d) Recall that there are short exact-sequences $0 \to \ker T \to \mathbb{R}^3 \to \operatorname{im} T \to 0$ and $0 \to \operatorname{im} T \to \mathbb{R}^3 \to \operatorname{coker} T \to 0$. Verify that $\dim \ker T + \dim \operatorname{im} T = 3$ and $\dim \operatorname{im} T + \dim \operatorname{coker} T = 3$.

There is a conic passing through any five points on the plane

Exercise 2.7.T. Let $A_1, \dots, A_5 \in \mathbb{R}^2$ be five points on the plane. Consider

$$V = \{\text{polynomials in } x \text{ and } y \text{ with coefficients in } \mathbb{R} \text{ and total degree} \leq 2\}.$$

(For instance, $x^2 + xy + y + 2 \in V$ while $xy^2 \notin V$, because total degree of xy^2 is 3.)

(a) Show that V is a vector space over $\mathbb{R}$, and compute its dimension.

(b) Consider the map $\Phi : V \to \mathbb{R}^5$ given by $p(x,y) \mapsto (p(A_1), \dots, p(A_5))$. Check that this map is linear.

(c) Show that there exists a nonzero polynomial $p(x,y) \in V$ such that $\Phi(p) = 0$.

(d) Conclude that there exists a conic (or two lines) that passes through all five points $A_1, \dots, A_5$.

A polynomial is uniquely determined by values at its degree plus one numbers

Exercise 2.7.U. Let $a_0 < a_1 < \cdots < a_n$ be distinct real numbers. Consider the space

$$V = \{\text{polynomials in } x \text{ with coefficients in } \mathbb{R} \text{ and degree} \leq n\}.$$

Compute $\dim_{\mathbb{R}} V$. Consider the map

$$V \to \mathbb{R}^{n+1}; \quad p(x) \mapsto (p(a_0), \dots, p(a_n)).$$

Show that this linear map is an isomorphism. (You only have to show either injectivity or surjectivity. You can go both ways: for injectivity, you can use polynomial division, and for surjectivity, you can use Lagrange interpolation.)

Picking an ordered basis is picking an isomorphism with a standard vector space

Exercise 2.7.V. Let V be a vector space with $\dim_k V = n$. Show that there is a bijection (of sets)

$$\begin{aligned} \{\text{isomorphisms } k^n \to V\} &\leftrightarrow \{(v_1, \dots, v_n) \in V^n : \{v_1, \dots, v_n\} \text{ is a basis}\} \\ f &\mapsto (f(e_1), \dots, f(e_n)). \end{aligned}$$

In other words, picking an isomorphism $k^n \cong V$ is equivalent to picking a basis and then ordering it. We shall say that $(v_1, \dots, v_n) \in V^n$ is an **ordered basis** if $\{v_1, \dots, v_n\}$ is a basis.

Let V and W be vector spaces with $\dim_k V = n$ and $\dim_k W = m$. Consider a linear map $T : V \to W$. We would like to say that T corresponds to a $m \times n$ matrix, but we cannot immediately say this because a matrix is a linear map $k^n \to k^m$. (The reason we want to write down a map as a matrix is because matrices are computable.) So what we do is choose ordered bases B of V and C of W, and consider the corresponding isomorphisms $\varphi_B : k^n \to V$ and $\varphi_C : k^m \to W$. Then the composition

$${}_C[T]^B = \varphi_C^{-1} \circ T \circ \varphi_B : k^n \xrightarrow{\varphi_B} V \xrightarrow{T} W \xrightarrow{\varphi_C^{-1}} k^m$$

is a $m \times n$ matrix. (This is not standard notation, by the way.) We can also write down vectors $v \in V$ as $n \times 1$ column matrices:

$${}_B[v] = \varphi_B^{-1}(v) \in k^n.$$

Exercise 2.7.W. Let V, W be finite-dimensional vector spaces with ordered bases B, C. For a linear map $T : V \to W$, show that ${}_C[Tv] = {}_C[T]^B{}_B[v]$ as matrices.

Exercise 2.7.X. Let V, W, U be finite-dimensional vector spaces with ordered bases B, C, D. For linear maps $T : V \to W$ and $S : W \to U$, show that ${}_D[S \circ T]^B = {}_D[S]^C{}_C[T]^B$.

$$\begin{tikzcd}
V \arrow[r, "T"] & W \arrow[r, "S"] \arrow[d, shift left, "\varphi_C^{-1}"] & U \\
k^n \arrow[u, "\varphi_B"] \arrow[r, "{}_C[T]^B"'] & k^m \arrow[u, shift left, "\varphi_C"] \arrow[r, "{}_D[S]^C"'] & k^l \arrow[u, "\varphi_D"']
\end{tikzcd}$$

Change of basis corresponds to multiplying appropriate matrices

Exercise 2.7.Y. Let V be a finite-dimensional vector space and consider two bases B and B'. Consider the matrix ${}_B[\mathrm{id}]^{B'}$ representing the identify map $\mathrm{id}_V : V \to V$, where the two V are considered using different bases.

(a) Show that ${}_B[\mathrm{id}]^B = I$ is the identity matrix and ${}_{B'}[\mathrm{id}]^B{}_B[\mathrm{id}]^{B'} = {}_B[\mathrm{id}]^{B'}{}_{B'}[\mathrm{id}]^B = I$, that is, changing B and B' gives the inverse matrix.

(b) If W is another vector space with ordered bases C, C', and $T : V \to W$ is a linear map, show that ${}_{C'}[T]^{B'} = {}_{C'}[\mathrm{id}]^C{}_C[T]^B{}_B[\mathrm{id}]^{B'}$.

(c) If $T : V \to V$ is a linear map, show that ${}_{B'}[T]^{B'} = P{}_B[T]^B P^{-1}$ where $P = {}_{B'}[\mathrm{id}]^B$.

Exercise 2.7.Z. Consider the linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix

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

Consider ordered bases $B = ((1,2),(-1,-1))$ and $B' = ((-1,0),(4,1))$ of $\mathbb{R}^2$.

(a) Compute the matrix ${}_B[T]^B$.

(b) Compute the matrix ${}_{B'}[T]^{B'}$.

(c) Compute the matrix $P = {}_{B'}[\mathrm{id}]^B$.

(d) Verify that ${}_{B'}[T]^{B'} = P{}_B[T]^B P^{-1}$.

2.8 Dual spaces

In this section, we are going to talk about linear maps $V \to k$. These are going to be ways to measure a vector in terms of a single number. Why are these interesting objects to study? First, it is a natural thing to try and measure an object. We might want to measure the length or area of some geometric object, and to do this, we need to know how to measure vectors, or maybe parallelograms formed by vectors. That is, we need some way of turning a vector into a number to measure. The second reason is formal. In mathematics, many objects or concepts appear with some form of duality and it is fundamental to understand what happens between them. The dual vector space is one such example.

Definition 2.8.1. Let V be a k-vector space. We define its **dual** as

$$V^* = \mathrm{Hom}_k(V, k).$$

An element of V^* is sometimes called a **linear functional** on V.

Example 2.8.2. What is $(k^n)^*$? This is the space of linear maps $k^n \to k$, in other words, the space of $1 \times n$ matrices. (Note that we have identified k^n as $n \times 1$ column vectors.) So there is an isomorphism

$$-^* : k^n \to (k^n)^*; \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \mapsto a^* = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}.$$

An interesting structure coming from this isomorphism is the **standard inner product** or **dot product** on k^n. Consider two vectors $a, b \in k^n$. Under

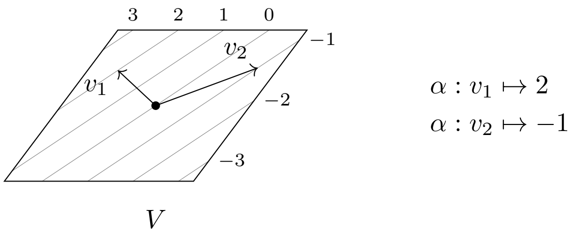

Figure 2.3: Visualizing an element of $\alpha \in V^*$: the vectors on the gray line labeled a are sent to a.

the identification, $a^* \in (k^n)^*$ is a linear map $k^n \to k$, so we can evaluate $a^*(w)$. In terms of matrices, it will be

$$a^*(b) = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \cdots + a_n b_n \in k.$$

From this formula, we immediately note that $a^*(b) = b^*(a)$, and define

$$a \cdot b = b \cdot a = a^*(b) = b^*(a) \in k.$$

Because $a^* : k^n \to k$ is linear, we immediately have some properties like

$$a \cdot (\beta b + \gamma c) = \beta(a \cdot b) + \gamma(a \cdot c)$$

for $a, b, c \in k^n$ and $\beta, \gamma \in k$.

The dual of a free vector space is the function vector space

Exercise 2.8.A. For each set S, show that there is a natural isomorphism $(k^{\oplus S})^* \cong k^S$. This should be the map defined above when S is finite.

A note of caution, however, is that we have made a somewhat arbitrary choice here when identifying $k^n \cong (k^n)^*$. We could have defined $v^* = [a_n \ \cdots \ a_1]$, and the map would have still been an isomorphism. What this suggests is that given an abstract vector space V, there is no *canonical* (i.e., free of arbitrary choices) isomorphism $V \to V^*$, even if V is finite-dimensional. Of course, you can choose a basis, i.e., an isomorphism $V \cong k^n$ and then identify $V \cong k^n \cong (k^n)^* \cong V^*$ through k^n. But without choosing a basis, there is just no way to identify V and V^*.

Let me spell out the map $V \cong V^*$ (after choosing a basis) more explicitly. Suppose $\dim_k V = n$ and choose an ordered basis $B = (v_1, v_2, \ldots, v_n)$. The assignment $\Phi(v_i) = e_i$ then gives an isomorphism

$$\Phi : V \to k^n; \quad \sum_{i=1}^n a_i v_i \mapsto \sum_{i=1}^n a_i e_i = (a_1, \ldots, a_n),$$

where e_i are the standard basis vectors. Using this isomorphism, we can construct a map

$$-^* : V \xrightarrow{\Phi} k^n \xrightarrow{\cong} (k^n)^* \xrightarrow{f \mapsto f \circ \Phi} V^*.$$

This is an isomorphism, because it is a composition of isomorphisms. At the least, we have the following.

A finite-dimensional vector space and its dual have the same dimension

Proposition 2.8.3. *If V is finite-dimensional, then V^* is finite-dimensional as well and $\dim_k V = \dim_k V^*$.*

Proof. We constructed an (non-canonical) isomorphism between V and V^*. (Check that $f \mapsto f \circ \Phi$ is an isomorphism.) □

Exercise 2.8.B. Let $v_i^* \in V^*$ be the image of v_i under the above isomorphism. Show that

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

This can also be taken as the definition of v_i^*. One thing to be careful is that v_i^* is not defined using only v_i. You need the entire basis $v_1, \dots, v_n$ in order to define even a single v_i^*. The notation is highly misleading in this sense, but unfortunately it is used widely in mathematics. The following exercise will demonstrate to you how careful you need to be when doing computations with the dual vector space.

The identification with its dual requires the choice of an entire ordered basis

Exercise 2.8.C. Let us take $V = \mathbb{R}^2$ with $k = \mathbb{R}$.

(a) Consider the ordered basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$. There is an isomorphism $-^{*e} : V \to V^*$ corresponding to this ordered basis. (We have put the superscript e to denote that it comes frome e_1, e_2.) Write e_1^{*e} and e_2^{*e} as 1×2 matrices.

(b) Change one vector and take $v_1 = e_1 = (1, 0)$ and $v_2 = (1, 1)$. Consider the isomorphism $-^{*v} : V \to V^*$ corresponding to this ordered basis. Write v_1^{*v} and v_2^{*v} as 1×2 matrices.

(c) Note that $v_1^{*v} \neq e_1^{*e}$ even though $e_1 = v_1$. Express e_1^{*v} and e_2^{*v} as 1×2 matrices.

I hope the discussion above gives enough intuition about how dual spaces behave. Let us now discuss the more formal properties of the dual. Suppose we have a linear map $f : V \to W$ between vector spaces. This induces a map

$$f^* : W^* \to V^*; \quad \alpha \mapsto \alpha \circ f.$$

We actually talked about this when discussing Hom_k. In fact, dualizing is a special case of Hom_k, so everything we have proven about Hom_k holds.

Exercise 2.8.D. For vector spaces V and W, show that there is a natural isomorphism $(V \oplus W)^* \cong V^* \oplus W^*$.

Dual of an injective map is surjective, dual of a surjective map is injective

Proposition 2.8.4. *Consider a linear map $f : V \to W$ between vector spaces. If f is injective, then f^* is surjective. If f is surjective, then f^* is injective.*

Proof. Let us first show that if f is surjective then f^* is injective. To check that $f^* : W^* \to V^*$ is injective, we only need to show that $\ker f^* = 0$. Consider any $\alpha \in W^*$ such that $f^*(\alpha) = \alpha \circ f = 0$. Then $\alpha(f(v)) = 0$ for all $v \in V$. Because f is surjective, $f(v)$ ranges over all vectors in W. This shows that $\alpha(w) = 0$ for all $w \in W$, and hence $\alpha = 0$.

We now show that if f is injective then f^* is surjective. Given an arbitrary $\beta \in V^*$, we need to show that there exists an $\alpha \in W^*$ such that $f^*(\alpha) = \alpha \circ f = \beta$. By this condition $\alpha \circ f = \beta$, such a map α restricted to $f(V) \subseteq W$ is uniquely determined. That is, the map $\alpha_0 = \alpha \circ f^{-1} : f(V) \to k$ is well-defined, and $\alpha \circ f = \beta$ is equivalent to $\alpha|_{f(V)} = \alpha_0$. What we need to do is to extend this linear map $\alpha_0 : f(V) \to k$ to a linear map $\alpha : W \to k$. We do this by picking a basis. Let $B_1 \subseteq f(V)$ be a basis of $f(V)$. By Exercise 2.7.F, there exists a $B_2 \subseteq W$, disjoint from B_1, such that $B_1 \cup B_2$ is a basis of W. Let define $\alpha : W \to k$ by

$$\alpha(b) = \alpha_0(b) \text{ if } \alpha \in B_1, \quad \alpha(b) = 0 \text{ if } \alpha \in B_2.$$

Because $\operatorname{span} B_1 = f(V)$ and $\alpha(b) = \alpha_0(b)$ for $b \in B_1$, we see that $\alpha|_{f(V)} = \alpha_0$. On the other hand, this is clearly gives a linear map $W \to k$ because $B_1 \cup B_2$ is a basis of W. □

Dualizing is exact

Exercise 2.8.E. Let $V_1 \to V_2 \to V_3$ be an exact sequence of vector spaces. Show that its dual $V_3^* \to V_2^* \to V_1^*$ is exact as well.

This actually holds in greater generality.

Contravariant Hom is exact

Exercise 2.8.F. Let $V_1 \to V_2 \to V_3$ be an exact sequence of vector spaces. For an arbitrary vector space W, show that the induced sequence

$$\operatorname{Hom}(V_3, W) \to \operatorname{Hom}(V_2, W) \to \operatorname{Hom}(V_1, W)$$

is exact. (So in Exercise 2.6.S, actually you can put a 0 at the end.)

Covariant Hom is exact

Exercise 2.8.G. This is not really relevant, but I think it's worth mentioning. Let $V_1 \to V_2 \to V_3$ be an exact sequence of vector spaces. For an arbitrary vector space W, show that the induced sequence

$$\operatorname{Hom}(W, V_1) \to \operatorname{Hom}(W, V_2) \to \operatorname{Hom}(W, V_3)$$

is exact. (So you can put a 0 at the end of Exercise 2.6.R as well.)

Exercise 2.8.H. Consider a vector space V and a vector $v \in V$. If $f(v) = 0$ for all linear maps $f : V \to k$, then show that $v = 0$.

Exercise 2.8.I. Consider a vector space V and a subset $S = \{v_i\}_{i \in I}$. For every vector $v \in V$, show that the following are equivalent:

(i) $v \notin \operatorname{span}(S)$.

(ii) There exists a $f \in V^*$ such that $f(v_i) = 0$ for all $v_i \in S$ but $f(v) \neq 0$.

A linear equation has a solution if and only if there is no contradiction

Exercise 2.8.J. Let V be a vector space and $S = \{v_i\}_{i \in I} \subseteq V$ be a set of vectors. For $c_i \in k$, show that the following conditions are equivalent:

(i) The equations $f(v_i) = c_i$ for $i \in I$ do not have a common solution $f \in V^*$.

(ii) There exist $i_1, \dots, i_n \in I$ and $a_1, \dots, a_n \in k$ such that $a_1 v_{i_1} + a_2 v_{i_2} + \cdots + a_n v_{i_n} = 0$ but $a_1 c_{i_1} + \cdots + a_n c_{i_n} \neq 0$.

(Hint: extend v_i to $\tilde{v}_i = (v_i, c_i) \in V \oplus k$. Then (i) is equivalent to the nonexistence of $f \in (V \oplus k)^*$ such that $f(\tilde{v}_i) = 0$ for all $i \in I$ and $f(0_V, 1) = -1$. Now use the previous exercise.)

Exercise 2.8.K. For a linear $f : V \to W$, recall that we have defined $\operatorname{rank} f = \dim_k \operatorname{im} f$. If $\operatorname{rank} f < \infty$, i.e., $\operatorname{im} f$ is finite-dimensional, show that $\operatorname{rank} f = \operatorname{rank} f^*$. (Hint: if f is decomposed as $f : V \twoheadrightarrow U \hookrightarrow W$, then show that $\operatorname{rank} f = \dim U$.)

The dual map of a matrix is the transpose matrix

Exercise 2.8.L. Consider a linear map $T : k^n \to k^m$, corresponding to an $m \times n$ matrix with entries in k. Write

$$T = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{m1} & \cdots & t_{mn} \end{bmatrix}.$$

Its dual is going to be $T^* : (k^m)^* \to (k^n)^*$, and under the isomorphisms $k^m \cong (k^m)^*$ and $k^n \cong (k^n)^*$, we can regard it as a linear map $T^* : k^n \to k^m$. Show that the $n \times m$ matrix corresponding to T^* is

$$T^* = \begin{bmatrix} t_{11} & \cdots & t_{m1} \\ \vdots & \ddots & \vdots \\ t_{1n} & \cdots & t_{mn} \end{bmatrix}.$$

This is also called the **transpose matrix** of T.

Row rank equals column rank

Exercise 2.8.M. Consider a linear map $T : k^n \to k^m$, and consider it as a $m \times n$ matrix. Consider the columns of T as elements of k^m, and denote them by $v_1, \dots, v_n \in k^m$. We define the **column rank** as $\dim_k \operatorname{span}(v_1, \dots, v_n)$. Show that the column rank of T is simply $\operatorname{rank} T$, where T is regarded as a linear map. Similarly, consider the rows of T as $w_1, \dots, w_m \in k^n$, and define the **row rank** as $\dim_k \operatorname{span}(w_1, \dots, w_m)$. Show that the row rank of T is $\operatorname{rank} T^*$. Conclude that

$$\operatorname{rank} T = (\text{column rank of } T) = (\text{row rank of } T)$$

for any matrix T.

An element $\alpha \in V^*$ is, by definition, a linear map $V \to k$, so there is a "pairing" between V and V^* given by

$$V^* \times V \to k; \quad (\alpha, v) \mapsto \alpha(v).$$

This map is *not* linear, because $(\alpha_1 + \alpha_2)(v_1 + v_2) \neq \alpha_1(v_1) + \alpha_2(v_2)$. Instead, it is linear when one component is fixed. This is called a "bilinear map", which we will look at in the next chapter. The point is that the map is linear on α when $v \in V$ is fixed. That is, for each $v \in V$, we get a linear map

$$-(v) : V^* \to k; \quad \alpha \mapsto \alpha(v).$$

This then gives a canonical map

$$\Psi_V : V \to (V^*)^*; \quad v \mapsto -(v) = (\alpha \mapsto \alpha(v)).$$

Exercise 2.8.N. Show that the map Ψ_V is injective for every vector space V.

A finite-dimensional vector space is isomorphic to its double dual

Exercise 2.8.O. Show that if V is finite-dimensional, then Ψ_V is an isomorphism. (Hint: look at the dimension)

Unfortunately, if V is infinite-dimensional, the map Ψ_V is always injective but not surjective. From Exercise 2.8.A, we see that $(k^{\oplus S})^*$ can be identified with k^S. But if S is infinite, there are many linear maps $k^S \to k$ that don't come from a pairing with an element of $k^{\oplus S}$. This is one reason we won't talk a lot about duals of infinite-dimensional vector spaces.

Exercise 2.8.P. Let V and W be finite-dimensional vector spaces, and let $f : V \to W$ be a linear map. The map f induces a map $f^* : W^* \to V^*$, which then induces $f^{**} : V^{**} \to W^{**}$. Show that f^{**} is equal to f under the identifications $V \cong V^{**}$ and $W \cong W^{**}$.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\Psi_V"] & W \arrow[d, "\Psi_W"] \\ V^{**} \arrow[r, "f^{**}"] & W^{**} \end{tikzcd}$$

Exercise 2.8.Q. Let V and W be finite-dimensional vector spaces. Show that the map

$$\mathrm{Hom}_k(V, W) \to \mathrm{Hom}_k(W^*, V^*); \quad f \mapsto f^*$$

is an isomorphism of vector spaces.

In a finite-dimensional vector space, the annihilator of the annihilator is itself

Exercise 2.8.R. Let V be a finite-dimensional vector space, and let $W \subseteq V$ be a subspace. We define the **annihilator** of W as the subspace

$$W^0 = \{\alpha \in V^* : \alpha(W) = 0\} \subseteq V^*.$$

(a) Show that $W^0 = \ker(i^* : V^* \to W^*)$, where $i : W \hookrightarrow V$ is the inclusion map.

(b) Show that $\dim V = \dim W + \dim W_0$.

(c) Show that $W^{00} \subseteq V^{**}$ is equal to W under the identification $V \cong V^{**}$.

Exercise 2.8.S. Let V be a finite-dimensional vector space and $S = \{f_i\}_{i \in I} \subseteq V^*$ be a set of linear functionals. For $c_i \in k$, show that the following conditions are equivalent:

(i) The equations $f_i(v) = c_i$ for $i \in I$ do not have a common solution $v \in V$.

(ii) There exist $i_1, \ldots, i_n \in I$ and $a_1, \ldots, a_n \in k$ such that $a_1 f_{i_1} + a_2 f_{i_2} + \cdots + a_n f_{i_n} = 0$ but $a_1 c_{i_1} + \cdots + a_n c_{i_n} \neq 0$.

2.9 Linear algebra in combinatorics

The theory of bases and dimension is difficult and delicate, as you have seen. This also means that it is a powerful tool that turn many difficult problems into trivialities. Linear algebra is fundamental to many areas of mathematics, such as algebraic geometry, differential geometry, real analysis, and so on. It is even used in combinatorics to solve seemingly unrelated problems. We will see some of those problems where linear algebra is a powerful tool. If you are not interested such applications, you may skip this section entirely. Also, the exercises in this section are going to be hard, so feel free to move on without solving all of them.

If there are vectors more than the dimension, they are linearly dependent

Exercise 2.9.A. Let V be a k-vector space with $\dim V = n$ finite. If $v_1, \ldots, v_{n+1} \in V$, there exist $a_1, \ldots, a_{n+1} \in k$, not all zero, such that

$$a_1 v_1 + a_2 v_2 + \cdots + a_{n+1} v_{n+1} = 0.$$

If there are more variables than homogeneous equations, there is a nontrivial solution

Exercise 2.9.B. Let $x_1, \ldots, x_n$ be n variables, and consider $n-1$ equations

$$a_{j,1} x_1 + a_{j,2} x_2 + \cdots + a_{j,n} x_n = 0$$

for $1 \leq j \leq n-1$, where $a_{j,i} \in k$. Show that there always exists a nonzero solution $(x_1, \ldots, x_n) \in k^n$ satisfying all the $n-1$ equations. (Nonzero means that $(x_1, \ldots, x_n) \neq (0, \ldots, 0)$.)

A lot of the examples will rely on these two results, that $n+1$ elements in an n-dimensional vector space are always linearly dependent, and that any $n-1$ equations with n variables have a nonzero solution. In each situation, a clever choice of the base field k and the vectors v_i will give us some useful information. The first example is called the "Oddtown problem".

Theorem 2.9.1 (Oddtown problem)**.** *Let $A_1, \ldots, A_m \subseteq \{1, 2, \ldots, n\}$ be subsets such that $|A_i|$ is odd for each $1 \leq i \leq n$ and $|A_i \cap A_j|$ is even for each $1 \leq i < j \leq n$. Then $m \leq n$.*

Proof. Consider $k = \mathbb{F}_2$, the field with two elements with addition and multiplication modulo 2. For each set A_i, consider the vector $v_i \in \mathbb{F}_2^n$ such that $(v_i)_j = 1$ if $j \in A_i$ and $(v_i)_j = 0$ if $j \notin A_i$. Using the standard inner product on $\mathbb{F}_2^n$, we can write the condition $|A_i|$ odd simply as $v_i \cdot v_i = 1 \in \mathbb{F}_2$, and the condition $|A_i \cap A_j|$ can be translated to $v_i \cdot v_j = 0 \in \mathbb{F}_2$.

We are now going to prove that the vectors $v_1, \ldots, v_m$ are linearly independent in $\mathbb{F}_2^n$. This will immediately show that $m \leq n$. Suppose that

$$a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0$$

in $\mathbb{F}_2^n$. Then taking the dot product with v_i, we get

$$0 = v_i \cdot 0 = v_i \cdot (a_1 v_1 + \cdots + a_m v_m) = \sum_{j=1}^{m} a_j (v_i \cdot v_j) = a_i.$$

Therefore $a_i = 0$ for all i, and this shows that $v_1, \ldots, v_m$ are indeed linearly independent. □

Exercise 2.9.C (Eventown problem)**.** Let $A_1, \ldots, A_m \subseteq \{1, 2, \ldots, n\}$ be subsets such that $|A_i|$ is even for each $1 \leq i \leq n$ and $|A_i \cap A_j|$ is odd for each $1 \leq i < j \leq n$. Show that $m \leq n$.

Exercise 2.9.D. Let $A_1, \ldots, A_m, B_1, \ldots, B_p \subseteq \{1, \ldots, n\}$ be subset such that $|A_i \cap B_j|$ is odd for all $1 \leq i \leq m$ and $1 \leq j \leq p$. Show that $mp \leq 2^{n-1}$.

Exercise 2.9.E. Let $A_1, \ldots, A_{n+1} \subseteq \{1, 2, \ldots, n\}$ be arbitrary nonempty subsets. Show that there exist nonempty disjoint subsets $I, J \subseteq \{1, 2, \ldots, n\}$ such that

$$\bigcup_{i \in I} A_i = \bigcup_{j \in J} A_j.$$

Theorem 2.9.2 (Graham–Pollak)**.** *Suppose that the edges of a complete graph K_n are partitioned into the edges of m complete bipartite graphs. Then $m \geq n - 1$.*

Proof. Label the vertices of K_n by $1, 2, \ldots, n$, and let A_i and B_i be the set of vertices on each side of the ith complete bipartite graph. Then we have the identity

$$\sum_{1 \leq i < j \leq n} x_i x_j = \sum_{t=1}^{m} \Bigg(\sum_{j \in A_t} x_j \Bigg) \Bigg(\sum_{j \in B_t} x_j \Bigg).$$

If $m \leq n - 2$, then $m + 1 \leq n - 1$ and so there exists a nonzero solution to the $m + 1$ equations

$$\sum_{j \in A_t} x_j = 0, \quad x_1 + x_2 + \cdots + x_n = 0,$$

where we work over $k = \mathbb{R}$. Then we have

$$\begin{aligned} 0 = (x_1 + \cdots + x_n)^2 &= \sum_{i=1}^n x_i^2 + 2 \sum_{1 \le i < j \le n} x_i x_j \\ &= \sum_{i=1}^n x_i^2 + 2 \sum_{t=1}^m 0 \cdot \sum_{j \in B_t} x_j = \sum_{i=1}^n x_i^2. \end{aligned}$$

But $x_i \in \mathbb{R}$, and so the only way this can be satisfied is $x_1 = \cdots = x_n = 0$. This contradicts that $(x_1, \ldots, x_n) \neq (0, \ldots, 0)$. Therefore $m \geq n - 1$. □

Exercise 2.9.F (Fisher's inequality)**.** Let $1 \leq k \leq n$ be positive integers. Suppose $A_1, \ldots, A_m \subseteq \{1, 2, \ldots, n\}$ are distinct subsets such that $|A_i \cap A_j| = k$ for each $1 \leq i < j \leq n$. Show that $m \leq n$.

Sometimes, it is very useful to set the base field as $\mathbb{Q}$, although the vector spaces become messy.

Theorem 2.9.3. *Let $r > 0$ be a positive real number. It is possible to partition a $1 \times r$ rectangle into finitely many squares, if and only if r is a rational number.*

Proof. If r is rational, it is clear that the $1 \times r$ rectangle can be cut into finitely many squares. Now suppose that r is irrational and there is a partition of the rectangle into finitely many squares. Consider a $\mathbb{Q}$-linear function $f : \mathbb{R} \to \mathbb{R}$ such that $f(1) = 1$ and $f(r) = -1$. (Such a map exists because we can take a $\mathbb{Q}$-linear map $\operatorname{span}(1, r) \to \mathbb{R}$ with $f(1) = 1, f(r) = -1$ and then extend it to $\mathbb{R}$ by choosing a basis.) Now if a rectangle of side-length a and b is partitioned into rectangles with side-length a_i and b_i, we have the equality

$$f(a)f(b) = \sum_i f(a_i)f(b_i)$$

because dividing along a side splits $f(a)$ into $f(x) + f(a - x) = f(a)$. Now if the $1 \times r$ rectangle is partitioned into squares of side-length a_i, then

$$-1 = f(1)f(-r) = \sum_i f(a_i)f(a_i) = \sum_i f(a_i)^2 \geq 0.$$

This is clearly a contradiction. □

Exercise 2.9.G. Let A be a finite set of real numbers strictly between 0 and 1, such that for each $x \in A$, there exist $a, b \in A \cup \{0, 1\}$ such that $a, b \neq x$ and $x = \frac{1}{2}(a + b)$. Show that all elements of A are rational numbers.

Exercise 2.9.H (Iran 1998)**.** Let $S \subseteq [0, 1]$ be a finite subset containing 0 and 1. Suppose that every distance between elements of S occurs in at least two different ways, except for the distance 1. Prove that S contains only rational numbers.

We end the chapter by a recent result that was proved in an elegant way using linear algebra.

Exercise 2.9.I. Let p be a prime and n be a positive integer. Consider the $\mathbb{F}_p$-vector space

$$V = \{\text{polynomials } P \text{ in } x_1, \dots, x_n \text{ with coefficients in } \mathbb{F}_p \text{ and } \deg_{x_i} P \le p-1\}.$$

(For instance, $x_1^{p-1}x_2^{p-1} \in V$ because separately the degree are at most $p-1$, even though the total degree is $2(p-1)$.) Consider the linear map

$$V \to \mathbb{F}_p^{(\mathbb{F}_p^n)}; \quad P \mapsto (P(A))_{A \in \mathbb{F}_p^n}.$$

Show that this map is an isomorphism, by showing that it is a surjection between vector spaces of equal dimension. Conclude that if a polynomial $P \in V$ vanishes at all points in $\mathbb{F}_p^n$ then $P = 0$ as a polynomial.

Theorem 2.9.4 (Dvir, 2009)**.** *Let p be a prime and n be a positive integer. Consider a subset $K \subseteq \mathbb{F}_p^n$ such that for each $x \in \mathbb{F}_p^n$, there exists an $y \in \mathbb{F}_p^n$ such that*

$$y, y+x, y+2x, \dots, y+(p-1)x \in K.$$

Then $|K| \ge \binom{p+n-1}{n}$.

Proof. Suppose that $|K| < \binom{p+n-1}{n}$. Consider the space

$$V = \{\text{polynomials in } x_1, \dots, x_n \text{ with coefficients in } \mathbb{F}_p \text{ and total degree} \le p-1\}.$$

This is a vector space over $\mathbb{F}_p$, and $\dim_{\mathbb{F}_p} V = \binom{p+n-1}{n}$. Consider the linear map

$$V \to \mathbb{F}_p^K; \quad P(x_1, \dots, x_n) \mapsto (P(A))_{A \in K}.$$

Because the dimension of the right hand side is smaller, it cannot be injective. Therefore there is a nonzero polynomial $P \in V$ such that $P(A) = 0$ for all $A \in K$.

Now for each x, find a $y \in \mathbb{F}_p^n$ such that $y + ix \in K$ for all $0 \le i < p$. Consider the polynomial

$$P_{x,y}(t) = P(y + tx),$$

which is a polynomial with one variable and coefficients in $\mathbb{F}_p$. We have $\deg P_{x,y} \le p-1$ but $P_{x,y}(t) = 0$ for all $t \in \mathbb{F}_p$. This shows that $P_{x,y}$ must be the zero polynomial. Let $0 \le \deg P = d \le p-1$, and consider the homogeneous degree d part of P, denoted by $\overline{P}$. If we look at the degree d coefficient of $P_{x,y}$, it is going to be $\overline{P}(x)$. This shows that $\overline{P}(x) = 0$ for all $x \in \mathbb{F}_p^n$, and by Exercise 2.9.I, we see that $\overline{P} = 0$ as a polynomial. This contradicts $\deg P = d$. □

Exercise 2.9.J (German TST 2004)**.** Consider a simple graph G, and suppose that all the vertices are colored white. We are allowed to do the following operation: pick a vertex v of G, change the color of v from white to black or from black to white, and also change the color of all neighbors of v. Show that it is possible to make all the vertices black after a finite number of operations.

Exercise 2.9.K (USAMO 2008 6)**.** At a certain mathematical conference, every pair of mathematicians are either friends or strangers. At mealtime, every participant eats in one of two large dining rooms. Each mathematician insists upon eating in a room which contains an even number of his or her friends. Prove that the number of ways that the mathematicians may be split between the two rooms is a power of two (i.e., is of the form 2^k for some positive integer k).

Exercise 2.9.L (Putnam 2017 A6)**.** The 30 vertices of a regular icosahedron are distinguished by labeling them $1, 2, \dots, 30$. How many ways are there to paint each edge red, white, or blue, such that each of the 20 triangular faces of the icosahedron has two edges of the same color and third edge of a different color?

Chapter 3

Multilinear algebra

We now proceed into a new territory. So far, we have been looking at linear structure, called vector spaces, and linear maps between them. When working with these spaces, we are allowed to add vectors and multiply vectors by scalars. But there are times where we would like to multiply vectors together. The inner product

$$v \cdot w = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

defined on k^n is one example. Another interesting example is the cross product

$$v \times w = (v_2 w_3 - v_3 w_2, v_3 w_1 - v_1 w_3, v_1 w_2 - v_2 w_1)$$

defined on k^3. More generally, notions such as "area" or "volume" necessarily involves multiplying vectors together in some way. All multiplications here satisfy the property that the map is linear if all but one component is fixed.

3.1 Bilinear maps and tensor products

Definition 3.1.1. Let V, W, and U be k-vector spaces. A (set) map $f : V \times W \to U$ is said to be **bilinear** if

(BL1) for each $v \in V$, the map $f(v, -) : W \to U$ is linear, and

(BL2) for each $w \in W$, the map $f(-, w) : V \to U$ is linear.

Here, $V \times W$ is not supposed to thought of as a vector space. If you regard $V \times W$ as a vector space, the bilinear map f isn't going to be linear in general. For example,

$$f(cv, cw) = cf(v, cw) = c^2 f(v, w)$$

for $c \in k$.

Bilinear maps form a vector space

Exercise 3.1.A. Show that the set $\{\text{bilinear } f : V \times W \to U\}$ is a k-vector space under the usual operations

$$(f + g)(v, w) = f(v, w) + g(v, w), \quad (cf)(v, w) = cf(v, w).$$

A bilinear map is a linear map to the Hom space

Exercise 3.1.B. Consider the natural map

$$\begin{aligned}\operatorname{Hom}_k(V, \operatorname{Hom}_k(W, U)) &\to \{\text{bilinear } f : V \times W \to U\};\\ f &\mapsto ((v, w) \mapsto f(v)(w)).\end{aligned}$$

Show that this map is an isomorphism of vector spaces. Likewise, show that the map

$$\begin{aligned}\operatorname{Hom}_k(W, \operatorname{Hom}_k(V, U)) &\to \{\text{bilinear } f : V \times W \to U\};\\ f &\mapsto ((v, w) \mapsto f(w)(v))\end{aligned}$$

is an isomorphism. In particular, we have a natural isomorphism

$$\operatorname{Hom}_k(W, \operatorname{Hom}_k(V, U)) \cong \operatorname{Hom}_k(V, \operatorname{Hom}_k(W, U)).$$

Exercise 3.1.C. For a vector space V, show that the map $V^* \times V \to k$ given by $(\alpha, v) \mapsto \alpha(v)$ is bilinear. For vector spaces V, W and $\alpha \in V^*$, $\beta \in W^*$, show that the map

$$V \times W \to k; \quad (v, w) \mapsto \alpha(v)\beta(w)$$

is bilinear.

Let us now try to classify bilinear maps. For instance, how is a bilinear map $f : k^2 \times k^2 \to V$ classified? We have

$$\begin{aligned}f(ae_1 + be_2, ce_1 + de_2) &= af(e_1, ce_1 + de_2) + bf(e_2, ce_1 + de_2)\\ &= acf(e_1, e_1) + adf(e_1, e_2) + bcf(e_2, e_1) + cdf(e_2, e_2),\end{aligned}$$

and thus f is completely determined by the four values $f(e_i, e_j) \in V$. Conversely, for arbitrary choices of $f(e_i, e_j)$, the above map is bilinear, because every coefficient has exactly one of a, b and exactly one of c, d. This shows that there is a correspondence

$$\{\text{bilinear } k^2 \times k^2 \to V\} \to V^4; \quad f \mapsto (f(e_i, e_j))_{1 \le i,j \le 2}.$$

Another way to think about this is that

$$\operatorname{Hom}_k(k^2, \operatorname{Hom}_k(k^2, V)) \cong \operatorname{Hom}_k(k^2, V^2) \cong (V^2)^2 \cong V^4.$$

Exercise 3.1.D. Show that the map

$$\{\text{bilinear } k^m \times k^n \to V\} \to V^{mn}; \quad f \mapsto (f(e_i, e_j))_{1 \le i \le m, 1 \le j \le n}$$

is an isomorphism of vector spaces.

These facts can also be stated as there being an isomorphism

$$\{\text{bilinear } k^m \times k^n \to V\} \cong \{\text{linear } k^{mn} \to V\}.$$

This motivates the definition of tensor products.

Tensor products are defined so that a bilinear map is the same as a linear map from the tensor product

Definition 3.1.2. Let V and W be vector spaces. If $\otimes : V \times W \to T$ is a bilinear map and satisfies the following property, we say that T (along with the data of the bilinear map $\otimes : V \times W \to T$) is a **tensor product**: for any vector space U and a bilinear map $b : V \times W \to U$, there exists a unique linear map $f : V \otimes W \to U$ such that $b = f \circ \otimes$.

$$\begin{tikzcd}
V \times W \arrow[r, "b"] \arrow[d, "\otimes"] & U \\
T \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

As with all universal properties, such a T, if exists, is unique up to isomorphism. So we call T *the* **tensor product** of V and W, and write $T = V \otimes W$. We also write $\otimes(v, w) = v \otimes w$. An element of $V \otimes W$ is called a **simple tensor** if it is of the form $v \otimes w$ for some $v \in V$ and $w \in W$. (Simple tensors generally don't form a subspace of T.)

This is a weird definition, because we have characterized the tensor product as something satisfying a property. Such a definition is possible because any two objects satisfying the same universal property are canonically isomorphic. But the problem is that we do not know if there exists such a vector space T. If there is no such T that satisfies the universal property, we would always have to worry about existence when writing $V \otimes W$.

The tensor product always exists

Proposition 3.1.3. *For arbitrary vector space V and W, their tensor product $V \otimes W$ always exists.*

Proof. We are going to do something crazy. Note that a (set) map $b : V \times W \to U$ is the same as a linear map

$$\Phi : k^{\oplus(V \times W)} \to U,$$

where $V \times W$ is regarded as a set in $k^{\oplus(V \times W)}$. (This is a super large vector space.) Here, $\Phi(\underline{v, w}) = b(v, w)$ by definition. We have the condition that b is bilinear, which means that

$$\begin{aligned}
b(v + v', w) &= b(v, w) + b(v', w), \quad & b(cv, w) &= cb(v, w), \\
b(v, w + w') &= b(v, w) + b(v, w'), \quad & b(v, cw) &= cb(v, w).
\end{aligned}$$

In terms of Φ, this can be translated to

$$\begin{aligned}
\Phi(\underline{v + v', w}) &= \Phi(\underline{v, w}) + \Phi(\underline{v', w}), \quad & \Phi(\underline{cv, w}) &= c\Phi(\underline{v, w}), \\
\Phi(\underline{v, w + w'}) &= \Phi(\underline{v, w}) + \Phi(\underline{v, w'}), \quad & \Phi(\underline{v, cw}) &= c\Phi(\underline{v, w}).
\end{aligned}$$

So if we consider the subspace spanned by all such relations,

$$X = \operatorname{span}\left\{\begin{matrix} \Phi(\underline{v + v', w}) - \Phi(\underline{v, w}) - \Phi(\underline{v', w}), & \Phi(\underline{cv, w}) - c\Phi(\underline{v, w}), \\ \Phi(\underline{v, w + w'}) - \Phi(\underline{v, w}) - \Phi(\underline{v, w'}), & \Phi(\underline{v, cw}) - c\Phi(\underline{v, w}) \end{matrix}\right\} \subseteq k^{\oplus(V \times W)},$$

the condition b bilinear is equivalent to $\Phi(X) = 0$. That is, a bilinear map corresponds to a linear map $k^{\oplus(V\times W)}/X \to U$.

Now define $T = k^{\oplus(V\times W)}/X$, with $v \otimes w = [\underline{(v,w)}] \in T$. Then any bilinear map $b : V \times W \to U$ factors uniquely through $\otimes : V \times W \to T$, by the discussion above. That is, T is the tensor product of V and W. □

This gives an explicit construction of $V \times W$, and it can be taken as the definition of $V \otimes W$. But we had to quotient a very large vector space by a very large subspace, so the definition by universal properties is more intuitive to understand. In fact, we are going to derive all properties of the tensor product from this universal property. But on the other hand, if you want a concrete idea, the above description can be useful. For instance, if

$$v_1 \otimes w_1 + v_2 \otimes w_2 = v_3 \otimes w_3,$$

this means that $\underline{(v_1, w_1)} + \underline{(v_2, w_2)} - \underline{(v_3, w_3)} \in X$ and so the relation can be built out of basic bilinear operations like $v \otimes (w + w') = v \otimes w + v \otimes w'$ or $(cv) \otimes w = c(v \otimes w)$.

A linera map to the Hom space is the same as a linear map from the tensor product

Exercise 3.1.E. Show that the natural map

$$\mathrm{Hom}_k(V \otimes W, U) \to \mathrm{Hom}_k(V, \mathrm{Hom}_k(W, U)); \quad f \mapsto (v \mapsto (w \mapsto f(v \otimes w)))$$

is an isomorphism of vector spaces.

Exercise 3.1.F. For an arbitrary vector space, show that the natural map

$$V \to k \otimes V; \quad v \mapsto 1 \otimes v$$

is an isomorphism.

Exercise 3.1.G. Show that $\dim(k^m \otimes k^n) = mn$.

The simple tensors formed by basis vectors form a basis for the tensor product

Proposition 3.1.4. *Let V and W be two vector spaces, with bases $\{v_i\}_{i\in I}$ and $\{w_j\}_{j\in J}$. Then $\{v_i \otimes w_j\}_{i\in I, j\in J}$ is a basis of $V \otimes W$.*

Proof. Let us first show that $v_i \otimes w_j$ generate $V \otimes W$. Because every $v \in V$ is a linear combination of v_is and every $w \in W$ is a linear combination of w_js, every $v \otimes w$ is a linear combinations of $v_i \otimes w_j$s. On the other hand, the construction in Proposition 3.1.3 immediately implies that $V \otimes W$ is generated by the simple tensors $v \otimes w$ for $v \in V$ and $w \in W$. This shows that $V \otimes W$ is generated by $v_i \otimes w_j$. (Here is another way to see this. Any linear map $f : V \otimes W \to U$ such that $f(v \otimes w) = 0$ has to be 0 because the corresponding bilinear map $V \times W \to U$ is 0. This shows that $V \otimes W$ is generated by $v \otimes w$.)

Now let us show that the vectors $v_i \otimes w_j$ are linearly independent. Suppose that

$$\sum_{i\in I}\sum_{j\in J} a_{ij}(v_i \otimes w_j) = 0 \in V \otimes W,$$

where $a_{ij} = 0$ except for finitely many $(i,j) \in I \times J$. For each $i \in I$ and $j \in J$, there are functionals $v_i^* \in V^*$ and $w_j^* \in W^*$. Thus the map

$$V \times W \to k; \quad (v,w) \mapsto v_i^*(v) w_j^*(w)$$

is bilinear, and corresponds to a linear map $f_{ij} : V \otimes W \to k$. This should satisfy $f_{ij}(v \otimes w) = v_i^*(v)w_j^*(w)$, and hence

$$\begin{aligned} 0 = f_{i_0 j_0}(0) &= f_{i_0 j_0}\left(\sum_{i\in I}\sum_{j \in J} a_{ij} v_i \otimes w_j\right) \\ &= \sum_{i\in I}\sum_{j\in J} a_{ij} f_{i_0 j_0}(v_i \otimes w_j) = \sum_{i\in I}\sum_{j\in J} a_{ij} v_{i_0}^*(v_i) w_{j_0}^*(w_j) = a_{i_0 j_0}. \end{aligned}$$

Therefore all a_{ij} are 0, which means that $v_i \otimes w_j$ are linearly independent. □

This also can be taken as a definition for the tensor product $V \otimes W$. But then the problem is that it depends on the choice of bases $\{v_i\}$ and $\{w_j\}$. One would have to check that a change of basis induces an isomorphism between the two tensor products, and this will not be fun.

The dimension of the tensor product is the product of the dimensions

Corollary 3.1.5. *If V and W are finite-dimensional vector spaces, then $V \otimes W$ is finite-dimensional and*

$$\dim_k(V \otimes W) = \dim_k V \cdot \dim_k W.$$

Let us now discuss some of the more formal properties of tensor products. Let $f : V \to W$ be a linear map and U be an arbitrary vector space. We would like to say that f induces a map

$$f \otimes \mathrm{id}_U : V \otimes U \to W \otimes U; \quad v \otimes u \mapsto f(v) \otimes u.$$

How can we construct this map? The above is not really a definition because $v \otimes u$ cover only very special elements in $V \otimes U$. So we appeal to the universal property. Such a map should correspond to a bilinear map

$$V \times U \to W \otimes U; \quad (v,u) \mapsto f(v) \otimes u,$$

and it is easy to check that this is indeed bilinear. By the universal property, it induces a linear map $f \otimes \mathrm{id}_U : V \otimes U \to W \otimes U$, and $f \otimes \mathrm{id}_U$ should satisfy $v \otimes u \mapsto f(v) \otimes u$ by definition.

$$\begin{tikzcd} V \times U \arrow[r, "f \times \mathrm{id}_U"] \arrow[d, "\otimes"] & W \times U \arrow[d, "\otimes"] \\ V \otimes U \arrow[r, dashed] & W \otimes U \end{tikzcd}$$

This construction suggests that whenever I have an expression $E(v,w)$ that is bilinear in v and w, the map

$$V \otimes W \to (\mathrm{sth}); \quad v \otimes w \mapsto E(v,w)$$

is always uniquely defined. Such a definition should be interpreted as the map $V \otimes W \to$ (sth) induced from the bilinear map $V \times W \to$ (sth) with $(v, w) \mapsto E(v, w)$.

For example, suppose I have $f : V_1 \to V_2$ and $g : W_1 \to W_2$. Then the map

$$f \otimes g : V_1 \otimes W_1 \to V_2 \otimes W_2; \quad v \otimes w \mapsto f(v) \otimes g(w)$$

is a well-defined linear map.

Exercise 3.1.H. Let $f_1 : V_1 \to V_2$, $f_2 : V_2 \to V_3$, $g_1 : W_1 \to W_2$, $g_2 : W_2 \to W_3$ be linear maps. Show that $(f_2 \otimes g_2) \circ (f_1 \otimes g_1) = (f_2 \circ f_1) \otimes (g_2 \circ g_1)$.

Exercise 3.1.I. For vector spaces V and W, show that the map

$$V \otimes W \to W \otimes V; \quad v \otimes w \mapsto w \otimes v$$

is an isomorphism of vector spaces.

Tensor product is associative

Exercise 3.1.J. Let V, W, and U be vector spaces. Rigorously define the linear map

$$V \otimes (W \otimes U) \to (V \otimes W) \otimes U; \quad v \otimes (w \otimes u) \mapsto (v \otimes w) \otimes u$$

and show that it is an isomorphism. Then we are allowed to write $V \otimes W \otimes U$ without ambiguity.

Tensor product distributes over direct sum

Exercise 3.1.K. Let $\{V_i\}_{i \in I}$ and W be vector spaces. Show that the map

$$\left(\bigoplus_{i \in I} V_i\right) \otimes W \to \bigoplus_{i \in I} (V_i \otimes W); \quad (v_i)_{i \in I} \otimes w \mapsto (v_i \otimes w)_{i \in I}$$

is an isomorphism. (The chain of isomorphisms

$$\begin{aligned} \operatorname{Hom}((\textstyle\bigoplus_i V_i) \otimes W, X) &\cong \operatorname{Hom}(\textstyle\bigoplus_i V_i, \operatorname{Hom}(W, X)) \cong \textstyle\prod_i \operatorname{Hom}(V_i, \operatorname{Hom}(W, X)) \\ &\cong \textstyle\prod_i \operatorname{Hom}(V_i \otimes W, X) \cong \operatorname{Hom}(\textstyle\bigoplus_i (V_i \otimes W), X) \end{aligned}$$

suggests that the two should be isomorphic.)

Tensor product is exact

Exercise 3.1.L. Let $f : V \to W$ be a linear map and U be a vector space, so that we have an induced linear map $f \otimes \mathrm{id}_U : V \otimes U \to W \otimes U$.

(a) Show that if f is surjective, then $f \otimes \mathrm{id}_U$ is surjective.

(b) Show that if f is injective, then $f \otimes \mathrm{id}_U$ is injective. (Hint: pick a basis of U.)

(c) Show that if $V \xrightarrow{f} W \xrightarrow{g} X$ is exact, then

$$V \otimes U \xrightarrow{f \otimes \mathrm{id}_U} W \otimes U \xrightarrow{g \otimes \mathrm{id}_U} X \otimes U$$

is exact. In particular, statements like $\ker(f \otimes \mathrm{id}_U) = (\ker f) \otimes U$ are true.

Here is another interesting construction. For V and W vector spaces, we can define a linear map.

$$V^* \otimes W \to \mathrm{Hom}_k(V, W); \quad \alpha \otimes w \mapsto (\alpha(-)w : v \mapsto \alpha(v)w).$$

Tensoring with the dual of a finite-dimensional space is the same as taking Hom from that space

Exercise 3.1.M. Assume that V and W are both finite-dimensional. Show that the natural map $V^* \otimes W \to \mathrm{Hom}_k(V, W)$ defined above is an isomorphism. (Hint: after comparing dimension, it suffices to show either injectivity of surjectivity.)

Exercise 3.1.N. Let V and W be finite-dimensional vector spaces. Show that the natural map

$$V^* \otimes W^* \to (V \otimes W)^*; \quad \alpha \otimes \beta \mapsto (v \otimes w \mapsto \alpha(v)\beta(w))$$

is an isomorphism. (This is the same as $V^* \otimes W^* \to \mathrm{Hom}(V, W^*) \cong \mathrm{Hom}(V, \mathrm{Hom}(W, k)) \cong \mathrm{Hom}(V \otimes W, k) = (V \otimes W)^*$.)

So we are working with finite-dimensional vector spaces, we can always write Hom_k out as tensor products. For instance, we can simplify a complicated expression like

$$\begin{aligned}\mathrm{Hom}(\mathrm{Hom}(V, W), \mathrm{Hom}(U, X)) &\cong \mathrm{Hom}(V, W)^* \otimes \mathrm{Hom}(U, X) \\ &\cong (V^* \otimes W)^* \otimes (U^* \otimes X) \cong V^{**} \otimes W^* \otimes U^* \otimes X \\ &\cong V \otimes W^* \otimes U^* \otimes X.\end{aligned}$$

But for infinite-dimensional vector spaces, the situation becomes complicated.

Tensor product with a dual of a vector space is the same as taking finite rank linear maps from the space

Exercise 3.1.O. Let V and W be arbitrary vector spaces.

(a) Show that the natural map $V^* \otimes W \to \mathrm{Hom}_k(V, W)$ is always injective.

(b) Show that the image of $V^* \otimes W \to \mathrm{Hom}_k(V, W)$ is the subspace of $\mathrm{Hom}_k(V, W)$ consisting of finite rank linear maps $V \to W$.

Here is another interesting thing you can do. For a vector space V, we have seen that $V^* \times V \to k$ given by $(\alpha, v) \mapsto \alpha(v)$ is a bilinear map. This defines a linear map

$$\mathrm{tr}_V : V^* \otimes V \to k; \quad \alpha \otimes v \mapsto \alpha(v).$$

The trace map is induced from bilinear pairing

Definition 3.1.6. For a finite-dimensional vector space V and a linear map $f : V \to V$, we define the **trace map**

$$\mathrm{tr}_V : \mathrm{Hom}_k(V, V) \cong V^* \otimes V \to k.$$

More generally, in view of Exercise 3.1.O, we can define the trace map as

$$\mathrm{tr}_V : \{\text{linear } f : V \to V \text{ with finite rank}\} \cong V^* \otimes V \to k$$

even if V is infinite-dimensional.

Because tr_V is linear, we have, in particular, $\mathrm{tr}(f+g) = \mathrm{tr}(f) + \mathrm{tr}(g)$ and $\mathrm{tr}(cf) = c\,\mathrm{tr}(f)$.

The trace of a matrix is the sum of the diagonal entries

Exercise 3.1.P. Consider a linear map $A : k^n \to k^n$ given by the matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

Show that $\mathrm{tr}(A) = a_{11} + \cdots + a_{nn}$.

Exercise 3.1.Q. For V a finite-dimensional vector space and $f : V \to V$ a linear map, consider its dual map $f^* : V^* \to V^*$. Show that $\mathrm{tr}(f) = \mathrm{tr}(f^*)$.

Exercise 3.1.R. Let V be a finite-dimensional vector space. Consider linear maps $f, g : V \to V$, corresponding to $f, g \in V^* \otimes V$. Then $g \circ f$ is a liner map $V \to V$, and hence corresponds to $g \circ f \in V^* \otimes V$. Show that $g \circ f \in V^* \otimes V$ is the image of $f \otimes g \in V^* \otimes V \otimes V^* \otimes V$ under the map

$$\mathrm{id} \otimes \mathrm{tr} \otimes \mathrm{id} : V^* \otimes (V \otimes V^*) \otimes V \to V^* \otimes k \otimes V \cong V^* \otimes V.$$

Trace of a composition does not depend on the order of composition

Exercise 3.1.S. Let V be a finite-dimensional vector space, and let $f, g : V \to V$ be linear maps. Show that $\mathrm{tr}(f \circ g) = \mathrm{tr}(g \circ f)$. (This can be done by picking a basis, but you can use the previous exercise to do it without picking bases.) This means that you can cyclically permute compositions, but not arbitrarily. For instance, $\mathrm{tr}(f \circ g \circ h) = \mathrm{tr}(g \circ h \circ f)$ but it is not equal to $\mathrm{tr}(f \circ h \circ g)$.

Exercise 3.1.T. Let V and W be finite-dimensional vector spaces, and let $f : V \to V$ and $g : W \to W$ be linear maps. Show that $\mathrm{tr}(f \otimes g) = \mathrm{tr}(f)\,\mathrm{tr}(g)$. (Again, you are welcome to pick a basis, but you can do this without picking bases as well.)

Exercise 3.1.U. Let k have characteristic zero. For a finite-dimensional vector space V and two maps $f, g : V \to V$, is it possible that $f \circ g - g \circ f = \mathrm{id}$? What if k is allowed have positive characteristic?

3.2 Symmetric and exterior algebras

In this section, we look at variants of the tensor product. We have defined $V \otimes V$ to be the vector space that classifies bilinear maps mapping out of $V \times V$. But because the first and second component are both V, we can ask for additional properties for such bilinear maps, and try to come up with spaces classifying them. But first let us extend the notion of bilinearity to multilinearity.

Definition 3.2.1. Given vector spaces $V_1, \ldots, V_n$ and W, we say that a (set) map $f : V_1 \times V_2 \times \cdots \times V_n \to W$ is **multilinear** if for every $v_1 \in V_1, \ldots, v_n \in V_n$, the map

$$f(v_1, \ldots, v_{i-1}, -, v_{i+1}, \ldots, v_n) : V_i \to W$$

is linear.

Exercise 3.2.A. Show that $V_1 \otimes \cdots \otimes V_n$ classifies multilinear maps out of $V_1 \times \cdots \times V_n$. More concretely, given any multilinear $m : V_1 \times \cdots \times V_n \to W$, show that there exists a unique linear map $f : V_1 \otimes \cdots \otimes V_n \to W$ such that $m(v_1, \ldots, v_n) = f(v_1 \otimes \cdots \otimes v_n)$.

$$\begin{tikzcd}
V_1 \times \cdots \times V_n \arrow[r, "m"] \arrow[d, "\otimes"] & W \\
V_1 \otimes \cdots \otimes V_n \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

As we have said earlier, if $V_1 = \cdots = V_n$, then there are additional properties we can impose. In particular, we shall consider symmetry and anti-symmetry.

Definition 3.2.2. Let V and W be vector spaces. A multilinear map $f : V^n \to W$ is said to be **symmetric** if

$$f(v_1, \ldots, v_n) = f(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$$

for all permutations σ of $\{1, \ldots, n\}$ and $v_1, \ldots, v_n \in V$.

Exercise 3.2.B. Let $\alpha, \beta \in V^*$ be two linear functionals. Show that

$$V \times V \to k; \quad (v, w) \mapsto \alpha(v)\beta(w)$$

is symmetric if and only if α and β are linearly dependent.

In a similar way, let us define anti-symmetric, or alternating, maps. When $n = 2$, a map alternating should mean something like $f(v, w) = -f(w, v)$. For more variables, we need the notion of a signature of a permutation.

Definition 3.2.3. For σ a permutation of the set $\{1, 2, \ldots, n\}$, we define its **signature** as

$$\operatorname{sgn}(\sigma) = (-1)^{\#\{(i,j):1 \le i < j \le n, \sigma(i) > \sigma(j)\}}.$$

Exercise 3.2.C. We say that a permutation τ of $\{1, \ldots, n\}$ is a **transposition** if there exist $1 \le i < j \le n$ such that $\tau(i) = j$ and $\tau(j) = i$ and $\tau(k) = k$ for $k \neq i, j$. Show that every permutation can be written as a composition of transpositions.

Exercise 3.2.D. If σ is a permutation of $\{1, \ldots, n\}$ and τ is a transposition, show that $\operatorname{sgn}(\sigma \circ \tau) = -\operatorname{sgn}(\sigma)$. Deduce that if σ is a composition of m transpositions, then $\operatorname{sgn}(\sigma) = (-1)^m$. Show that $\operatorname{sgn}(\sigma_1 \circ \sigma_2) = \operatorname{sgn}(\sigma_1) \operatorname{sgn}(\sigma_2)$ for permutations σ_1 and σ_2 of $\{1, \ldots, n\}$.

The signature is the parity of the number of transpositions needed to make a permutation

Definition 3.2.4. Let V and W be vector spaces. A multilinear map $f : V^n \to W$ is said to be **skew-symmetric** or **anti-symmetric** or **alternating** if

$$f(v_1, \ldots, v_n) = 0$$

for $v_1, \ldots, v_n \in V$ with $v_i = v_j$ for some $i < j$.

Exercise 3.2.E. Show that if $f : V^n \to W$ is an alternating map, then

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

That is, switching two elements change the sign. (Hint: consider $f(\ldots, v_i + v_j, \ldots, v_i + v_j, \ldots)$.) Deduce that

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

for vectors $v_1, \ldots, v_n \in V$ and a permutation σ of $\{1, \ldots, n\}$.

But the condition $f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = \operatorname{sgn}(\sigma) f(v_1, \ldots, v_n)$ does not imply that f is alternating. If k has characteristic 2, that is if $2 = 0$ inside k, then $\operatorname{sgn}(\sigma) = 1$ for all σ and thus there is no guarantee that $f(v, v) = 0$. On the other hand if $\operatorname{char} k \neq 2$, it is straightforward to see that the two conditions are equivalent.

We are now ready to define the symmetric and exterior powers.

A multilinear symmetric map is the same as a linear map from the symmetric power

Definition 3.2.5. Let V be a vector space. Suppose S is a vector space along with a multilinear symmetric map $\cdot : V^e \to S$ such that for any other vector space W and a multilinear symmetric map $s : V^d \to W$, there exists a unique linear map $f : S \to W$ such that $s = f \circ \cdot$.

$$\begin{tikzcd}
V^d \arrow[r, "s"] \arrow[d, "\cdot"] & W \\
S \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

Then we say that S is the **symmetric power** of V, and write $S = \operatorname{Sym}^d V$. We also write $\cdot(v_1, \ldots, v_d) = v_1 \cdots v_d$, because permuting $v_1, \ldots, v_d$ does not change the product.

Exercise 3.2.F. Show that the symmetric power $\operatorname{Sym}^d V$ always exists. (Hint: Imitate the construction of the tensor product.)

The monomials formed by a basis forms a basis of the symmetric power

Exercise 3.2.G. Let $\{v_i\}_{i \in I}$ be a basis of V, where we assume that there is a total ordering of I. (Feel free to assume that I is finite.) Show that the set

$$\{v_{i_1} v_{i_2} \cdots v_{i_d} : i_1 \leq \ldots \leq i_d\}$$

is a basis of $\operatorname{Sym}^d V$. (Hint: for linear independence, imitate the proof of Proposition 3.1.4.)

Exercise 3.2.H. Let V be a finite-dimensional vector space with dimension $\dim_k V = n$. Show that $\operatorname{Sym}^d V$ is finite-dimensional as well and that $\dim_k \operatorname{Sym}^d V = \binom{d+n-1}{d}$.

The way to think about symmetric powers is as polynomials. For instance, let us consider $V = k^n$ with basis $\{e_1, \ldots, e_n\} \subseteq V$. The symmetric power $\operatorname{Sym}^d n$ will have basis

$$\{e_1^{d_1} e_2^{d_2} \cdots e_n^{d_n} : d_1 + d_2 + \cdots + d_n\} \subseteq \operatorname{Sym}^d V.$$

That is, $\operatorname{Sym}^d V$ looks like the homogeneous total degree d polynomials in the variables $e_1, \ldots, e_n$. This analogy can be pushed further.

Definition 3.2.6. Let V be a vector space. We define the **symmetric algebra** on V as

$$\operatorname{Sym}^\bullet V = \bigoplus_{d=0}^{\infty} \operatorname{Sym}^d V.$$

(Here, $\operatorname{Sym}^0 V = k$ if you think over the definition carefully.)

Then this is the direct sum of all the homogeneous polynomials, and therefore is like the space of all polynomials in the variables $e_1, \ldots, e_n$. A typical element in $\operatorname{Sym}^\bullet V$ will look like $1 + 2e_2^2 + 3e_1^2 e_3 - 2e_1 e_4^2$.

We can even multiply elements in $\operatorname{Sym}^\bullet V$ together. Consider the linear map

$$\operatorname{Sym}^d V \otimes \operatorname{Sym}^e V \to \operatorname{Sym}^{d+e} V; \quad (v_{i_1} \cdots v_{i_d}, v_{j_1} \cdots v_{j_e}) \mapsto v_{i_1} \cdots v_{i_d} v_{j_1} \cdots v_{j_e}.$$

This is well-defined, and is going to be the analogue of “multiplication” of polynomials. We then get an isomorphism

$$\{\text{polynomials in the variables } e_1, \ldots, e_n\} \quad \longleftrightarrow \quad \operatorname{Sym}^\bullet k^n$$

which not only is an isomorphism of vector spaces, but also preserves the “multiplication” structure. (A vector space with a multiplication structure is called an algebra, so this is an isomorphism of algebras.)

If V is just a finite-dimensional vector space, without a canonical isomorphism to k^n, we can still talk about $\operatorname{Sym}^\bullet V$. But this symmetric algebra will not be identified with a polynomial algebra without choosing a basis for V. This can be a useful construction, for instance, in algebraic geometry.

The symmetric power is functorial

Exercise 3.2.I. Show that a linear map $f : V \to W$ induces linear maps $f^d : \operatorname{Sym}^d V \to \operatorname{Sym}^d W$ on each degree d. Also show that the map $f^\bullet : \operatorname{Sym}^\bullet V \to \operatorname{Sym}^\bullet W$ preserves multiplication. Show that this induced map behaves well with composition, i.e., if $g : W \to U$ is another linear map then $(g \circ f)^\bullet = g^\bullet \circ f^\bullet$.

Exercise 3.2.J. Let $0 \to V \to W \to U \to 0$ be an exact sequence. Is

$$0 \to \operatorname{Sym}^d V \to \operatorname{Sym}^d W \to \operatorname{Sym}^d U \to 0$$

necessarily exact?

To us, the exterior power, classifying alternating maps, will be more important. The definition of the exterior power is almost identical to the definition of the symmetric power, except for that symmetric is replaced with alternating.

An multilinear alternating map is the same as a linear map from the exterior power

Definition 3.2.7. Let V be a vector space. Suppose E is a vector space along with a multilinear alternating map $\wedge : V^d \to E$ such that for any other vector space W and a multilinear alternating map $a : V^d \to W$, there exists a unique linear map $f : E \to W$ such that $a = f \circ \wedge$.

$$\begin{tikzcd}
V^d \arrow[r, "a"] \arrow[d, "\wedge"] & W \\
E \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

Then we say that E is the **exterior power** of V, and write $E = \bigwedge^d V$. We also write $\wedge(v_1, \dots, v_d) = v_1 \wedge \cdots \wedge v_d$.

Exercise 3.2.K. Show that the exterior power $\bigwedge^d V$ always exists.

Exercise 3.2.L. Let V be a vector space with basis $\{v_i\}_{i \in I}$, and assume that I is totally ordered. (Again, you can assume that V is finite-dimensional.) Show that

$$\{v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_d} : i_1 < i_2 < \cdots < i_d\}$$

is a basis of $\bigwedge^d V$.

Exercise 3.2.M. Let V be a finite-dimensional vector space with $\dim_k V = n$. Show that $\bigwedge^d V$ is finite-dimensional with dimension $\dim_k \bigwedge^d V = \binom{n}{d}$.

Exercise 3.2.N. Let $V = k^3$. The vector space V has basis $\{e_1, e_2, e_3\}$, and the vector space $\bigwedge^2 V$ has basis $\{e_1 \wedge e_2, e_2 \wedge e_3, e_3 \wedge e_1\}$. We then define a linear map, called the **Hodge star operator**, as

$$\star : \textstyle\bigwedge^2 V \to V; \quad e_1 \wedge e_2 \mapsto e_3, \quad e_2 \wedge e_3 \mapsto e_1, \quad e_3 \wedge e_1 \mapsto e_2.$$

Show that for two vectors $v, w \in k^3$, their cross product $v \times w$ is equal to $\star(v \wedge w)$. (If you don't know what a cross product is, take this as a definition and compute the components of $v \times w$ in terms of v_1, v_2, v_3 and w_1, w_2, w_3.)

How should we think about the exterior power $\bigwedge^d V$? The point of the exterior power is that there is some notion of orientation to it; if you switch two vectors, you flip the orientation and pick up a sign. The interpretation is that $\bigwedge^d V$ is supposed to be a d-dimensional volume element in the vector space V. You can think of $v_1 \wedge \cdots \wedge v_d$ as the volume element corresponding to the d-dimensional parallelotope with $v_1, \dots, v_d$ as sides. If two of these vectors are equal, the parallelotope becomes degenerate, and so the corresponding volume element becomes zero. If two vectors are switched, the orientation of the parallelotope is flipped and we get a minus sign.

Like with the symmetric powers, we can define multiplication between elements of exterior powers. Consider the linear map

$$\textstyle\bigwedge^d V \otimes \bigwedge^e V \to \bigwedge^{d+e} V; \quad (v_1 \wedge \cdots \wedge v_d, w_1 \wedge \cdots \wedge w_e) \mapsto v_1 \wedge \cdots \wedge v_d \wedge w_1 \wedge \cdots \wedge w_e.$$

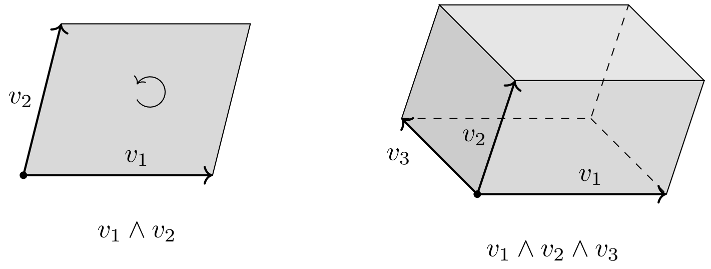

Figure 3.1: Visualizing elements of $\bigwedge^2 V$ and $\bigwedge^3 V$. Orientation of a 3-dimensional volume element is harder to describe geometrically.

Definition 3.2.8. For a vector space V, we define its **exterior algebra** as the vector space

$$\bigwedge\nolimits^\bullet V = \bigoplus_{d=0}^{\infty} \bigwedge\nolimits^d V$$

along with the multiplication maps defined above. (Here, $\bigwedge^0 V = k$ if you go back to the definition.)

For $\alpha \in \bigwedge^d V$ and $\beta \in \bigwedge^e V$, we shall simply denote by $\alpha \wedge \beta$ their product in $\bigwedge^{d+e} V$. For instance, we can do computations like

$$\begin{aligned}&(v_1 \wedge v_2 - 2v_2 \wedge v_3) \wedge (v_1 + v_3)\\ &\qquad = v_1 \wedge v_2 \wedge v_1 + v_1 \wedge v_2 \wedge v_3 - 2v_2 \wedge v_3 \wedge v_1 - 2v_2 \wedge v_3 \wedge v_3\\ &\qquad = v_1 \wedge v_2 \wedge v_3 - 2v_1 \wedge v_2 \wedge v_3 = -v_1 \wedge v_2 \wedge v_3.\end{aligned}$$

for $d = 2$ and $e = 1$.

The exterior algebra is commutative up to a sign

Exercise 3.2.O. Let $\alpha \in \bigwedge^d V$ and $\beta \in \bigwedge^e V$. Show that $\alpha \wedge \beta = (-1)^{de} \beta \wedge \alpha$ as elements of $\bigwedge^{d+e} V$.

Exercise 3.2.P. Let d be an odd positive integer. Show that $\alpha \wedge \alpha = 0$ for every $\alpha \in \bigwedge^d V$, even if k has characteristic 2 so that division by 2 is not allowed.

Exercise 3.2.Q. Show that vectors $v_1, \dots, v_d \in V$ are linearly independent if and only if $v_1 \wedge \cdots \wedge v_d \neq 0$ in $\bigwedge^d V$. (Hint: Exercise 2.7.F.) The geometric interpretation is that a parallelotope is degenerate if and only if the side vectors are linearly independent.

Of course, a linear map $f : V \to W$ induces a linear map

$$f^{\wedge d} : \bigwedge\nolimits^d V \to \bigwedge\nolimits^d W; \quad v_1 \wedge \cdots \wedge v_d \mapsto f(v_1) \wedge \cdots \wedge f(v_d).$$

Exterior algebra is functorial

Exercise 3.2.R. Verify that if $f : V \to W$ and $g : W \to U$ are linear maps, then $g^{\wedge d} \circ f^{\wedge d} = (g \circ f)^{\wedge d}$. Also show that $f^{\wedge \bullet} : \bigwedge^\bullet V \to \bigwedge^\bullet W$ preserves multiplication.

Symmetric power and exterior power of a direct sum decomposes into bidegrees

Exercise 3.2.S. For vector spaces V and W, find natural isomorphisms

$$\operatorname{Sym}^d(V \oplus W) \cong \bigoplus_{i=0}^{d} \operatorname{Sym}^i V \otimes \operatorname{Sym}^{d-i} W,$$
$$\bigwedge^d(V \oplus W) \cong \bigoplus_{i=0}^{d} \bigwedge^i V \otimes \bigwedge^{d-i} W.$$

Exterior power of a dual is the dual of the exterior power

Exercise 3.2.T. For a vector space V, show that the linear map

$$\bigwedge^d V^* \to (\bigwedge^d V)^*;$$
$$\alpha_1 \wedge \cdots \wedge \alpha_d \mapsto \left(v_1 \wedge \cdots \wedge v_d \mapsto \sum_{\sigma} \operatorname{sgn}(\sigma)\alpha_1(v_{\sigma(1)}) \cdots \alpha_d(v_{\sigma(d)}) \right)$$

where σ ranges over all permutations of $\{1, \ldots, d\}$, is well-defined and an isomorphism. Thus an element of $\bigwedge^d V^*$ inputs a d-dimensional volume form and outputs a number. It can then be thought of as something that can measure volume.

3.3 The determinant

We finally get to define the determinant. Intuitively, the determinant measures the change of volume in the linear transformation. Let us look at an example. Consider the linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the matrix

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 0 \end{bmatrix}.$$

What this map does to the standard basis vectors is depicted in Figure 3.2. The area of the parallelogram defined by e_1 and e_2 gets orientation-reversed, and the area becomes twice. In fact, given any reasonable finite shape in $\mathbb{R}^2$, the map A is going to reverse its orientation and double its volume. This suggests that the "factor of volume change" is some invariant of the map A that does not depend on the choice of a basis.

The determinant is the action on the top exterior power

Definition 3.3.1. Consider a finite-dimensional vector space V of dimension $\dim_k V = n$, and a linear map $f : V \to V$. Its nth exterior power $\bigwedge^n V$ is a 1-dimensional vector space, and

$$f^{\wedge n} : \bigwedge^n V \to \bigwedge^n V$$

is a linear map. Thus this map $f^{\wedge n}$ is just multiplication by some scalar in k. We define the **determinant** of f as the element $\deg f \in k$ such that

$$f^{\wedge n}(\omega) = (\det f) \cdot \omega.$$

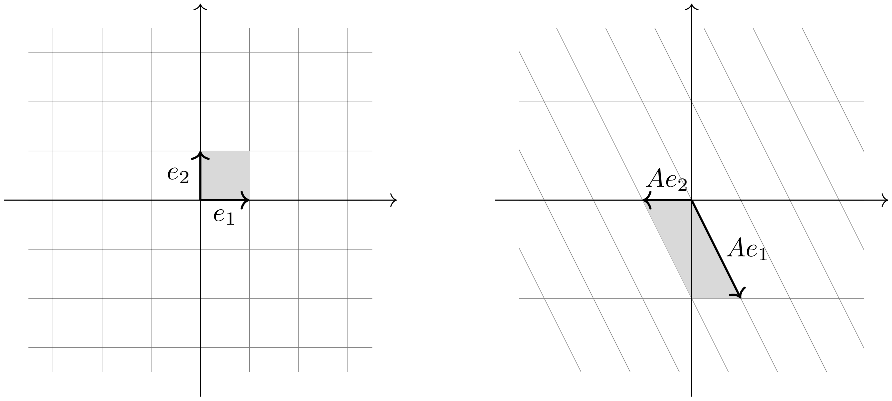

Figure 3.2: A linear map $\mathbb{R}^2 \to \mathbb{R}^2$: the map has determinant -2 because it doubles the volume and reverses orientation.

There are some obvious facts directly following from this definition. Let $f, g : V \to V$ be linear maps, where V is a finite-dimensional vector space of dimension n. Then

$$\bigwedge^n V \xrightarrow{f^{\wedge n}=\times(\det f)} \bigwedge^n V \xrightarrow{g^{\wedge n}=\times(\deg g)} \bigwedge^n V$$

and $g^{\wedge n} \circ f^{\wedge n} = (g \circ f)^{\wedge n}$ shows that

$$\det(g \circ f) = (\det g)(\det f).$$

This should be intuitively clear, because det is supposed to tell you how volume changes under the linear map.

A linear map is invertible if and only if it has nonzero determinant

Exercise 3.3.A. Let V be a finite-dimensional vector space with basis $v_1, \ldots, v_n$, and let $T : V \to V$ be a linear map.

$$(Tv_1) \wedge (Tv_2) \wedge \cdots \wedge (Tv_n) = (\det A) v_1 \wedge v_2 \wedge \cdots \wedge v_n.$$

Using Exercise 3.2.Q, conclude that T is an isomorphism (i.e., invertible) if and only if $\det T \neq 0$.

Exercise 3.3.B. Let σ be a permutation of $\{1, 2, \ldots, n\}$, and consider the map $f : k^n \to k^n$ defined by $f(e_i) = e_{\sigma(i)}$. Show that $\det f = \operatorname{sgn}(\sigma)$.

The determinant of a map is the same as the determinant of its dual map

Exercise 3.3.C. Let V be a finite-dimensional vector space, and let $f : V \to V$ be a linear map. Show that $\det(f) = \det(f^*)$.

Exercise 3.3.D. Compute the determinant of the map $A : k^2 \to k^2$ given by the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

in terms of a, b, c, d.

Now it would be nice to have a formula of a general $n \times n$ matrix, in terms of the entries. The purpose for looking for this formula is two-fold. Firstly, in terms of computations, it would be nice to have a general formula. If we ever get to do computations, such as checking that some identity holds, we could just write down the determinant, expand everything, and prove that the two sides are equal. Secondly, the mere existence of a formula is helpful from a theoretical perspective as well. If a formula that does not involve division exists, we could extend the definition to "number systems without division". If we find out that the formula is a polynomial, we find out that the set matrices with nonzero determinant forms an algebraic variety, for instance.

Hence let us look for the formula. We write the ith column of an $n \times n$ matrix A by $a_{\bullet i}$, so that A looks like

$$A = \begin{bmatrix} \\ a_{\bullet 1} & \cdots & a_{\bullet n} \\ \\ \end{bmatrix}.$$

Another way to say this is $a_{\bullet i} = Ae_i \in k^n$. But anyways, from Exercise 3.3.A applied to the standard basis vectors, we have

$$(\det A)e_1 \wedge \cdots \wedge e_n = (Ae_1) \wedge \cdots \wedge (Ae_n) = a_{\bullet 1} \wedge \cdots \wedge a_{\bullet n}.$$

Let us now expand the right hand side. We know, by definition, that $a_{\bullet j} = a_{1j}e_1 + \cdots + a_{nj}e_n$. So

$$\begin{aligned}(\det A)(e_1 \wedge \cdots \wedge e_n) &= \left(\sum_{i_1=1}^{n} a_{i_1 1} e_{i_1}\right) \wedge \cdots \wedge \left(\sum_{i_n=1}^{n} a_{i_n n} e_{i_n}\right) \\ &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} (a_{i_1 1} a_{i_2 2} \cdots a_{i_n n}) e_{i_1} \wedge \cdots \wedge e_{i_n}.\end{aligned}$$

We know that $e_{i_1} \wedge \cdots \wedge e_{i_n} = 0$ if any two of $i_1, \ldots, i_n$ are equal. Thus we may consider the sum as over $(i_1, i_2, \ldots, i_n)$ that is a permutation of $1, \ldots, n$. Then we can write

$$(\det A)(e_1 \wedge \cdots \wedge e_n) = \sum_{\sigma} a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} e_{\sigma(1)} \wedge e_{\sigma(2)} \wedge \cdots \wedge e_{\sigma(n)},$$

where σ runs over permutations of $\{1, \ldots, n\}$, i.e., bijective maps $\{1, \ldots, n\} \to \{1, \ldots, n\}$. But because taking $\wedge$ is alternating, we have $e_{\sigma(1)} \wedge \cdots \wedge e_{\sigma(n)} = \operatorname{sgn}(\sigma) e_1 \wedge \cdots \wedge e_n$. Therefore we get the following formula.

Proposition 3.3.2 (Formula for the determinant). *For $A : k^n \to k^n$ a matrix, we have*

$$\det A = \sum_{\sigma} \operatorname{sgn}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n},$$

where σ runs over all permutations of $\{1, \ldots, n\}$.

Corollary 3.3.3. *The determinant of a matrix A is an integer-coefficient polynomial in its entries a_{ij}.*

Exercise 3.3.E. Show that for any permutation σ, we have $a_{\sigma(1)1} \cdots a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)}$. Deduce that the formula for the determinant can also be written as

$$\det A = \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

Using the matrix interpretation of the dual (i.e., transpose, see Exercise 2.8.L), show that $\det A = \det A^*$. This is another proof of Exercise 3.3.C.

Exercise 3.3.F. Try to prove $\det(g \circ f) = \det(g)\det(f)$ for linear maps $f, g : k^n \to k^n$, directly from the formula. This is going to be a nice combinatorial exercise.

In a map between short exact sequences, the determinant of the middle map is the product of the determinants of the side maps

Exercise 3.3.G. Let V be a finite-dimensional vector space and let $f : V \to V$ be a linear map. Assume $W \subseteq V$ is a subspace satisfying $f(W) \subseteq W$. Let $g = f|_W : W \to W$ be the restriction of f to W and let $h : V/W \to V/W$ be the linear map defined by $[v] \mapsto [f(v)]$. (Check that this is well-defined.) Show that $\det(f) = \deg(g)\det(h)$. (You can do this in two ways: abstractly, or by picking a basis and looking at the corresponding matrix.)

Here is another way to put the same statement: if $0 \to V \to W \to U \to 0$ is a short exact sequence of finite-dimensional vector spaces and $f_V : V \to V$, $f_W : W \to W$, $f_U : U \to U$ are linear maps making the following diagram commute, then $\det(f_W) = \det(f_V)\det(f_U)$.

$$\begin{tikzcd}
0 \arrow[r] & V \arrow[r] \arrow[d, "f_V"] & W \arrow[r] \arrow[d, "f_W"] & U \arrow[r] \arrow[d, "f_U"] & 0 \\
0 \arrow[r] & V \arrow[r] & W \arrow[r] & U \arrow[r] & 0
\end{tikzcd}$$

The coefficients of the characteristic polynomial are the traces of exterior powers

Exercise 3.3.H. Let V be a n-dimensional vector space, and let $f : V \to V$ be a linear map. For any $c \in k$, show that

$$\det(c \cdot \operatorname{id} + f) = \sum_{i=0}^{n} c^{n-i} \operatorname{tr}(f^{\wedge i} : \textstyle\bigwedge^i V \to \bigwedge^i V).$$

(Hint: picking a basis sometimes makes life much easier.)

The determinant of a skew-symmetric matrix is the square of the Pfaffian

Exercise 3.3.I. Let k be a field of characteristic 0. Consider a matrix $A : k^{2n} \to k^{2n}$, and assume that the matrix A satisfies $a_{ij} = -a_{ji}$, i.e., $A^* = -A$ as matrices. Define the **Pfaffian** as

$$\operatorname{pf}(A) = \frac{1}{2^n n!} \sum_{\sigma} a_{\sigma(1)\sigma(2)} a_{\sigma(3)\sigma(4)} \cdots a_{\sigma(2n-1)\sigma(2n)},$$

where σ runs over all permutations of $\{1, \ldots, 2n\}$.

(a) Show that if we define $\omega = \sum_{1\le i<j\le 2n} a_{ij} e_i \wedge e_j \in \bigwedge^2 k^{2n}$, then

$$\frac{1}{n!}\omega^{\wedge n} = \mathrm{pf}(A) e_1 \wedge e_2 \wedge \cdots \wedge e_{2n}.$$

(b) Show that $\det(A) = \mathrm{pf}(A)^2$.

Exercise 3.3.J. Let $A, B, C, D : k^n \times k^n$ be $n\times n$ matrices such that $CD = DC$. Show that either $\det D = 0$ or the determinant of the $2n \times 2n$ matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is equal to $\det(AD - BC)$. (Hint: multiply and appropriate matrix N on the right so that $\det N$ and $\det MN$ are easy to compute.)

Exercise 3.3.K (Sylvester's determinant identity)**.** Let V and W be finite-dimensional vector spaces and $f : V \to W$ and $g : W \to V$ be matrices. Show that $\det(\mathrm{id}_V + g \circ f) = \det(\mathrm{id}_W + f \circ g)$. (Hint: pick bases so that f and g are represented by $n \times m$ and $m \times n$ matrices F and G. We then want to show that $\det(I_m + GF) = \det(I_n + FG)$. Show that both sides are equal to the determinant of $[\begin{smallmatrix} I_n & F \\ -G & I_m \end{smallmatrix}]$.)

Exercise 3.3.L (J. R. Sylvester, 2000)**.** Let $A, B, C, D : k^n \times k^n$ be $n \times n$ matrices such that $CD = DC$. Show that the determinant of the $2n \times 2n$ matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

is equal to $\det(AD - BC)$. (Hint: Show that the identity for $M = [\begin{smallmatrix} A & B \\ C & (D+xI) \end{smallmatrix}]$ as a polynomial in x. Here, the determinant is a polynomial, so everything should be a polynomial in the variable x and coefficients in k.)

3.4 Computing the inverse matrix

With our discussion of the determinant, we can now devise a formula for the inverse matrix. For a finite-dimensional vector space V, we have shown in Exercise 3.3.A that a linear map $f : V \to V$ is an isomorphism (i.e., invertible) if $\det(f) \neq 0$. For $V = k^2$, this is reflected in the formula for the inverse matrix

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{where } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

because $\det(A) = ad - bc$ appears in the denominator.

Definition 3.4.1. Consider a set $S \subseteq \{1, 2, \ldots, n\}$ and write $S = \{s_1, \ldots, s_a\}$ where $a = |S|$ and $s_1 < \cdots < s_a$. We define the linear maps

$$\iota_S : k^a \to k^n; \quad (x_1, \ldots, x_a) \mapsto x_1 e_{s_1} + \cdots + x_a e_{s_a} = (\ldots, 0, x_1, 0, \ldots)$$

and

$$\pi_S : k^n \to k^a; \quad (x_1, \ldots, x_n) \mapsto x_{s_1} e_1 + \cdots + x_{s_a} e_a = (x_{s_1}, \ldots, x_{s_a}).$$

Definition 3.4.2. For a matrix $M : k^n \to k^m$ and subsets $S \subseteq \{1, \ldots, m\}$ and $T \subseteq \{1, \ldots, n\}$, we define the **minor** of A with respect to S and T by

$$M_{ST} = \pi_S \circ M \circ \iota_T : k^{|T|} \to k^{|S|}.$$

If $S = \{1, \ldots, m\} \setminus \{i\}$ and $T = \{1, \ldots, n\} \setminus \{j\}$, we will also write

$$M_{ST} = M_{\hat{i}\hat{j}}.$$

The hat means that that index is missing.

Exercise 3.4.A. If $S = \{s_1, \ldots, s_a\}$ and $T = \{t_1, \ldots, t_b\}$ with $s_1 < \cdots < s_a$ and $t_1 < \cdots < t_b$, show that

$$M_{ST} = \begin{bmatrix} m_{s_1t_1} & m_{s_1t_2} & \cdots & m_{s_1t_b} \\ m_{s_2t_1} & m_{s_2t_2} & \cdots & m_{s_2t_b} \\ \vdots & \vdots & \ddots & \vdots \\ m_{s_at_1} & m_{s_at_2} & \cdots & m_{s_at_b} \end{bmatrix}.$$

Rank of a matrix is the maximal size of an invertible minor

Exercise 3.4.B. Let V be a finite-dimensional vector space of dimension m.

(a) For a injective linear map $f : V \hookrightarrow k^n$, show that there exists a subset $S \subseteq \{1, \ldots, n\}$ with $|S| = m$ such that $\pi_S \circ f : V \to k^m$ is an isomorphism.

(b) For a surjective linear map $f : k^n \twoheadrightarrow V$, show that there exists a subset $S \subseteq \{1, \ldots, n\}$ with $|S| = m$ such that $f \circ \iota_S : k^m \to V$ is an isomorphism.

(c) Consider a linear map $M : k^n \to k^m$. Show that the rank of M is equal to the largest integer r such that there exist subsets $S \subseteq \{1, \ldots, m\}$ and $T \subseteq \{1, \ldots, n\}$ with $|S| = |T| = r$ and M_{ST} an isomorphism.

Our goal is to find the inverse matrix, and let us think about what this means. Consider a matrix $A : k^n \to k^n$, and assume it is an isomorphism. Let us look at the rows of A^{-1}, which we will denote by α_i.

$$A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

These rows α_i are elements in $(k^n)^*$, and if we plug in the columns $Ae_i = a_{\bullet i}$ of A, from the identity $A^{-1}A = I$ it follows that

$$\alpha_i(a_{\bullet j}) = \begin{cases} 1 & j = i \\ 0 & j \neq i. \end{cases}$$

That is, α_i is a linear functional on k^n that sends $a_{\bullet i}$ to 1 and all other $a_{\bullet j}$ to 0. What function can this possibly be? It is immediate that

$$k^n \to \bigwedge\nolimits^n k^n; \quad v \mapsto a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge v \wedge a_{\bullet(i+1)} \wedge \cdots \wedge a_{\bullet n}$$

sends $a_{\bullet j}$ to 0 if $j \neq i$, and sends $a_{\bullet i}$ to $a_{\bullet 1} \wedge \cdots \wedge a_{\bullet n} = (Ae_1) \wedge \cdots \wedge (Ae_n) = \det(A)e_1 \wedge \cdots \wedge e_n$. So the functional α_i should satisfy

$$\det(A)\alpha_i(v)e_1 \wedge \cdots \wedge e_n = a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge v \wedge a_{\bullet(i+1)} \wedge \cdots \wedge a_{\bullet n}.$$

Let's describe this in a more concrete way. In particular, let us plug in $v = e_j$ and see what happens. We have

$$\begin{aligned}\alpha_i(e_j)e_1 \wedge \cdots \wedge e_n &= \frac{1}{\det(A)} a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge e_j \wedge a_{\bullet(i+1)} \cdots \wedge a_{\bullet n} \\ &= \frac{(-1)^{j-1}}{\det(A)} e_j \wedge a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge a_{\bullet(i+1)} \wedge \cdots \wedge a_{\bullet n}\end{aligned}$$

and so

$$\begin{aligned}&\alpha_i(e_j)e_j \wedge e_1 \wedge \cdots \wedge e_{j-1} \wedge e_{j+1} \wedge \cdots \wedge e_n \\ &\quad = \frac{(-1)^{i+j}}{\det(A)} e_j \wedge a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge a_{\bullet(i+1)} \wedge \cdots \wedge a_{\bullet n}.\end{aligned}$$

But on the right hand side, there's a e_j among the wedges, so we get to ignore all the e_j components of $a_{\bullet l}$. This means that we are looking at $a_{\bullet l}$ as elements in $k^n/\operatorname{span}(e_j)$, under the projection map $\pi_{\hat{j}}$. This shows that

$$\begin{aligned}&e_j \wedge a_{\bullet 1} \wedge \cdots \wedge a_{\bullet(i-1)} \wedge a_{\bullet(i+1)} \wedge \cdots \wedge a_{\bullet n} \\ &\quad = (\det A_{\hat{j}\hat{i}})e_j \wedge e_1 \wedge \cdots \wedge e_{j-1} \wedge e_{j+1} \wedge \cdots \wedge e_n.\end{aligned}$$

This immediately gives the formula for A^{-1}. By definition, the ith row jth column entry of A^{-1} is $\alpha_i(e_j)$, which is

$$\alpha_i(e_j) = \frac{(-1)^{i+j}}{\det A} \det A_{\hat{j}\hat{i}}$$

The inverse matrix is the matrix of determinants of minors, divided by the determinant

Theorem 3.4.3 (Cramer's rule)**.** *The inverse matrix of $A : k^n \to k^n$ is given by*

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} \det A_{\hat{1}\hat{1}} & -\det A_{\hat{2}\hat{1}} & \cdots & (-1)^{n+1}\det A_{\hat{n}\hat{1}} \\ -\det A_{\hat{1}\hat{2}} & \det A_{\hat{2}\hat{2}} & \cdots & (-1)^{n+2}\det A_{\hat{n}\hat{2}} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}\det A_{\hat{1}\hat{n}} & (-1)^{n+2}\det A_{\hat{2}\hat{n}} & \cdots & \det A_{\hat{n}\hat{n}} \end{bmatrix}.$$

In particular, for $n = 2$ we immediately recover

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Corollary 3.4.4. *There is a matrix* $\operatorname{adj}(A)$ *whose entries are integer-coefficient polynomials in the entries of A, such that*

$$\operatorname{adj}(A)A = A\operatorname{adj}(A) = \det(A)I.$$

This matrix $\operatorname{adj}(A)$ is called the **adjugate** of A. It is clear from the construction that if A is $n \times n$, then all the entries of $\operatorname{adj}(A)$ is homogeneous of degree $n-1$ in the entries of A.

Exercise 3.4.C. From Cramer's rule deduce that

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{\hat{i}\hat{j}}.$$

for every fixed i. Give another proof of this formula by directly using the formula for the determinant.

There is another way of computing the inverse matrix. Suppose you want to compute A^{-1}, or even $\det A$ for a given $n \times n$ matrix. If you decide to use the explicit formula for the determinant you would already need to add $n!$ terms together. If, say $n = 30$, we need to add $30! \approx 2.65 \times 10^{32}$ terms. This is clearly impractical.

Definition 3.4.5. Let A be a $m \times n$ matrix, and consider the rows $a_{i\bullet}$ of A. The following operations are called **elementary row operations** on A:

(i) Switching rows—switch $a_{i\bullet}$ and $a_{j\bullet}$ so that $a'_{i\bullet} = a_{j\bullet}$ and $a'_{j\bullet} = a_{i\bullet}$ for $i \neq j$.

(ii) Multiplying rows—multiply $a_{i\bullet}$ by a nonzero constant $c \in k^{\times}$ so that $a'_{i\bullet} = c a_{i\bullet}$.

(iii) Adding rows—add $a_{j\bullet}$ times a constant $c \in k$ to $a_{i\bullet}$ so that $a'_{i\bullet} = a_{i\bullet} + c a_{j\bullet}$.

The definition of the elementary row operations is supposed to be that they are reversible operations. If if we switch the ith and jth row, we can switch them back again. If we multiply a row by a nonzero constant c, multiplying it by c^{-1} brings it back to the original row. If we add c times the jth row to the ith row, adding $-c$ times the jth row to the ith row, we recover the original ith row.

Definition 3.4.6. We define the **elementary matrices** as

- $E^{(1)}_{i,j} = \begin{bmatrix} 1 &&&&&& \\ & \ddots &&&&& \\ && 0 && 1 && \\ &&& \ddots &&& \\ && 1 && 0 && \\ &&&&& \ddots & \\ &&&&&& 1 \end{bmatrix}$,
- $E^{(2)}_{i}(c) = \begin{bmatrix} 1 &&&&&& \\ & \ddots &&&&& \\ && 1 &&&& \\ &&& c &&& \\ &&&& 1 && \\ &&&&& \ddots & \\ &&&&&& 1 \end{bmatrix}$,

- $E^{(3)}_{i,j}(c) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & c & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}.$

Elementary row operation is multiplication on the left by an elementary matrix

Exercise 3.4.D. Show that the mth elementary row operation is the same as changing the matrix A to some $E^{(m)}A$. That is, elementary row operations are just multiplying elementary matrices on the left.

Exercise 3.4.E. Show that $\det E^{(1)}_{ij} = -1$, $\det E^{(2)}_{i}(c) = c$, and $\det E^{(3)}_{ij}(c) = 1$.

Inverses of elementary matrices are elementary matrices

Exercise 3.4.F. Show that $(E^{(1)}_{ij})^{-1} = E^{(1)}_{ij}$, $(E^{(2)}_{i}(c))^{-1} = E^{(2)}_{i}(c^{-1})$, and $(E^{(3)}_{ij}(c))^{-1} = E^{(3)}_{ij}(-c)$.

So elementary matrices have simple formulas for determinants and inverses. This means that finding the determinant or inverse of EA is almost equally hard as finding the determinant or inverse of A, because

$$\det(EA) = \det(E)\det(A), \quad (EA)^{-1} = A^{-1}E^{-1}.$$

The idea is that if we can make A into a very simple form through elementary row operations, i.e., by multiplying elementary matrices on the left, then we would be able to backtrack the process and compute the determinant and inverse of A.

We are now going to describe two algorithms, together which is called **Gaussian elimination**. The idea is that we may simplify a matrix by using elementary row operations. Let us call the following algorithm Alg1.

1. Look at the first column $a_{\bullet 1}$. If this is a zero vector, no elementary row operation can do anything, so move to the next column.
2. If this column is not all zero, we would like to make sure that a_{11} is nonzero. In order to do this, pick any j such that $a_{j1} \neq 0$ and switch the 1st and jth row.
3. Divide the 1st row by a_{11} so that after this step $a_{11} = 1$.
4. For each $j > 1$, subtract a_{j1} times the 1st row from the jth row. After this step, we will have $a_{j1} = 0$ for $j > 0$. Now we won't deal with the 1st row anymore, so ignore it from now on, and move to the next column.
5. Repeat the process until you get to the last column.

Maybe I have done a terrible job of explaining the algorithm, so I included an example of a matrix processed under Alg1 in Figure 3.3.

Alg1 puts a matrix in row echelon form

Exercise 3.4.G. Let A be an $m \times n$ matrix, and let $B = (b_{ij})$ be the result when Alg1 is run on A. Show that there exist integers $1 \leq n_1 < \cdots < n_k \leq n$ with $k \leq m$ such that

$$\begin{tikzcd}[ampersand replacement=\&]
\begin{bmatrix}2&-2&0&2\\-3&3&0&-1\\2&-1&3&0\end{bmatrix} \arrow[r, "E_1^{(2)}(\frac{1}{2})"] \& \begin{bmatrix}1&-1&0&1\\-3&3&0&-1\\2&-1&3&0\end{bmatrix} \arrow[r, "E_{2,1}^{(3)}(3)"] \& \begin{bmatrix}1&-1&0&1\\0&0&0&2\\2&-1&3&0\end{bmatrix} \arrow[dll, "E_{3,1}^{(3)}(-2)"] \\
\begin{bmatrix}1&-1&0&1\\0&0&0&2\\0&1&3&-2\end{bmatrix} \arrow[r, "E_{2,3}^{(1)}"] \& \begin{bmatrix}1&-1&0&1\\0&1&3&-2\\0&0&0&2\end{bmatrix} \arrow[r, "E_3^{(2)}(\frac{1}{2})"] \& \begin{bmatrix}1&-1&0&1\\0&1&3&-2\\0&0&0&1\end{bmatrix}
\end{tikzcd}$$

Figure 3.3: Example of a 3×4 matrix going through Alg1

(i) $b_{in_i} = 1$ for all $1 \leq i \leq k$, and

(ii) $b_{ix} = 0$ when $x < n_i$ or $i > k$.

A matrix in this form is sometimes said to be in **row echelon form**.

Elementary row operations preserve the rank

Exercise 3.4.H. Let A be an $m \times n$ matrix, and let E be an $m \times m$ elementary matrix. Show that $\operatorname{rank}(EA) = \operatorname{rank}(A)$. Conclude that $\operatorname{rank}(A)$ is the number of nonzero rows when A is made into a row echelon form by Alg1.

Actually, we can further simplify the matrix by using the 1's we have obtained to cancel out all the terms above it. Consider the following algorithm, which we call Alg2.

1. Start from the second row, with first nonzero entry $a_{2j} = 1$.
2. Cancel out the entry a_{1j} by adding $-a_{1j}$ times the second row to the first row.
3. Move to the next row, whose first nonzero entry is $a_{3j'} = 1$.
4. Again, cancel out the entries $a_{1j'}$ and $a_{2j'}$ by adding the third row times a constant to the first and second rows.
5. Repeat the process until we get to the last nonzero row.

Again, as an example, I have run this algorithm on the result of Figure 3.3 in Figure 3.4.

By running Alg1 and Alg2 on the given matrix, I have shown that

$$\begin{bmatrix}1&0&3&0\\0&1&3&0\\0&0&0&1\end{bmatrix} = E_{2,3}^{(3)}(2) \cdot E_{1,3}^{(3)}(1) \cdot E_{1,2}^{(3)}(1) \cdot E_3^{(2)}(\tfrac{1}{2})$$

$$\cdot\, E_{2,3}^{(1)} \cdot E_{3,1}^{(3)}(-2) \cdot E_{2,1}^{(3)}(3) \cdot E_1^{(2)}(\tfrac{1}{2}) \cdot \begin{bmatrix}2&-2&0&2\\-3&3&0&-1\\2&-1&3&0\end{bmatrix}.$$

If we run Alg1 and Alg2 on a square matrix, and the resulting matrix happens to be the identity matrix, then we would get $I = E_1 \cdots E_n \cdot A$, and then $E_1 \cdots E_n$ will be the inverse of A.

$$\begin{tikzcd}
\begin{bmatrix}1 & -1 & 0 & 1\\ 0 & 1 & 3 & -2\\ 0 & 0 & 0 & 1\end{bmatrix} \arrow[r, "E^{(3)}_{1,2}(1)"] & \begin{bmatrix}1 & 0 & 3 & -1\\ 0 & 1 & 3 & -2\\ 0 & 0 & 0 & 1\end{bmatrix} \arrow[dl, "E^{(3)}_{1,3}(1)"'] \\
\begin{bmatrix}1 & 0 & 3 & 0\\ 0 & 1 & 3 & -2\\ 0 & 0 & 0 & 1\end{bmatrix} \arrow[r, "E^{(3)}_{2,3}(2)"] & \begin{bmatrix}1 & 0 & 3 & 0\\ 0 & 1 & 3 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}
\end{tikzcd}$$

Figure 3.4: Example of a 3×4 matrix going through Alg2

Gaussian elimination puts a matrix in row reduced echelon form

Exercise 3.4.I. Let A be an $m \times n$ matrix, and let $B = (b_{ij})$ be the result when Alg1 and then Alg2 is run on A. Show that there exist integers $1 \leq n_1 < \cdots < n_k \leq n$ with $k \leq m$ such that

(i) $b_{in_i} = 1$ for $1 \leq i \leq k$,

(ii) $b_{ix} = 0$ for $x < n_i$ or $i > k$, and

(iii) $b_{xn_i} = 0$ for $1 \leq i \leq k$ and $x \neq n_i$.

A matrix in this form is said to be in **reduced row echelon form**.

Exercise 3.4.J. Let A be an invertible $n \times n$ matrix. Show that the result of Alg1 and Alg2 run on A is the $n \times n$ identity matrix I. Use this to show that the following algorithm calculates the inverse matrix of A:

1. First consider the $n \times 2n$ matrix $\tilde{A} = \begin{bmatrix} A & I \end{bmatrix}$ defined by putting the $n \times n$ identity matrix on the right side of A.
2. Run Alg1 and then Alg2 on $\tilde{A}$.
3. Discard the first n columns to get a $n \times n$ matrix.

Doing Gaussian elimination on an augmented matrix solves a linear equation

Exercise 3.4.K. Let A be an invertible $n \times n$ matrix, and let $b \in k^n$ be a $n \times 1$ column vector. Show that the following algorithm solves the equation $Ax = b$:

1. First consider the $n \times (n+1)$ matrix $\tilde{A} = \begin{bmatrix} A & b \end{bmatrix}$ defined by augmenting A with the $n \times 1$ vector b to the right.
2. Run Alg1 and then Alg2 on $\tilde{A}$.
3. Take the $(n+1)$th column.

Chapter 4

Linear algebra without division

So far we have been studying vector spaces. These are defined over a field, which in particular has division. But we might want to do look at a more general situation. For instance, we might want to think of $\mathbb{Z}^n$ as a "vector space" of "dimension" n over $\mathbb{Z}$, even though $\mathbb{Z}$ is not a field. This can be made precise, although in this setting the trade-off is losing many the nice theorems we had in the case of fields.

Other than being just a generalization of the theory we have developed so far, this theory will also have an important application for vector spaces over fields. From a structure theory of modules over $k[t]$, we will immediately deduce Jordan normal form, in Section 4.4.

4.1 Commutative rings

A commutative ring is like a vector space, except that we do not have multiplicative inverses.

A ring is a set with addition, subtraction, and multiplication

Definition 4.1.1. A **commutative ring** R is a set with the choice of two elements $0, 1 \in R$ and two maps $+, \cdot : R \times R \to R$ satisfying the following conditions: (Here, we write $+(a, b) = a + b$ and $\cdot(a, b) = a \cdot b$.)

(R1) For all $a \in R$ we have $a + 0 = 0 + a = a$.

(R2) For all $a, b \in R$ we have $a + b = b + a$.

(R3) For all $a, b, c \in R$ we have $a + (b + c) = (a + b) + c$.

(R4) For all $a \in R$ there exists an $(-a) \in R$ such that $a + (-a) = (-a) + a = 0$.

(R5) For all $a \in R$ we have $a \cdot 1 = 1 \cdot a = a$

(R6) For all $a, b \in R$ we have $a \cdot b = b \cdot a$.

(R7) For all $a, b, c \in R$ we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

(R8) For all $a, b, c \in R$ we have $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

Here, are some standard examples.

- The ring $\{0 = 1\}$ with one element is a ring, with the obvious addition and multiplication maps.
- Any field is a ring.
- $\mathbb{Z}$ is a ring with usual addition and multiplication.
- Take any ring R, and look at the set of polynomials
$$R[t] = \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n : a_0, \dots, a_n \in R\}$$
in the variable t with coefficients in R. This is a ring with addition and multiplication given by
$$\sum_i a_i t^i + \sum_i b_i t^i = \sum_i (a_i + b_i) t^i, \quad \Big(\sum_i a_i t^i\Big)\Big(\sum_i b_i t^i\Big) = \sum_i \Big(\sum_{j+k=i} a_j b_k\Big) t^i.$$
- Take any ring R, but now look at the set of formal power series
$$R[[t]] = \{a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots : a_0, a_1, \ldots \in R\}$$
in the variable t with coefficients in R. (Here, you don't worry about convergence just take the set of all series that can possibly be written down.) This is going to be a ring with addition and multiplication given similarly.
- The set
$$\mathbb{Z}/n\mathbb{Z} = \{0, 1, 2, \dots, n - 1\}$$
is a ring with addition and multiplication defined modulo n, i.e., taking remainder of division by n after doing ordinary addition or multiplication.
- For an integer $n > 0$, the set
$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\} \subseteq \mathbb{C}$$
is a ring with usual addition and multiplication.
- The set
$$\mathbb{Z}[\tfrac{1}{2}] = \{a 2^{-b} : a \in \mathbb{Z}, b \in \mathbb{Z}_{\geq 0}\} \subseteq \mathbb{Q}$$
is a ring with usual addition and multiplication.

As you can see, there are lots and lots of rings, and the theory of commutative rings can be pretty delicate. Here is one concept in commutative ring theory (also called commutative algebra) that helps studying rings.

Definition 4.1.2. An **ideal** of a ring R is a subset $\mathfrak{a} \subseteq R$ such that

(I1) $0 \in \mathfrak{a}$,

(I2) $a, b \in \mathfrak{a}$ implies $a + b, -a \in \mathfrak{a}$,

(I3) for any $a \in \mathfrak{a}$ and $r \in R$, we have $ra \in \mathfrak{a}$.

This is not the same as a subring, because (I3) is something stronger than saying that $a, b \in \mathfrak{a}$ implies $ab \in \mathfrak{a}$.

Exercise 4.1.A. For each element $a \in R$, show that the set

$$(a) = aR = \{ra : r \in R\}$$

is an ideal of R. An ideal that can be written in this form is called a **principal ideal**.

The ring of integers is a PID

Exercise 4.1.B. Show that the ideals of $\mathbb{Z}$ are precisely $(0), (1) = \mathbb{Z}, (2), (3), \ldots$. (Hint: take the minimal positive element in the ideal.)

Definition 4.1.3. A commutative ring R is called a **principal ideal domain**, or **PID** for short, if

(PID1) $ab = 0$ implies either $a = 0$ or $b = 0$,

(PID2) every ideal of R is principal.

So for example $\mathbb{Z}$ is a principal ideal domain.

The polynomial ring over a field is a PID

Exercise 4.1.C. For k a field, show that the polynomial ring $k[t]$ is a principal ideal domain.

Exercise 4.1.D. Show that $\mathbb{Z}[i]$ is a principal ideal domain. On the other hand, show that

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\} \subseteq \mathbb{C}$$

is not a principal ideal domain.

To satisfy number-theorists, let me define the following divisibility relation.

Definition 4.1.4. Let R be a ring in general. For $x, y \in R$, we say that x **divides** y or write $x \mid y$ if there exists an $z \in R$ such that $xz = y$.

For instance, we always have $1 \mid x$ or $x \mid 0$ because $x = 1 \cdot x$ and $0 = x \cdot 0$. Of course, the motivation comes from $R = \mathbb{Z}$. If R is a field, this relation is not very useful because for instance $x \mid y$ for arbitrary $x, y \in R \setminus \{0\}$.

In a principal ideal domain, we may take the greatest common divisor by adding ideals.

The sum of ideal is an ideal

Exercise 4.1.E. Let $\mathfrak{a}, \mathfrak{b} \subseteq R$ be two ideals. Show that

$$\mathfrak{a} + \mathfrak{b} = \{a + b : a \in \mathfrak{a}, b \in \mathfrak{b}\} \subseteq R$$

is again an ideal of R.

Exercise 4.1.F. Assume that R is a principal ideal domain. Consider two elements $a, b \in R$ so that (a) and (b) are two ideals of R. Because $(a) + (b)$ is an ideal of R, we may find a $d \in R$ (it is not uniquely determined!) such that

$$(a) + (b) = (d).$$

Show that d divides both a and b, and show that if some $e \in R$ divides both a and b, then e divides d. Check that this agrees with the usual notion of a greatest common divisor when $R = \mathbb{Z}$.

We can quotient rings by ideals.

The additive cosets of an ideal in a ring form a ring

Exercise 4.1.G. Let R be a commutative ring and $\mathfrak{a} \subseteq R$ be an ideal. Show that

$$R/\mathfrak{a} = R/(x \sim a + x \text{ for } x \in R, a \in \mathfrak{a})$$

inherits a structure of a ring from R. (You first need to check that the relation is an equivalence relation. Then you will need to check that the addition and multiplication maps $[x] + [y] = [x + y]$ and $[x][y] = [xy]$ are well-defined. After this, you should check the ring axioms.)

The example $\mathbb{Z}/n\mathbb{Z}$ is actually an instance of this. Because $(n) = n\mathbb{Z}$ is an ideal of $\mathbb{Z}$, the quotient $\mathbb{Z}/n\mathbb{Z}$ naturally is a ring.

4.2 Modules

We now define an analogue of vector spaces over a ring.

A module is a set with addition and scalar multiplication over a ring

Definition 4.2.1. Fix a commutative ring R. A **module over** R, or an R**-module** is a set M with the choice of an element $0 \in M$ and two maps $+ : M \times M \to M$ and $\cdot : R \times M \to M$ satisfying the following conditions:

(M1) For all $x \in M$ we have $x + 0 = 0 + x = x$.

(M2) For all $x, y \in M$ we have $x + y = y + x$.

(M3) For all $x, y, z \in M$ we have $(x + y) + z = x + (y + z)$.

(M4) For all $x \in M$ there exists an $(-x) \in M$ such that $x+(-x) = (-x)+x = 0$.

(M5) For all $x \in M$ we have $0 \cdot x = 0$ and $1 \cdot x = x$.

(M6) For all $x \in M$ and $a, b \in R$ we have $(a \cdot b) \cdot x = a \cdot (b \cdot x)$.

(M7) For all $x \in M$ and $a, b \in R$ we have $(a + b) \cdot x = a \cdot x + b \cdot x$.

(M8) For all $x, y \in M$ and $a \in R$ we have $a \cdot (x + y) = a \cdot x + a \cdot y$.

Note that the axioms for modules look exactly the same as the axioms for vector spaces. The only difference is that we are working over a commutative ring R instead of a field k. Because a field k is also a ring, we may call a k-vector space a k-module instead.

Example 4.2.2. Take $R = \mathbb{Z}$. What are $\mathbb{Z}$-modules? If M has a structure of addition $+ : M \times M \to M$ satisfying (M1)–(M4), this is already a $\mathbb{Z}$-module. This is because for $x \in M$ and $n > 0$ an integer, we can recover multiplication as

$$nx = \overbrace{x + x + \cdots + x}^{n}.$$

Then we can define $(-n)x = -(nx)$.

So anything with addition is a $\mathbb{Z}$-module.

- $\mathbb{Z}^n$ is a $\mathbb{Z}$-module under the usual addition

  $$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n).$$

  Then scalar multiplication will be

  $$m(a_1, \dots, a_n) = (ma_1, \dots, ma_n).$$

- $\mathbb{Z}/n\mathbb{Z}$ is also a $\mathbb{Z}$-module, with addition modulo n. Scalar multiplication will then be described as

  $$a \cdot b = (ab \text{ modulo } n) \in \mathbb{Z}/n\mathbb{Z}$$

  for $a \in \mathbb{Z}$ and $b \in \mathbb{Z}/n\mathbb{Z}$.

In a module, it is not unusual that $ax = 0$ for $a \neq 0$ and $x \neq 0$. For $M = \mathbb{Z}/n\mathbb{Z}$ over $R = \mathbb{Z}$, we have

$$n \cdot 1 = 0 \in \mathbb{Z}/n\mathbb{Z}$$

while both $n \in \mathbb{Z}$ and $1 \in \mathbb{Z}/n\mathbb{Z}$ are nonzero. Here is one more example.

- Take $R = \mathbb{Z}[i]$, and take $M = \mathbb{Z}/13\mathbb{Z}$. If we define addition on M normally, and scalar multiplication as

  $$(a + bi)x = (ax + 5bx \bmod 13),$$

  then M is an R-module. This is because $5^2 \equiv -1 \bmod 13$. Here, we have $(5 - i) \cdot 1 = 0$.

Definition 4.2.3. Let M and N be R-modules. A map $f : M \to N$ is called ***R*-linear** or a ***R*-module homomorphism** if it satisfies

(L0) $f(0) = 0$,

(L1) $f(x + y) = f(x) + f(y)$ for all $x, y \in M$,

(L2) $f(ax) = af(x)$ for all $a \in R$ and $x \in M$.

Linear maps are closed under composition

Exercise 4.2.A. Check that the identity map $\mathrm{id}_M : M \to M$ is always R-linear. Also, check that the composition $g \circ f$ of two R-linear maps $f : M_1 \to M_2$ and $g : M_2 \to M_3$ is R-linear.

When we work with modules, we always need to be careful about multiplying nonzero elements to get zero. But other than this, most of the construction we have discussed in Chapter 2 works fine.

Definition 4.2.4. For as set S, we define the **free module** as

$$R^{\oplus S} = \{(f : S \to R) : f(s) \neq 0 \text{ only for finitely many } s\}.$$

(Compare with Definition 2.4.2.)

A map from a set is the same as a linear map from the free module

Exercise 4.2.B. There is a natural inclusion map $\iota : S \to R^{\oplus S}$ of sets. Prove the universal property for free modules. That is, if $f : S \to M$ is any set map, show that there exists a unique R-linear map $\tilde{f} : R^{\oplus S} \to M$ that extends f, i.e., satisfies $f = \tilde{f} \circ \iota$.

$$\begin{tikzcd}
S \arrow[r, "f"] \arrow[d, hook] & M \\
R^{\oplus S} \arrow[ur, dashed, "\tilde{f}"'] &
\end{tikzcd}$$

(Compare with Proposition 2.4.3.)

Definition 4.2.5. Let $\{M_i\}_{i \in I}$ be a set of vector spaces, where I is an indexing set. Define their **product** as

$$\prod_{i \in I} M_i = \{(x_i)_{i \in I} : x_i \in M_i\}.$$

Similarly, define their **direct sum** as

$$\bigoplus_{i \in I} M_i = \{(x_i)_{i \in I} : x_i \in M_i, x_i \neq 0 \text{ only for finitely many } i \in I\}.$$

Exercise 4.2.C. Verify the universal properties for products and for direct sums. (See Proposition 2.4.3 and Proposition 2.4.8.)

Definition 4.2.6. Let M be an R-module. Suppose that a subset $N \subseteq M$ satisfies

(SM1) $0 \in N$,

(SM2) $x_1, x_2 \in N$ implies $x_1 + x_2 \in N$,

(SM3) $x \in N$ and $a \in R$ implies $ax \in N$.

In this case, N inherits a R-module structure from M, and we call N a **submodule** of M.

An ideal is same as a submodule of the ring

Exercise 4.2.D. Check that a submodule of R (considered as an R-module) is the same thing as an ideal of R.

Definition 4.2.7. For $N \subseteq M$ a submodule, we define the **quotient** as

$$M/N = M/(x \sim x + y \text{ for } y \in N).$$

The module structure here is defined as $[x] + [y] = [x + y]$ and $a[x] = [ax]$.

Exercise 4.2.E. Check that the quotient module is well-defined, and is indeed a module.

So for instance, if $\mathfrak{a} \subseteq R$ is an ideal, the quotient $R/\mathfrak{a}$ is an R-module while it is also a commutative ring itself.

Exercise 4.2.F. State and prove the universal property for quotients, which is going to be the direct analogue of Exercise 2.5.H.

Definition 4.2.8. For an R-linear map $f : M \to N$, we define their **kernel**, **image**, **cokernel** as

$$\begin{aligned} \ker(f) &= \{x \in M : f(x) = 0\} \subseteq M, \\ \operatorname{im}(f) &= \{f(x) \in N : x \in M\} \subseteq N, \\ \operatorname{coker}(f) &= N/\operatorname{im}(f). \end{aligned}$$

Exercise 4.2.G. Prove the first isomorphism theorem for modules: if $f : M \to N$ is an R-linear map, there exists a canonical isomorphism $M/\ker(f) \cong \operatorname{im}(f)$.

Exercise 4.2.H. State and prove the universal properties of the kernel and the cokernel. (See Exercise 2.6.E and Exercise 2.6.F.)

Definition 4.2.9. For M and N two R-modules, define

$$\operatorname{Hom}_R(M, N) = \{R\text{-linear maps } f : M \to N\}$$

as an R-module, with addition and scalar multiplication defined as

$$(f + g)(x) = f(x) + g(x), \quad (af)(x) = af(x).$$

Linear maps from the ring is the module itself

Exercise 4.2.I. Show that $\operatorname{Hom}_R(R, M) \cong M$ canonically, for all R-modules M.

Hom is functorial

Exercise 4.2.J. Show that an R-linear map $f : M_1 \to M_2$ naturally induces R-linear maps $f_* : \operatorname{Hom}_R(N, M_1) \to \operatorname{Hom}_R(N, M_2)$ and $f^* : \operatorname{Hom}_R(M_2, N) \to \operatorname{Hom}_R(M_1, N)$.

In fact, everything in Sections 2.4, 2.5, and 2.6 can be done in the context of modules. If you are bored, you can always pick an arbitrary exercise in these three sections and redo it in the context of modules.

4.3 Classification of finitely generated modules over a PID

For vector spaces, we had this classification theorem.

Theorem 4.3.1. *Every vector space V over a field k is isomorphic to $V \cong k^{\oplus S}$ for some set S. (In other words, every vector space has a basis.)*

Can we expect this to be true for rings as well? For instance, take $R = \mathbb{Z}$. There are modules like $\mathbb{Z}^{\oplus 3}$, but there are also modules like $\mathbb{Z}/10\mathbb{Z}$. You can also mix these sorts of modules and have $\mathbb{Z}^{\oplus 2} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$. Such modules are clearly not free, because a nonzero scalar like 4 can annihilate nonzero elements of the module. But we still can expect some nice things to happen, in nice cases.

Definition 4.3.2. An R-module M is said to be **finitely generated** if there exists a finite number of elements $x_1, \dots, x_n \in M$ such that every $x \in M$ can be written as

$$x = a_1 x_1 + \cdots + a_n x_n$$

for $a_1, \dots, a_n \in R$.

The result we are going to prove is that if the base ring R is a principal ideal domain and M is finitely generated over R, then M has a particular structure.

A finitely generated module over a PID is a finite direct sum of quotients by ideals

Theorem 4.3.3 (Classification of finitely generated modules over a PID)**.** *Let R be a principal ideal domain. For every finitely generated module M over R, there exists a nonnegative integer $n \geq 0$ and a sequence of ideals*

$$(0) \subseteq \mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots \subseteq \mathfrak{a}_n \subsetneq R$$

such that

$$M \cong (R/\mathfrak{a}_1) \oplus (R/\mathfrak{a}_2) \oplus \cdots \oplus (R/\mathfrak{a}_n).$$

Moreover, n and $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ is uniquely determined by M.

Before start proving it, let us see what the theorem implies. Take $R = \mathbb{Z}$. Then every ideal $\mathfrak{a} \subseteq \mathbb{Z}$ looks like $(d) \subseteq \mathbb{Z}$. So the sequence of ideals look like

$$(0) \subseteq (0) \subseteq \cdots \subseteq (0) \subseteq (d_r) \subseteq (d_{r-1}) \subseteq \cdots \subseteq (d_1)$$

for $d_1, \dots, d_r > 0$. The conditions $(d_{i+1}) \subseteq (d_i)$ imply that $d_i \mid d_{i+1}$. So then M can be written like

$$M \cong \mathbb{Z}^{\oplus(n-r)} \oplus (\mathbb{Z}/d_1\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/d_r\mathbb{Z}),$$

where $d_1 \mid d_2 \mid \cdots \mid d_r$ where $d_1, \dots, d_r$ are positive integers. Once we have this structure theorem, it can be used for explicit computations.

Exercise 4.3.A. Let M be a $\mathbb{Z}$-module that is finite (as a set). Assume that for each positive integer $k > 0$, there are at most k elements $x \in M$ such that $kx = 0$. Show that M must be of the form $M \cong \mathbb{Z}/n\mathbb{Z}$ for some $n \geq 0$.

Every prime number has a primitive root

Exercise 4.3.B. Let p be a prime number, and consider the field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Take its multiplicative group

$$\mathbb{F}_p^\times = (\mathbb{Z}/p\mathbb{Z})^\times = (\mathbb{Z}/p\mathbb{Z}) \setminus \{[0]\} = \{[1], \dots, [p-1]\}.$$

(a) For any positive integer k, show that the equation $x^k \equiv 1 \bmod p$ has at most k solutions $x \in (\mathbb{Z}/p\mathbb{Z})^\times$.

(b) Using the previous exercise, show that $(\mathbb{Z}/p\mathbb{Z})^\times$ is isomorphic to $\mathbb{Z}/(p-1)\mathbb{Z}$ as a group. In other words, the prime p has a primitive root.

Let us now start proving the theorem. Because M is a finitely generated module, there exists a surjective R-linear homomorphism

$$R^{\oplus n} \twoheadrightarrow M \to 0.$$

Then we can look at the kernel

$$0 \to N \hookrightarrow R^{\oplus n} \twoheadrightarrow M \to 0.$$

The kernel N is going to be a submodule of $R^{\oplus n}$.

A submodule of a free module is free

Theorem 4.3.4. *If R is a principal ideal domain, then every submodule of a free module is free. Moreover, every submodule of a free module with finite basis is free with finite basis.*

By this theorem, we can find an isomorphism $N \cong R^{\oplus k}$ for some k. Hence we may write our short exact sequence as

$$0 \to R^{\oplus k} \xrightarrow{A} R^{\oplus n} \twoheadrightarrow M \to 0.$$

The module M is the cokernel of the linear map A, so it will be useful to analyze the linear map A up to composition by isomorphisms. Here, $A : R^{\oplus k} \to R^{\oplus n}$ can be regarded as a $n \times k$ matrix with entries in R.

Every matrix can be made into a diagonal matrix by multiplying invertible matrices

Theorem 4.3.5 (Smith normal form)**.** *Let R be a principal ideal domain, and let A be a $n \times k$ matrix with entries in R. Then there exists an invertible $n \times n$ matrix (i.e., an invertible linear map) P and an invertible $k \times k$ matrix Q such that PAQ is of the form*

$$PAQ = \begin{bmatrix} d_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & d_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix},$$

*where $d_1 \mid d_2 \mid \cdots \mid d_r$. This is sometimes called the **Smith normal form** of A.*

Recall that we have a short exact sequence

$$0 \to R^{\oplus k} \xrightarrow{A} R^{\oplus n} \xrightarrow{B} M \to 0.$$

Because P and Q are isomorphisms from $R^{\oplus k}$ and $R^{\oplus n}$ to themselves, it is not hard to see that

$$0 \to R^{\oplus k} \xrightarrow{PAQ} R^{\oplus n} \xrightarrow{BP^{-1}} M \to 0$$

is also short exact.

Exercise 4.3.C. Verify the above claim, that the modified sequence is short exact.

As a consequence, we have an isomorphism

$$M \cong R^{\oplus n} / \operatorname{im}(PAQ).$$

Because we know exactly what PAQ looks like, the right hand side can be computed. It follows that

$$M \cong R^{\oplus(n-r)} \oplus (R/d_1R) \oplus \cdots \oplus (R/d_rR).$$

If we let $\mathfrak{a}_1 = \cdots = \mathfrak{a}_{n-r} = (0)$ and $\mathfrak{a}_{n-r+i} = (d_i)$, then we can write this also as

$$M \cong \bigoplus_{i=1}^{n} (R/\mathfrak{a}_i).$$

This proves the existence part of the classification theorem (Theorem 4.3.3).

Exercise 4.3.D. Compute $R^{\oplus n} / \operatorname{im}(PAQ)$ and show that it is isomorphic to $R^{\oplus(n-r)} \oplus (R/d_1R) \oplus \cdots \oplus (R/d_rR)$.

I now owe you the proof of two theorems: Theorem 4.3.4 and Theorem 4.3.5.

Over a PID, a submodule of a free module is free

Theorem 4.3.4. *If R is a principal ideal domain, then every submodule of a free module is free. Moreover, every submodule of a free module with finite basis is free with finite basis.*

We need to prove that any submodule of $R^{\oplus S}$ is free. But here, we will only only prove this in the case when S is a finite set.

Proof. We show that any submodule of $R^{\oplus n}$ is free, by induction on n. If $n = 0$, we have $R^{\oplus n} \cong 0$ and so there is nothing to prove.

Assume that the claim is true for $n-1$. For an arbitrary submodule $N \subseteq R^{\oplus n}$, we want to prove that N is free. If we consider

$$N' = N \cap (\{0\} \oplus R^{\oplus(n-1)}) \subseteq \{0\} \oplus R^{\oplus(n-1)},$$

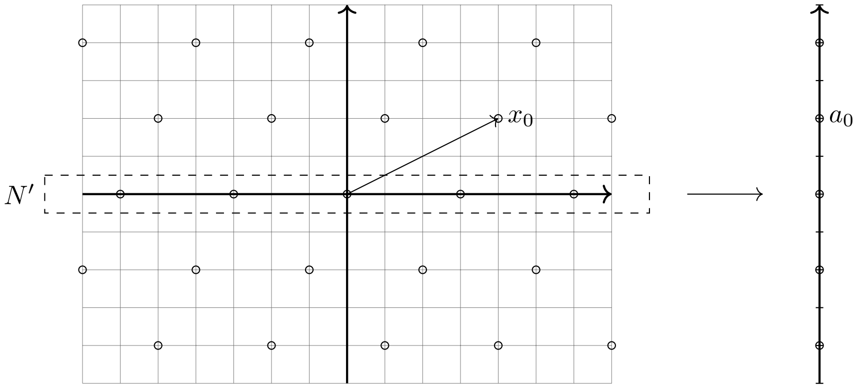

Figure 4.1: Submodule of a free module is free

this is a submodule of $R^{\oplus(n-1)}$. Then N' is a free module, because N' is a submodule of $R^{\oplus(n-1)}$.

Consider the composite R-linear map

$$\varphi : N \hookrightarrow R^{\oplus n} \xrightarrow{\pi_1} R$$

where $\pi_1 : R^{\oplus n} \to R$ is given by projection $(x_1, \ldots, x_n) \mapsto x_1$. Then the image of $N \to R$ is going to be submodule of R, which is going to be an ideal. Because R is a principal ideal domain, $\operatorname{im}\varphi = (a_0)$ for some $a_0 \in R$. Note that we have a short exact sequence

$$0 \to N' \hookrightarrow N \xrightarrow{\varphi} \operatorname{im}(\varphi) = (a_0) \to 0.$$

If $a_0 = 0$, then $\operatorname{im}\varphi = (0)$ and so N is actually contained in $\{0\} \oplus R^{\oplus(n-1)}$. Then $N = N'$ is a free module.

Suppose now that $a_0 \neq 0$. Then $a_0 \in \operatorname{im}\varphi$ shows that there exists a $x_0 \in N$ such that $\pi_1(x_0) = a_0$. Using this, we may define the map

$$N' \oplus R \to N; \quad (x, a) \mapsto x + ax_0.$$

This is clearly an R-linear map. We claim that it is bijective. To show that it is injective, suppose that $x + ax_0 = 0$. Applying π_1 to both sides give $0 = \pi_1(x) + a\pi_1(x_0) = aa_0$ because $\pi_1(x) = 0$ for $x \in N'$ and $\pi_1(x_0) = a_0$. Then $a_0 \neq 0$ and $aa_0 = 0$ implies $a = 0$ because R is a PID. It follows that $0 = x + ax_0 = x$.

To show that it is surjective, we consider an arbitrary $y \in N$ and check that it can be written as $y = x + ax_0$ for $x \in N'$ and $a \in R$. Because $\pi_1(y) \in \operatorname{im}\varphi = (a_0)$, we can write $\pi_1(y) = aa_0$ for some $a \in R$. Then $\pi_1(y - ax_0) = \pi_1(y) - a\pi_1(x_0) = 0$ and $y - ax_0 \in N$ because $y \in N$ and $x_0 \in N$. Therefore $x = y - ax_0 \in N'$ so that $y = x + ax_0$ for $x \in N'$ and $a \in R$. This shows that $\varphi : N' \oplus R \to N$ is an isomorphism of R-modules. Because N' is free, $N \cong N' \oplus R$ is also free. □

The intuition is that we can find a basis for N by looking at the sequence

$$\begin{aligned}
&N \cap (\{0\} \oplus \cdots \oplus \{0\}) = 0, \\
&N \cap (\{0\} \oplus \cdots \oplus \{0\} \oplus R), \dots, \\
&N \cap (\{0\} \oplus R \oplus \cdots \oplus R), \\
&N \cap (R \oplus \cdots \oplus R) = N \cap R^{\oplus n} = N
\end{aligned}$$

and extending the basis one at a time. To prove the theorem in the infinite rank case, we need to use Zorn's lemma.

Exercise 4.3.E. Prove the above theorem as follows. Let $R^{\oplus S}$ be our free module and $N \subseteq R^{\oplus S}$ be the submodule. We want to show that N is free. For a subset $T \subseteq S$, we have a submodule $R^{\oplus T} \subseteq R^{\oplus S}$. Consider the set

$$\{(T, \mathcal{B}) : \mathcal{B} \text{ is a basis of } N \cap R^{\oplus T}\}$$

and equip it with the partial order defined by $(T_1, \mathcal{B}_1) \prec (T_2, \mathcal{B}_2)$ if and only if $T_1 \subseteq T_2$ and $\mathcal{B}_1 \subseteq \mathcal{B}_2$. Apply Zorn's lemma (Lemma 2.7.5) to this partially ordered set.

Exercise 4.3.F. For n a nonnegative integer, consider a submodule $N \subseteq R^{\oplus n}$. Show that there exists an integer $k \leq n$ such that $N \cong R^{\oplus k}$. (We do not know yet that k is uniquely determined by N.)

Recall the second ingredient of the proof.

Every matrix can be made into a diagonal matrix by multiplying invertible matrices

Theorem 4.3.5. *Let R be a principal ideal domain, and let A be a $n \times k$ matrix with entries in R. Then there exists an invertible $n \times n$ matrix (i.e., an invertible linear map) P and an invertible $k \times k$ matrix Q such that PAQ is of the form*

$$PAQ = \begin{bmatrix} d_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & d_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix},$$

where $d_1 \mid d_2 \mid \cdots \mid d_r$.

Recall from Section 3.4 that multiplying an elementary matrix from the left is doing an elementary row operation.

An elementary column operation is right multiplication by an elementary matrix

Exercise 4.3.G. Similarly, define what an elementary column operation is, and verify that multiplying an elementary matrix from the right is doing an elementary column operation.

So what we need to do is to take an arbitrary matrix A, multiply invertible matrices or do elementary row and elementary column operations, and make it into a diagonal matrix. One thing to be careful is that we need all our operations to be invertible. So when we do the second operation, i.e., multiplying a row/column by a constant, we are allowed to only multiply by an element that divides 1. (These are called units.) To keep things simple, we just won't use the second elementary operation.

The proof is really combinatorial in nature. We first study some basic properties of princial ideal domains.

Lemma 4.3.6. *Let R be a principal ideal domain, and consider $x_1, \dots, x_n \in R$. Then there exists an element $d \in R$ satisfying the following:*

(a) d divides $x_1, \dots, x_n$.

(b) there exist $a_1, \dots, a_n \in R$ such that $d = a_1x_1 + \cdots + a_nx_n$.

(c) if $e \in R$ divides $x_1, \dots, x_n$, then e divides also d.

Proof. Consider the ideal

$$I = \{a_1x_1 + \cdots + a_nx_n : a_1, \dots, a_n \in R\} \subseteq R.$$

Because R is a principal ideal, there exists a $d \in R$ such that $I = (d)$. Then (a) follows from $x_1, \dots, x_n \in I = (d)$, and (b) follows from $d \in (d) = I$. It is also clear that (b) implies (c). □

Definition 4.3.7. Let R be a principal ideal domain. If $x_1, \dots, x_n \in R$, we will call this $d = \gcd(x_1, \dots, x_n)$ from Lemma 4.3.6 a **greatest common divisor** of $x_1, \dots, x_n$. (Note that a greatest common divisor is not unique, since it is possible that $(d) = (d')$ for $d \neq d'$.)

A PID is Noetherian, i.e., every ascending chain of ideals converges

Lemma 4.3.8. *Let R be a principal ideal domain. If*

$$\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \mathfrak{a}_3 \subseteq \cdots \subseteq R$$

is a sequence of ideals, then there exists a positive integer n such that $\mathfrak{a}_n = \mathfrak{a}_{n+1} = \cdots$. (This is saying that R is Noetherian.)

Proof. If we define

$$\mathfrak{a} = \bigcup_{i=1}^{\infty} \mathfrak{a}_i,$$

this is an ideal of R. Because R is a principal ideal domain, we have $\mathfrak{a} = (x)$ for some x. This means that $x \in \mathfrak{a} = \bigcup_{i=1}^{\infty} \mathfrak{a}_i$, so $x \in \mathfrak{a}_n$ for some n. It follows that $(x) \subseteq \mathfrak{a}_n \subseteq \mathfrak{a} = (x)$, so $\mathfrak{a}_n = \mathfrak{a}_{n+1} = \cdots$. □

Let us now describe the process.

Lemma 4.3.9. *Let R be a principal ideal domain and consider $x_1, \dots, x_n \in R$. Let $d = \gcd(x_1, \dots, x_n)$. Then there exists an $n \times n$ invertible matrix P such that*

$$P \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Proof. We do this by induction on n. If $n = 1$, there is nothing to do. For $n = 2$, we can do this explicitly. Let $x_1 = dy_1$ and $x_2 = dy_2$, so that $\gcd(y_1, y_2) = 1$. Then what we are finding is a 2×2 invertible matrix P such that

$$dP \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = d \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Because $\gcd(y_1, y_2) = 1$, there exist $z_1, z_2 \in R$ such that $y_1 z_1 + y_2 z_2 = 1$. Then we have

$$d \begin{bmatrix} z_1 & z_2 \\ -y_2 & y_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = d \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We check that the matrix is invertible because

$$\begin{bmatrix} z_1 & z_2 \\ -y_2 & y_1 \end{bmatrix} \begin{bmatrix} y_1 & -z_2 \\ y_2 & z_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now assume we can find P for $n - 1$. Then there exists a matrix P_0 such that

$$P_0 \begin{bmatrix} x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} \gcd(x_{n-1}, x_n) \\ 0 \end{bmatrix}.$$

This means that if we put 1s on the diagonal, we can build a $n \times n$ matrix P_1 such that

$$P_1 \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ \gcd(x_{n-1}, x_n) \\ 0 \end{bmatrix}.$$

Now we can apply the inductive hypothesis and find another $n \times n$ matrix P_2 so that

$$(P_2 P_1) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = P_2 \begin{bmatrix} x_1 \\ \vdots \\ \gcd(x_{n-1}, x_n) \\ 0 \end{bmatrix} = \begin{bmatrix} \gcd(x_1, \dots, x_{n-2}, \gcd(x_{n-1,x_n})) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This is what we want, by the next exercise. □

Exercise 4.3.H. Assume that R is a principal ideal domain. Then for $x_1, \dots, x_n \in R$ we have

$$\gcd(x_1, \dots, x_n) = \gcd(x_1, \dots, x_{n-2}, \gcd(x_{n-1}, x_n)).$$

Theorem 4.3.10. *Let R be a principal ideal domain, and let A be a $n \times k$ matrix with entries in R. By multiplying invertible matrices on the left and right of A, one can make A into the form described in Theorem 4.3.5.*

Proof. Let me describe the algorithm for doing this.

Step 1. First, we use Lemma 4.3.9 on the first column of A. Then we can multiply an invertible matrix on the left of A to make it into

$$\begin{bmatrix} d_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

If we use Lemma 4.3.9 on the first row, we can multiply an invertibel matrix on the right and make it into

$$\begin{bmatrix} d_2 & 0 & \cdots & 0 \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix}.$$

Here, $d_2 = \gcd(d_1, *, \ldots, *)$ and so $d_2 \mid d_1$. Now we do this again and make it into

$$\begin{bmatrix} d_3 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

As we repeat this process, we get a sequence of ideals

$$(d_1) \subseteq (d_2) \subseteq (d_3) \subseteq \cdots$$

and thus by Lemma 4.3.8, the sequence stabilizes. In other words, $(d_{a+1}) = (d_a)$ for some a. Because d_{a+1} is the greatest common divisor of everything in the first column and row, this means that when d_a is the first row first column entry, every entry in the first row or column is divisible by d_a. Without loss of generality, we may assume that the matrix looks like

$$\begin{bmatrix} d_a & d_a x_2 & \cdots & d_a x_k \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

By doing the elementary column operation that adds $(-x_i)$ times the first column to the ith column, we can make the matrix into

$$\begin{bmatrix} d_a & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

Step 2. Rewrite $e_1 = d_a$. Our matrix currently looks like

$$\begin{bmatrix} e_1 & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

If all the $*$ are divisible by e_1, we are happy. If not, assume that the ith row jth column entry is not divisible by e_1. Add the jth column to the first column, so that not all entries in the first column is divisible by e_1. That is, the matrix is

$$\begin{bmatrix} e_1 & 0 & \cdots & 0 \\ \bullet_2 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ \bullet_n & * & \cdots & * \end{bmatrix}$$

where $\bullet_2, \dots, \bullet_n$ are not all divisble by e_1. Now do Step 1 at this point. At the first application of Lemma 4.3.9, the first row first column entry becomes $\gcd(e_1, \bullet_2, \dots, \bullet_n)$.

Let us denote the result of Step 1 by

$$\begin{bmatrix} e_2 & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}.$$

Because $e_2 \mid \gcd(e_1, \bullet_2, \dots, \bullet_n)$, we have $(e_1) \subsetneq (e_2)$. If we repeat this process, we get a strictly ascending chain of ideals

$$(e_1) \subsetneq (e_2) \subsetneq \cdots.$$

This contradicts Lemma 4.3.8, and this means that the process cannot continue infinitely. That is, we arrive at a point where that matrix is

$$\begin{bmatrix} e_b & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \cdots & * \end{bmatrix}$$

where all $*$ are divisible by e_b.

Step 3. Once we have this, we do this on the $(n-1) \times (k-1)$ matrix of $*$. Then what we inductively get is a matrix that looks like

$$\begin{bmatrix} f_1 & 0 & \cdots \\ 0 & f_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

with $f_1 \mid f_2 \mid \cdots$. □

Exercise 4.3.I. Let $R = \mathbb{Z}$, which is a principal ideal domain. Using the algorithm described in the proof, find invertible matrices P and Q such that

$$P \begin{bmatrix} 1 & -1 & 3 \\ 3 & 1 & -1 \\ -3 & 1 & -3 \end{bmatrix} Q$$

is diagonal with entries dividing one another. If you're energetic, repeat this for other matrices or try to write a compute program that does this.

This is what we have proven so far.

Proposition 4.3.11. *Let R be a principal ideal domain. For every finitely generated module M over R, there exists a nonnegative integer $n \geq 0$ and a sequence of ideals*

$$(0) \subseteq \mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots \subseteq \mathfrak{a}_n \subsetneq R$$

such that

$$M \cong (R/\mathfrak{a}_1) \oplus (R/\mathfrak{a}_2) \oplus \cdots \oplus (R/\mathfrak{a}_n).$$

What we are missing is the uniqueness part. We need to show that M uniquely determine the integer n and the ideals $\mathfrak{a}_1, \dots, \mathfrak{a}_n$. Suppose we have a $\mathbb{Z}$-module

$$M = \mathbb{Z}^{\oplus 2} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}.$$

Note that the module M can be generated by the four elements $(1, 0, 0, 0), \dots, (0, 0, 0, 1)$, but it cannot be generated by three elements. We will prove that this minimal number is equal to n. Then, if we look at $\operatorname{im}(\times 2 : M \to M)$, this is isomorphic to $\mathbb{Z}^{\oplus 2} \oplus 0 \oplus \mathbb{Z}/2\mathbb{Z}$ and now can be generated by three elements. We will see how M determines the ideals $\mathfrak{a}_k$ using this marvelous idea that we can extract a lot of information by looking at the minimal number of generators of $\operatorname{im}(\times x)$, where x varies in R.

Definition 4.3.12. Let R be a ring and let M be a finitely generated R-module. The **minimal number of generators of** M is defined as the minimal nonnegative integer n such that there exist generators $x_1, \dots, x_n$ of M.

Every proper ideal is contained in a maximal ideal, in a PID

Lemma 4.3.13. *Let R be a principal ideal domain,*[1] *and let $\mathfrak{a} \subsetneq R$ be an ideal. Then there exists a larger ideal $\mathfrak{a} \subseteq \mathfrak{b} \subsetneq R$ such that $R/\mathfrak{b}$ is a field.*

Proof. If $R/\mathfrak{a}$ is a field, we can set $\mathfrak{b} = \mathfrak{a}$ and we are done. If not, there exists a nonzero element $[x_1] \in R/\mathfrak{a}$ such that $[x_1]$ does not divide $[1]$. This means that we have

$$\mathfrak{a} \subsetneq \mathfrak{a}_1 = \mathfrak{a} + (x_1) \subsetneq R.$$

[1] We actually don't need this assumption. R can be an arbitrary ring, but then we need to use Zorn's lemma.

(We have $\mathfrak{a} \subsetneq \mathfrak{a}_1$ because $[x_1] \neq [0]$ in $R/\mathfrak{a}$, and we have $\mathfrak{a}_1 \subsetneq R$ because $[x_1]$ does not divide $[1]$.) If $R/\mathfrak{a}_1$ is a field, we are done, otherwise we can find a larger ideal

$$\mathfrak{a} \subsetneq \mathfrak{a}_1 \subsetneq \mathfrak{a}_2 \subsetneq R.$$

This process cannot continue forever by Lemma 4.3.8. This means that $R/\mathfrak{a}_i$ is a field for some i. □

The number of summands in the classification of finitely generated modules is the minimal number of generators

Lemma 4.3.14. *Let R be a principal ideal domain, and let*

$$(0) \subseteq \mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots \subseteq \mathfrak{a}_n \subsetneq R$$

be a sequence of ideals. Define the R-module

$$M = (R/\mathfrak{a}_1) \oplus \cdots \oplus (R/\mathfrak{a}_n).$$

Then the minimal number of generators of M is exactly n.

Proof. It is clear that there exists a set of generators of size n, namely $([1],[0],\ldots,[0])$ through $([0],[0],\ldots,[1])$.

We now have to show that there is no set of generators of size $n-1$. Suppose there exists such a set of generators. Then by definition there is a surjective homomorphism

$$R^{\oplus(n-1)} \twoheadrightarrow (R/\mathfrak{a}_1) \oplus \cdots \oplus (R/\mathfrak{a}_n).$$

Using Lemma 4.3.13, we find an ideal

$$\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots \subseteq \mathfrak{a}_n \subseteq \mathfrak{b} \subsetneq R.$$

Now there is a surjective linear map $R/\mathfrak{a}_i \twoheadrightarrow R/\mathfrak{b}$. We can direct sum these maps together and define

$$\varphi : R^{\oplus(n-1)} \twoheadrightarrow (R/\mathfrak{a}_1) \oplus \cdots \oplus (R/\mathfrak{a}_n) \twoheadrightarrow (R/\mathfrak{b})^{\oplus n}.$$

So there is this surjective R-linear map $\varphi : R^{\oplus(n-1)} \twoheadrightarrow (R/\mathfrak{b})^{\oplus n}$. Let us write $\mathfrak{b} = (e)$. Then for any $x \in R^{\oplus(n-1)}$ we have

$$\varphi(ex) = e\varphi(x) = 0$$

because multiplication by e turns everything in $(R/\mathfrak{b})^{\oplus n}$ to zero. This means that $\ker(\varphi)$ contains the module $\mathfrak{b}^{\oplus(n-1)} \subseteq R^{\oplus(n-1)}$. Using the module version of Exercise 2.5.H, we can factor φ as

$$\varphi : R^{\oplus(n-1)} \to R^{\oplus(n-1)}/\mathfrak{b}^{\oplus(n-1)} = (R/\mathfrak{b})^{\oplus(n-1)} \xrightarrow{\psi} (R/\mathfrak{b})^{\oplus n}.$$

Here, the R-linear map $\psi : (R/\mathfrak{b})^{\oplus(n-1)} \twoheadrightarrow (R/\mathfrak{b})^{\oplus n}$ is surjective because φ is surjective.

But recall that we have set $k = R/\mathfrak{b}$ to be a field. This means that $(R/\mathfrak{b})^{\oplus(n-1)}$ and $(R/\mathfrak{b})^{\oplus n}$ are vector spaces over k. The map $\psi : k^{\oplus(n-1)} \twoheadrightarrow k^{\oplus n}$ is R-linear, and for any $[a] \in k$ we have

$$[a]\psi(x) = a\psi(x) = \psi(ax) = \psi([a]x).$$

So $\psi : k^{n-1} \to k^n$ is a surjective k-linear morphism. This contradicts Corollary 2.7.13 because $\dim_k k^{n-1} = n-1$ is smaller than $\dim_k k^n = n$. □

This lemma can be used to recover the ideals from the module. For an ideal $\mathfrak{a}_i \subseteq R$, consider the R-module $R/\mathfrak{a}_i$. For some element $d \in R$, what is the image of $\times d : R/\mathfrak{a}_i \to R/\mathfrak{a}_i$? We have a surjective map

$$R \twoheadrightarrow R/\mathfrak{a}_i \xrightarrow{\times d} \operatorname{im}(\times d : R/\mathfrak{a}_i \to R/\mathfrak{a}_i).$$

The kernel of this is the ideal

$$\mathfrak{b}_i = \{x \in R : dx \in \mathfrak{a}_i\} \subseteq R.$$

Thus the first isomorphism theorem (Exercise 4.2.G) gives an isomorphism

$$\operatorname{im}(\times d : R/\mathfrak{a}_i \to R/\mathfrak{a}_i) \cong R/\mathfrak{b}_i.$$

Now let M be the R-module

$$M = (R/\mathfrak{a}_1) \oplus \cdots \oplus (R/\mathfrak{a}_n)$$

where $\mathfrak{a}_1 \subseteq \cdots \subseteq \mathfrak{a}_n \subsetneq R$. Then the image of $\times d : M \to M$ is given by

$$\operatorname{im}(\times d : M \to M) \cong (R/\mathfrak{b}_1) \oplus \cdots \oplus (R/\mathfrak{b}_n)$$

where $\mathfrak{b}_i$ is as defined above. Also, $\mathfrak{b}_1 \subseteq \cdots \subseteq \mathfrak{b}_n \subseteq R$. Lemma 4.3.14 immediately implies that the number of minimal generators of $\operatorname{im}(\times d : M \to M)$ is the number of i such that $\mathfrak{b}_i \subsetneq R$. Also, it is not hard to see from the definition that $\mathfrak{b}_i = R$ if and only if $d \in \mathfrak{a}_i$. Therefore the minimal number of generators of $\operatorname{im}(\times d : M \to M)$ is

$$\#\{1 \leq i \leq n : d \notin \mathfrak{a}_i\}.$$

Exercise 4.3.J. From the above discussion, deduce the following. If R is a principal ideal domain, and $\mathfrak{a}_1 \subseteq \cdots \subseteq \mathfrak{a}_n \subsetneq R$, and consider the R-module

$$M = (R/\mathfrak{a}_1) \oplus \cdots \oplus (R/\mathfrak{a}_n).$$

Then n is the minimal number of generators of M, and the ideal $\mathfrak{a}_i$ can be identified as

$$\mathfrak{a}_i = \{d \in R : \text{minimal number of generators of } \operatorname{im}(\times d : M \to M) \text{ is} < i\}.$$

It immediately follows that M uniquely determines n and the ideals $\mathfrak{a}_i$. This finishes the proof of Theorem 4.3.3.

4.4 Frobenius and Jordan normal form

Let us come back to the situation of a finite-dimensional vector space over a field. Let k be a field, and let V be a finite-dimensional vector space over k. Fix a linear map $T : V \to V$. We want to know how T acts on V. For instance, can we choose a nice basis of V so that T is represented by a relatively simple matrix?

Here is another way to think about what we are trying to do. First choose an arbitrary basis of V, so that T becomes a square matrix. A change of basis is given by a conjugation by an invertible matrix, i.e., PTP^{-1}. So we are trying to find an invertible matrix P such that PTP^{-1} is a "relatively simple matrix".

Our strategy is to combine the data of a vector space V along with the data of a linear map $T : V \to V$, and make it into a data of a single module. Recall the polynomial ring.

Definition 4.4.1. Let k be a field. The **ring of polynomials** $k[t]$ is defined as

$$k[t] = \{a_0 + a_1 t + \cdots + a_n t^n : a_0, \ldots, a_n \in k\}$$

with natural addition and multiplication.

Given a vector space V and a linear map $T : V \to V$, we can define a $k[t]$-module structure on V as

$$(a_0 + \cdots + a_n t^n) \cdot v = a_0 v + a_1 T(v) + a_2 T(T(v)) + \cdots + a_n T^n(v)$$

for $x \in V$ and $a_0, \ldots, a_n \in k$.

Exercise 4.4.A. Verify that this is indeed a $k[t]$-module structure.

Conversely, suppose we are given a $k[t]$-module V. Then V is naturally a k-vector space because $k \subseteq k[t]$. Also, we can recover the linear map $T : V \to V$ by $T(v) = t \cdot v$. These are inverse constructions, and thus we can translate between these two structures.

A vector space with a linera map is the same as a module over the polynomial ring

$$\left\{\begin{matrix} \text{a } k\text{-vector space } V \\ \text{plus a linear map } T : V \to V \end{matrix}\right\} \quad \longleftrightarrow \quad \left\{\text{a } k[t]\text{-module } V\right\}$$

Exercise 4.4.B. Let V_1 be a $k[t]$-module, corresponding to a k-linear map $T_1 : V_1 \to V_1$, and let V_2 be a $k[t]$-module corresponding to a k-linear map $T_2 : V_2 \to V_2$. If $P : V_1 \to V_2$ is $k[t]$-linear and bijective, show that $PT_1P^{-1} = T_2$, as k-linear maps. This means that analyzing the $k[t]$-module structure of V_1 amounts to studying the k-linear map T_1 up to conjugation.

Exercise 4.4.C. Let A be an $n \times n$ matrix with entries in k. Then $tI_n - A$ can be considered as an $n \times n$ matrix with entries in $k[t]$. Show that the $k[t]$-module corresponding to $A : k^n \to k^n$ is isomorphic to

$$\operatorname{coker}(k[t]^{\oplus n} \xrightarrow{tI_n - A} k[t]^{\oplus n}).$$

Because V is a finite-dimensional k-vector space, it is finitely generated as a $k[t]$-module. We would like to apply the classification theory of finitely generated modules over a principal ideal domain.

The polynomial ring over a field is a PID

Theorem 4.4.2. *Let k be a field. Then $k[t]$ is a principal ideal domain.*

Proof. Take an ideal $\mathfrak{a} \subseteq k[t]$. We need to show that $\mathfrak{a} = (f)$ for some $f \in k[t]$. If $\mathfrak{a} = \{0\}$, we can set $f = 0$. If $\mathfrak{a} \supsetneq (0)$, take a nonzero polynomial $f \in \mathfrak{a}$ with minimal degree. Then we first have $(f) \subseteq \mathfrak{a}$ because $f \in \mathfrak{a}$. For the other containment, take an arbitrary element $g \in \mathfrak{a}$. By polynomial division, we can find $q, r \in k[t]$ such that

$$g = fq + r, \quad \deg r < \deg f.$$

Because $f, g \in \mathfrak{a}$, we have $r = g - fq \in \mathfrak{a}$. But $\deg r < \deg f$ contradicts the minimality of f, unless $r = 0$. This means that $g = fq \in (f)$, for every $g \in \mathfrak{a}$. This shows that $(f) = \mathfrak{a}$. □

By Theorem 4.3.3, it immediately follows that any V(a $k[t]$-module that is a finite-dimensional k-vector space) is isomorphic to

$$V \cong k[t]^{\oplus r} \oplus k[t]/(p_1(t)) \oplus \cdots \oplus k[t]/(p_s(t))$$

as a $k[t]$-module. But $k[t]$ is an infinite-dimensional k-vector space. Because V is finite-dimensional, we must have $r = 0$. Moreover, note that $(cp_i) = (p_i)$ for $c \in k^\times \setminus \{0\}$ a nonzero constant. This shows that we may assume that p_i are all monic, i.e., have leading coefficient 1.

Exercise 4.4.D. Consider $k[t]$-module $V = k[t]/(p(t))$. If $d = \deg p$, then show that $[1], [t], [t^2], \dots, [t^{d-1}]$ form a basis of the k-vector space V.

Exercise 4.4.E. Consider the $k[t]$-module $V = k[t]/(p_1(t)) \oplus \cdots \oplus k[t]/(p_s(t))$. Show that

$$\dim_k V = \deg p_1 + \cdots + \deg p_s.$$

What does this structure theory imply? Consider the ith component $V_i = k[t]/(p_i(t))$, which is itself a $k[t]$-module and thus corresponds to a vector space plus a linear map. Write

$$p_i(t) = t^{d_i} + a_{i,d_i-1}t^{d_i-1} + \cdots + a_{i,1}t + a_{i,0}.$$

and consider the basis $v_{i,0} = [1]$, $v_{i,1} = [t]$, $\dots$, $v_{i,d_i-1} = [t^{d_i-1}]$ of V_i. Recall that the k-linear map $T : V \to V$ is multiplication by t. So we have

$$\begin{aligned} Tv_{i,0} = v_{i,1}, \quad Tv_{i,1} &= v_{i,2}, \quad \dots, \quad Tv_{i,d_i-2} = v_{i,d_i-1}, \\ Tv_{d_i-1} = [t^{d_i}] &= -a_{i,d_i-1}[t^{d_i-1}] - \cdots - a_{i,1}[t] - a_{i,0}[1] \\ &= -a_{i,d_i-1}v_{i,d_i-1} - \cdots - a_{i,1}v_{i,1} - a_{i,0}v_{i,0}. \end{aligned}$$

With this basis $v_{i,0}, \dots, v_{i,d_i-1}$, the linear map $T : V \to V$ can be represented by the matrix

$$_B[T]^B = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_{i,0} \\ 1 & 0 & \cdots & 0 & -a_{i,1} \\ 0 & 1 & \cdots & 0 & -a_{i,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{i,d_i-1} \end{bmatrix}.$$

Definition 4.4.3. A square matrix A is said to be in **Frobenius normal form** or **rational canonical form** if it is of the block form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_s \end{bmatrix}$$

with each $d_i \times d_i$ square matrix A_i being of the form

$$A_i = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_{i,0} \\ 1 & 0 & \cdots & 0 & -a_{i,1} \\ 0 & 1 & \cdots & 0 & -a_{i,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{i,d_i-1} \end{bmatrix}.$$

Using the classification theorem and applying the above discussion to each component, we obtain the following result.

Every linear map on a finite-dimensional can be put in Frobenius normal form

Theorem 4.4.4. *Let V be a finite-dimensional vector space over k, and let $T : V \to V$ be a linear map. Then T can be put in a Frobenius normal form, by choosing a suitable basis.*

This also means that any square matrix can be put in a Frobenius normal form by conjugating it with an invertible matrix.

Exercise 4.4.F. Consider $k = \mathbb{Q}$. Find a 3×3 matrix (with entries in $k = \mathbb{Q}$!) such that

$$P \begin{bmatrix} 1 & 0 & 4 \\ 2 & -3 & 1 \\ -1 & 2 & 0 \end{bmatrix} P^{-1}$$

is in Frobenius normal form.

Every matrix is conjugate to its transpose

Exercise 4.4.G. Let A be an $n \times n$ matrix with entries in k. Show that there exists an $n \times n$ invertible matrix P (with entries in k) such that

$$PAP^{-1} = A^*.$$

(Hint: Use Exercise 4.4.C and put the matrix $tI_n - A$ into Smith normal form. The transpose of a diagonal matrix is itself.)

We now look at when k has a particularly nice property.

Definition 4.4.5. Let k be a field. We say that k is **algebraically closed** if any polynomial $p \in k[t]$ of degree at least 1 has a root, that is, for every nonzero $p(t) \in k[t]$ with $\deg p \geq 1$ there exists a $\alpha \in k$ such that $p(\alpha) = 0$.

Here is one example, although we will not prove it.

The field of complex numbers is algebraically closed

Theorem 4.4.6 (Fundamental theorem of algebra)**.** *The field $k = \mathbb{C}$ is algebraically closed.*

Exercise 4.4.H. Let $p \in k[t]$ be a nonzero polynomial with $\deg p \geq 1$. If $p(\alpha) = 0$, then we can factorize

$$p(t) = (t - \alpha)q(t)$$

for some $q \in k[t]$.

Any polynomial over an algebraically closed field can be factored into linear terms

Exercise 4.4.I. Let k be an algebraically closed field. Show that any nonzero polynomial $p \in k[t]$ can be written in the form

$$p(t) = c(t - \alpha_1) \cdots (t - \alpha_n).$$

So if k is algebraically closed, any $k[t]$-module that is a finite-dimensional k-vector space looks like

$$V \cong k[t]/((t - \alpha_1)^{d_1} \cdots (t - \alpha_k)^{d_r}) \oplus \cdots .$$

At this point, we cannot do much more than the Frobenius normal form. But we are going to use a generalized version of the Chinese remainder theorem to further decomose the module.

The Chinese remainder theorem works for relatively prime ideals

Theorem 4.4.7 (Chinese remainder theorem)**.** *Let R be a commutative ring, and let $\mathfrak{a}, \mathfrak{b} \subseteq R$ be ideals such that $\mathfrak{a} + \mathfrak{b} = R$ as ideals. There are natural projection maps $R/(\mathfrak{a} \cap \mathfrak{b}) \to R/\mathfrak{a}$ and $R/(\mathfrak{a} \cap \mathfrak{b}) \to R/\mathfrak{b}$. The induced map*

$$R/(\mathfrak{a} \cap \mathfrak{b}) \to (R/\mathfrak{a}) \times (R/\mathfrak{b})$$

is an isomorphism.

Proof. Because $\mathfrak{a} + \mathfrak{b} = 1$, there exist $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$ such that $a + b = 1$. Consider the R-linear projection map

$$\varphi : R \to (R/\mathfrak{a}) \times (R/\mathfrak{b}).$$

We claim that this is surjective. Consider any $([x], [y])$ in $(R/\mathfrak{a}) \times (R/\mathfrak{b})$. Because $a + b = 1$, we have $x - y = (x - y)(a + b)$ and hence

$$z = x + (y - x)a = y + (x - y)b.$$

This means that z is mapped to $([z], [z]) = ([x], [y])$. Hence φ is surjective.

By the first isomorphism theorem(Exercise 4.2.G), we have

$$(R/\mathfrak{a}) \times (R/\mathfrak{b}) \cong R/\ker(\varphi).$$

Here it is clear that $\ker(\varphi) = \mathfrak{a} \cap \mathfrak{b}$. □

Let's apply this theorem to $R = k[t]$.

Exercise 4.4.J. Let k be a field. For nonzero polynomials $f, g \in k[t]$ with $\gcd(f, g) = 1$, show that $(f) + (g) = (1)$ and $(f) \cap (g) = (fg)$. As a consequence, show that

$$R/(fg) \cong (R/(f)) \oplus (R/(g)).$$

Over an algebraically closed field, two polynomials are coprime if and only if they have no common roots

Exercise 4.4.K. Let k be an algebraically closed field. If $f, g \in k[t]$ are nonzero polynomials, show that $\gcd(f, g) = 1$ if and only if $f(t) = 0$ and $g(t) = 0$ do not share a common root.

Exercise 4.4.L. Let k be an algebraically closed field. If $f \in R = k[t]$ is factorized as

$$f(t) = (t - \alpha_1)^{d_1} \cdots (t - \alpha_k)^{d_k}$$

with $\alpha_1, \dots, \alpha_k$ pairwise distinct, show that we may express

$$R/(f) \cong R/((t - \alpha_1)^{d_1}) \oplus \cdots \oplus R/((t - \alpha_k)^{d_k}).$$

So if k is algebraically closed, we can write any $k[t]$-module V that is finite-dimensional over k as

$$V \cong k[t]/((t - \alpha_1)^{d_1}) \oplus \cdots k[t]/((t - \alpha_k)^{d_k}).$$

Here, the numbers α_i need not be distinct from each other. This is because V is a direct sum of several modules that look like $k[t]/(f)$.

Let us now represent the $k[t]$-module $V = k[t]/((t - \alpha_i)^{d_i})$ in a nice matrix form. Let $T : V \to V$ be multiplication by t. We pick the basis

$$v_{i,1} = [(t - \alpha_i)^{d_i - 1}], \quad v_{i,2} = [(t - \alpha_i)^{d_i - 2}], \quad \dots, \quad v_{i,d_i} = [1].$$

Then the matrix T acts on the vectors as

$$\begin{aligned} Tv_{i,1} &= [t(t - \alpha_i)^{d_i - 1}] = [(t - \alpha_i)^{d_i}] + [\alpha_i (t - \alpha_i)^{d_i - 1}] = 0 + \alpha_i v_{i,1}, \\ Tv_{i,2} &= [t(t - \alpha_i)^{d_i - 2}] = [(t - \alpha_i)^{d_i - 1}] + [\alpha_i (t - \alpha_i)^{d_i - 2}] = v_{i,1} + \alpha_i v_{i,2}, \\ &\vdots \\ Tv_{i,d_i} &= [t] = [t - \alpha_i] + [\alpha_i] = v_{i,d_i - 1} + \alpha_i v_{i,d_i}. \end{aligned}$$

It follows that if we choose the basis $v_{i,1}, \dots, v_{i,d_i}$, we get the matrix

$$T = \begin{bmatrix} \alpha_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \alpha_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_i \end{bmatrix}.$$

Definition 4.4.8. A square matrix A is said to be in **Jordan normal form** if it is of the block form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}$$

where each $d_i \times d_i$ matrix A_i being of the form

$$A_i = \begin{bmatrix} \alpha_i & 1 & 0 & \cdots & 0 \\ 0 & \alpha_i & 1 & \cdots & 0 \\ 0 & 0 & \alpha_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_i \end{bmatrix}.$$

From the discussion above, we obtain the following theorem.

Every matrix can be put in Jordan normal form

Theorem 4.4.9. *Let V be a finite-dimensional vector space over k, where k is algebraically closed. Let $T : V \to V$ be a linear map. Then T can be put in a Jordan normal form, by choosing a suitable basis.*

Exercise 4.4.M. Consider $k = \mathbb{C}$. Find a 3×3 matrix such that

$$P \begin{bmatrix} 1 & 0 & 4 \\ 2 & -3 & 1 \\ -1 & 2 & 0 \end{bmatrix} P^{-1}$$

is in Jordan normal form.

4.5 Eigenvalues and eigenvectors

Eigenvectors and eigenvalues are useful tools for analyzing how a linear map acts on a vector space.

Definition 4.5.1. Let V be a vector space over k and let $T : V \to V$ be a k-linear map. We say that $\lambda \in k$ is an **eigenvalue** of T if there exists a vector $v \in V$ with $v \neq 0$ such that $Tv = \lambda v$. In this case, we say that v is an **eigenvector** for λ.

If $v \in V$ is an eigenvector for T with eigenvalue λ, then we inductively have

$$T^n v = \lambda^n v$$

for all $n \geq 0$. Moreover, for any polynomial $p(t) \in k[t]$ with coefficients in k, we have

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

There are linear maps without eigenvectors

Exercise 4.5.A. An eigenvector need not exist. For instance, consider the matrix $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Verify that T does not have an eigenvector.

But if the base field k is algebraically closed, the theorem about Jordan normal form implies that every vector space (of positive finite dimension) has an eigenvector. It moreover tells us exactly what the eigenvectors and eigenvalues are.

Exercise 4.5.B. Let V be a finite-dimensional vector space over k, where k is algebraically closed. Show that the set of eigenvalues of $A : V \to V$ is the same as the set of diagonal entries of any Jordan normal form of A.

Before analyzing the situation for algebraically closed field, let us set up a bit of notation. Fix a finite-dimensional vector space V over a field k (not necessarily algebraically closed), and fix a linear map $T : V \to V$.

Definition 4.5.2. For each scalar $\lambda \in k$, define the **eigenspace** as

$$V_\lambda = \ker(\lambda \cdot \mathrm{id} - T) = \{v \in V : Tv = \lambda v\}.$$

Exercise 4.5.C. Check that V_λ is a subspace of V, as a k-vector space. Verify also that λ is an eigenvalue if and only if $\dim_k V_\lambda > 0$.

The interesting fact is that the spaces V_λ are all linearly independent from each other. Let me make this statement precise.

The eigenspaces of a linear map are linearly independent

Exercise 4.5.D. Fix a finite-dimensional k-vector space V and a linear map $T : V \to V$. Consider the eigenspaces $V_\lambda \subseteq V$ for each $\lambda \in k$.

(a) Show that if $\lambda_1, \dots, \lambda_n \in k$ are different scalars, and $v_i \in V_{\lambda_i}$ for all $1 \le i \le n$, then

$$v_1 + v_2 + \cdots + v_n = 0$$

implies $v_1 = v_2 = \cdots = v_n = 0$. (Hint: use Lagrange interpolation)

(b) The family of inclusion maps $V_\lambda \hookrightarrow V$ induce a k-linear map

$$\bigoplus_{\lambda \in k} V_\lambda \to V.$$

Show that this map is injective.

It will be great if the linear map $\bigoplus_{\lambda \in k} V_\lambda \to V$ is an isomorphism. However, we should not expect this to hold in generality even if k is algebraically closed. For instance,

$$T : \mathbb{C}^2 \to \mathbb{C}^2, \quad T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

has $V_\lambda = 0$ for $\lambda \neq 0$ and $V_0 = \operatorname{span}((1,0))$. So $\dim \bigoplus_{\lambda \in k} V_\lambda = 1$ while $\dim V = 2$. On the other hand, in Chapter 5, we prove a theorem that $\bigoplus_{\lambda \in k} V_\lambda \to V$ is indeed an isomorphism if T is self-adjoint. But without an assumption on the linear map T, we do not have such a result.

The solution to this problem is to look at a bigger subspace, called the generalized eigenspace.

Definition 4.5.3. Let V be a finite-dimensional vector space and $T : V \to V$ be a linear map. For each $\lambda \in k$, define the **generalized eigenspace** as

$$V_{(\lambda)} = \bigcup_{n \geq 1} \ker((\lambda \cdot \mathrm{id} - T)^n) = \{v \in V : (\lambda - T)^n v = 0 \text{ for some } n \geq 1\}.$$

Exercise 4.5.E. Check that $V_{(\lambda)}$ is a linear subspace of V. Also verify that λ is an eigenvalue if and only if $\dim V_\lambda > 0$ if and only if $\dim V_{(\lambda)} > 0$.

Although we have enlarged each of the spaces V_λ to $V_{(\lambda)}$, generalized eigenspaces still satisfy the linear independence property.

Generalized eigenspaces of a linear map are linearly independent

Exercise 4.5.F. Let V be a finite-dimensional space, and $T : V \to V$ be a linear map.

(a) Show that if $0 \neq v \in V_{(\lambda)}$ and $\lambda \neq \kappa \in k$, then $(\kappa - T)v \in V_{(\lambda)}$ and $(\kappa - T)v \neq 0$.

(b) Show that the linear map

$$\bigoplus_{\lambda \in k} V_{(\lambda)} \to V$$

induced by the inclusions $V_{(\lambda)} \hookrightarrow V$ is injective. (Hint: suppose it is not injective, and look at nonzero vectors $v_1, \dots, v_n$ that add up to 0 and contained in different generalized eigenspaces, such that n is smallest as possible. Derive a contradiction by applying a suitable $(T - \lambda)^n$ to it.)

Exercise 4.5.G. Find an example of a vector space V and a linear map $T : V \to V$ such that the linear map $\bigoplus_{\lambda \in k} V_{(\lambda)} \to V$ is not an isomorphism.

Let us now assume that k is an algebraically closed closed field. In this case, we will show that the map $\bigoplus_{\lambda \in k} V_{(\lambda)} \to V$ is an isomorphism.

Generalized eigenspaces form a decomposition of the original vector space

Theorem 4.5.4. *Let V be a finite-dimensional vector space over k, where k is algebraically closed. Consider a linear map $T : V \to V$ and the generalized eigenspaces $V_{(\lambda)}$ with repsect to T. Then the linear map*

$$\bigoplus_{\lambda \in k} V_{(\lambda)} \to V$$

induced by the inclusion maps $V_{(\lambda)} \hookrightarrow V$ is an isomorphism.

Proof. Consider the vector space V with linear map T as a $k[t]$-module, where multiplication by t acts as applying T. By the discussion around the Jordan normal form, we see that V has a decomposition

$$V \cong \bigoplus_{i=1}^{n} k[t]/((t-\alpha_i)^{d_i})$$

as $k[t]$-modules, where $\alpha_i \in k$ and $d_i \geq 1$. For each i, we may consider $k[t]/((t-\alpha_i)^{d_i}) \subseteq V$ as a subspace. Then for any $v \in k[t]/((t-\alpha_i)^{d_i})$, applying $(T-\alpha_i)^{d_i}$ to v gives

$$(T-\alpha_i)^{d_i} v = (t-\alpha_i)^{d_i} v = 0 \in k[t]/((t-\alpha_i)^{d_i}).$$

This implies that $k[t]/((t-\alpha_i)^{d_i}) \subseteq V_{(\alpha_i)}$. Because the subspaces $k[t]/((t-\alpha_i)^{d_i})$ generate V, it follows that the subspaces $V_{(\lambda)}$ also generate V. Combining with Exercise 4.5.F, we obtain the desired result. □

Exercise 4.5.H. In the above proof, show that actually we can identify $V_{(\lambda)}$ directly as

$$V_{(\lambda)} = \bigoplus_{\alpha_i=\lambda} k[t]/((t-\alpha_i)^{d_i}).$$

Hence it immediately follows that $\bigoplus_\lambda V_{(\lambda)} \cong V$.

There is an application of the above discussion, called the Jordan–Chevalley decomposition. The theorem is useful in representation theory, but we use it only as a demonstration of the theory we developed so far. To state the theorem, we make some definitions.

Definition 4.5.5. Let k be an algebraically closed field, and let V be a finite-dimensional vector space over k. A linear map $T : V \to V$ is said to be **diagonalizable** or **semisimple** if the linear map

$$\bigoplus_{\lambda \in k} V_\lambda \to V$$

induced by the inclusions is an isomorphism of vector spaces. (V_λ are the ordinary eigenspaces for T.)

We have seen an example of linear map that is not semisimple, namely $T = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. On the other hand, most matrices (if the entries are chosen randomly) are semisimple.

Exercise 4.5.I. Show that T is semisimple if and only if, the matrix in Jordan normal form obtained by conjugating T is a diagonal matrix.

Most linear maps are semisimple

Exercise 4.5.J. Let T be an $n \times n$ matrix with entries in k, where k is algebraically closed. If the polynomial $\det(tI-T)$ in the variable t has n distinct roots in k, show that T is semisimple.

Definition 4.5.6. Let k be an algebraically closed field, and let V be a finite-dimensional vector space over k. A linear map $T : V \to V$ is said to be **nilpotent** if $T^n = 0$ for some $n \geq 1$.

Exercise 4.5.K. Show that $T : V \to V$ is nilpotent if and only if $V_{(0)} = V$.

We now state the main theorem.

Every linear map decomposes into a sum of a semisimple and a nilpotent that commute

Theorem 4.5.7 (Jordan–Chevalley decomposition)**.** *Let V be a finite-dimensional vector space over an algebraically closed field k. Then for every linear map $T : V \to V$ there is a decomposition*

$$T = T_{ss} + T_n$$

into a semisimple map T_{ss} and a nilpotent map T_n such that $T_{ss}T_n = T_nT_{ss}$. Moreover, such a decomposition is unique.

We prove this theorem in a series of exercises.

Exercise 4.5.L. Consider the decomposition $V \cong \bigoplus V_{(\lambda)}^{(T)}$ with respect to T. Define T_{ss} by setting

$$T_{ss}v = \lambda v$$

if $v \in V_{(\lambda)}^{(T)}$, and then extending linearly.

(a) Show that the eigenspaces with respect to T_{ss} are $V_{\lambda}^{(T_{ss})} = V_{(\lambda)}^{(T)}$. Deduce that T_{ss} is indeed semisimple.

(b) Show that $T_{ss}T = TT_{ss}$.

(c) Define $T_n = T - T_{ss}$. Show that for every $v \in V_{(\lambda)}^{(T)}$, there exists a $k \geq 1$ such that $T_n^k v = 0$. Conclude that T_n is nilpotent.

(d) Show that $T_{ss}T_n = T_nT_{ss}$.

Exercise 4.5.M. In the other direction, suppose that $T = T_{ss} + T_n$, where T_{ss} is semisimple, T_n is nilpotent, and they commute.

(a) Show that T and T_{ss} commute, i.e., $TT_{ss} = T_{ss}T$.

(b) Show that if $v \in V_{\lambda}^{(T_{ss})}$, then $v \in V_{(\lambda)}^{(T)}$. That is, $V_{\lambda}^{(T_{ss})} \subseteq V_{(\lambda)}^{(T)}$.

(c) From the decompositions $V \cong \bigoplus_{\lambda} V_{(\lambda)}^{(T)}$ and $V \cong \bigoplus_{\lambda} V_{\lambda}^{(T_{ss})}$, show that $V_{\lambda}^{(T_{ss})} = V_{(\lambda)}^{(T)}$.

(d) Conclude that T_{ss} has to be the linear map constructed in the previous exercise.

There is a similar multiplicative version of the Jordan–Chevalley decomposition.

Definition 4.5.8. Let k be an algebraically closed field, and let V be a finite-dimensional vector space over k. A linear map $T : V \to V$ is said to be **unipotent** if $T - \mathrm{id}$ is nilpotent.

Every invertible linear map decomposes into a product of a semisimple and a unipotent that commute

Exercise 4.5.N (multiplicative Jordan–Chevalley decomposition)**.** Let V be a finite-dimensional vector space over an algebraically closed field k. Prove that for every linear map $T : V \to V$ there is a decomposition

$$T = T_{ss}T_u$$

into a semisimple map T_{ss} and a unipotent map T_u such that $T_{ss}T_u = T_uT_{ss}$. Moreover, show that such a decomposition is unique.

Chapter 5

Linear algebra over $\mathbb{R}$ and $\mathbb{C}$

When the base field is $k = \mathbb{R}$ or $k = \mathbb{C}$, we can do analysis in the vector spaces. Well, not really hard analysis like differentiation or integration, but stuff like comparing sizes of numbers or taking the supremum. When we mix linear algebra with a bit of analysis, we get interesting consequences.

In this chapter, we will only deal with finite-dimensional vector spaces. There is a theory of infinite-dimensional topological vector spaces, and many of the things we shall discuss will be generalizable to that case. But we will not worry about such spaces. If you are interested in the infinite-dimensional case, go and study functional analysis. You will deal with spaces like the vector space of continuous functions $[0,1] \to \mathbb{R}$.

5.1 A bit of analysis

Given two real numbers, we can compare them, and so there is a notion of smallness.

Definition 5.1.1. Let $\{a_n\}_{n\geq 0}$ be an infinite sequence of real numbers. We say that its **limit** is L if for every $\epsilon > 0$, there exists a sufficiently large $N_\epsilon > 0$ such that

$$|a_n - L| < \epsilon$$

for all $n \geq N_\epsilon$. In this case, we write

$$\lim_{n\to\infty} a_n = L.$$

Of course, some sequences do not have limits. For instance, $a_n = (-1)^n$ does not have a limit. But if a limit exists, it is unique.

Limits commute with termwise sums

Exercise 5.1.A. Show that if $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$, then $\lim_{n\to\infty}(a_n + b_n) = a + b$.

Limits commute with termwise products

Exercise 5.1.B. Show that if $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$, then $\lim_{n\to\infty}(a_n b_n) = ab$.

It is a sort of definition that the real numbers are complete.

Definition 5.1.2. A **Cauchy sequence** is a sequence of real numbers $\{a_n\}_{n\geq 0}$ such that for any $\epsilon > 0$, there exists a sufficiently large $N_\epsilon > 0$ such that

$$|a_n - a_m| < \epsilon$$

for all $n, m \geq N_\epsilon$.

Cauchy sequences converge

Theorem 5.1.3 (Completeness of $\mathbb{R}$)**.** *Every Cauchy sequence in $\mathbb{R}$ has a limit.*

Bounded monotone sequences have limits

Exercise 5.1.C. Let $a_0, a_1, \ldots$ be a bounded nondecreasing sequence of real numbers. In other words, $a_i \leq a_{i+1}$ for all $i \geq 0$, and there exists a universal constant C such that $a_i \leq C$ for all $i \geq 0$. Show that the limit $\lim_{n\to\infty} a_n$ exists.

Bounded sets have supremums

Exercise 5.1.D. Let $S \subseteq \mathbb{R}$ be a nonempty subset such that there exists a constant C such that $x < C$ for all $x \in S$. Show that there exists a real number M such that

(i) $x \leq M$ for all $x \in S$,

(ii) for any $M' < M$, there exists a $x \in S$ such that $x > M'$.

Such M is called the **supremum** of S and we write $M = \sup S$.

We have thus defined the limit of a sequence of real numbers. Let us now define a limit of vectors.

Definition 5.1.4. Let $\{v_k\}_{k\geq 0}$ be a sequence of vectors in $\mathbb{R}^n$. Write $v_k = (v_{k,1}, \ldots, v_{k,n}) \in \mathbb{R}^n$. We say that the **limit** of the sequence is $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ if

$$x_i = \lim_{k\to\infty} v_{k,i}$$

and all the limits exist. In this case, we write

$$x = \lim_{k\to\infty} v_k.$$

Exercise 5.1.E. Consider $\{z_k\}_{k\geq 0}$ be a sequence of complex numbers. Show that the following are equivalent:

(1) When $\mathbb{C}$ is considered as a real vector space, $\lim_{k\to\infty} z_k = z$.

(2) For every $\epsilon > 0$, there exists a suffuciently large $N_\epsilon > 0$ such that $|z_n - z| < \epsilon$ for all $n > N_\epsilon$.

But what about in general vector spaces that are not canonically isomorphic to $\mathbb{R}^n$?

Limits commute with linear maps

Proposition 5.1.5. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be an $\mathbb{R}$-linear map. If $\{v_k\}_{k\geq 0}$ is a sequence with $x = \lim_{k\to\infty} v_k$, then*

$$Tx = \lim_{k\to\infty} Tv_k.$$

Proof. Write T as a matrix $T = (t_{ij})$, and let $v_k = (v_{k,1}, \dots, v_{k,n})$. Then the ith component of Tv_k is given by

$$(Tv_k)_i = \sum_j t_{ij} v_{k,j}.$$

Because $\lim_{k\to\infty} v_{k,j} = x_j$, we have

$$\lim_{k\to\infty} (Tv_k)_i = \sum_j t_{ij} \lim_{k\to\infty} v_{k,j} = \sum_j t_{ij} x_j = (Tx)_i$$

by the commutation of limits with addition and multiplication. This shows that $Tx = \lim_{k\to\infty} Tv_k$. □

Definition 5.1.6. Let V be a finite-dimensional vector space over $\mathbb{R}$. For a sequence $\{v_k\}_{k\geq 0}$ of vectors in V, we say that

$$v = \lim_{k\to\infty} v_k$$

if $Tv = \lim_{k\to\infty} Tv_k$ for a choice of isomorphism $T : V \to \mathbb{R}^n$.

What if we choose another isomorphism $S : V \to \mathbb{R}^n$? Because $ST^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map, applying ST^{-1} shows that $Tv = \lim_{k\to\infty} Tv_k$ implies $Sv = \lim_{k\to\infty} Sv_k$. Also, applying TS^{-1} implies the other direction. Therefore $Tv = \lim_{k\to\infty} Tv_k$ if and only if $Sv = \lim_{k\to\infty} Sv_k$. This means that the limit does not depend on the choice of isomorphism $T : V \to \mathbb{R}^n$.

Definition 5.1.7. Let V and W be finite-dimensional vector spaces over $\mathbb{R}$. We say that a function $f : V \to W$ (not necessarily a linear map) is **continuous** if $v = \lim_{k\to\infty} v_k$ implies $f(v) = \lim_{k\to\infty} f(v_k)$.

What we proved in Proposition 5.1.5 is that a linear map is always continuous.

The composition of continuous maps is continuous

Exercise 5.1.F. Let V, W, U be finite-dimensional vector spaces over $\mathbb{R}$. Let $f : V \to W$ be a continuous map, and let $g : W \to U$ be a continuous map. Show that $g \circ f : V \to U$ is continuous as well.

The square root is continuous

Exercise 5.1.G. Show that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sqrt{|x|}$ is continuous.

We end with a useful theorem.

A continuous function on a closed bounded set has a maximum

Theorem 5.1.8 (Heine–Borel)**.** *Let V be a finite-dimensional vector space over $\mathbb{R}$, and let $S \subseteq V$ be a subset satisfying the following two properties:*

(a) The set S is bounded, i.e., there exists an isomorphism $T : V \to \mathbb{R}^n$ and a large real number L such that $S \subseteq T^{-1}([-L, L]^n)$.

(b) The set S is closed, i.e., for every sequence $\{v_k\}_{k\geq 0}$ with a limit, if $v_k \in S$ then $\lim_{k\to\infty} v_k \in S$.

Then for any continuous function $V \to \mathbb{R}$, the functions has a maximal value on S. That is, there exists a $x \in S$ such that $f(x) \geq f(s)$ for all $s \in S$.

5.2 Inner products

We all know about inner products in $\mathbb{R}^3$, or maybe $\mathbb{R}^n$. Given two vectors $\vec{a} = (a_1, \dots, a_n)$ and $\vec{b} = (b_1, \dots, b_n)$, we define their inner product as

$$\langle \vec{a}, \vec{b} \rangle = a_1 b_1 + \cdots + a_n b_n.$$

But given a general finite-dimensional vector space V over $\mathbb{R}$, what is the inner product? One cannot define an inner product canonically, but we can try to think about what it means to be an inner product.

Definition 5.2.1. Let V be a finite-dimensional vector space over a field k. A **symmetric bilinear form** is an element of $\operatorname{Sym}^2 V^* \cong (\operatorname{Sym}^2 V)^*$. In other words, it is a bilinear map

$$\langle -, - \rangle : V \times V \to k$$

such that $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$.

This can be defined over a general field k. But for $k = \mathbb{R}$, we can further impose a positivity condition.

Definition 5.2.2. Let V be a finite-dimensional vector space over $\mathbb{R}$, and let $\langle -, - \rangle$ be a symmetric bilinear form. We say that $\langle -, - \rangle$ is **positive definite** or an **inner product** if

$$\langle v, v \rangle > 0$$

for all $v \neq 0$.

Example 5.2.3. The **standard inner product** on $\mathbb{R}^n$ defined by

$$\langle v, w \rangle = v_1 w_1 + \cdots + v_n w_n$$

is indeed an inner product on $\mathbb{R}^n$. This is because $v_i^2 \geq 0$ and equality holds if and only if $v_i = 0$.

This is for real vector spaces V. For complex vector spaces, we could regard it as a real vector space and look at inner products. But then we forget the complex structure of V, and we do not want to do this. So we make a somewhat peculiar definition.

Definition 5.2.4. Let V be a finite-dimensional vector space over $\mathbb{C}$. A **Hermitian sesquilinear form** is a $\mathbb{R}$-bilinear map

$$\langle -, - \rangle : V \times V \to \mathbb{C}$$

such that

$$\langle w, v \rangle = \overline{\langle v, w \rangle}, \quad \langle cv, w \rangle = \bar{c}\langle v, w \rangle, \quad \langle v, cw \rangle = c\langle v, w \rangle$$

for all $c \in \mathbb{C}$ and $v, w \in V$.

The reason we are putting in complex conjugation is because this is when it makes sense to put a positive-definite condition. If $\langle -, - \rangle$ was bilinear, we would always have $\langle iv, iv \rangle = i^2 \langle v, v \rangle = -\langle v, v \rangle$.

Definition 5.2.5. Let V be a finite-vector space over $\mathbb{C}$, and let $\langle -, - \rangle$ be a sesquilinear form. We say that $\langle -, - \rangle$ is an **inner product** (sometimes a **Hermitian inner product** to emphasize that we are working over $\mathbb{C}$) if

$$\langle v, v \rangle \geq 0$$

for all $v \neq 0$. (The condition $\langle v, v \rangle \geq 0$ includes the fact that $\langle v, v \rangle$ is a real number.)

Example 5.2.6. The **standard inner product** on $\mathbb{C}^n$ defined by

$$\langle v, w \rangle = \bar{v}_1 w_1 + \cdots + \bar{v}_n w_n$$

is indeed an inner product on $\mathbb{C}^n$ as a complex vector space. This is because $\bar{v}_i v_i = |v_i|^2 \geq 0$ with equality holding if and only if $v_i = 0$.

Technically speaking, the definitions of an inner product on a $\mathbb{R}$-vector space and of an inner product over a $\mathbb{C}$-vector space are different. We are giving them the same name because they express the same idea and share useful properties.

Inner products form a cone

Exercise 5.2.A. Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$. Show that the sum of any two inner products on V is again an inner product of V. Here, we add inner products as elements of $\operatorname{Sym}^2 V^*$, so that the sum of $\langle -, - \rangle_a$ and $\langle -, - \rangle_b$ is given by $\langle v, w \rangle_{a+b} = \langle v, w \rangle_a + \langle v, w \rangle_b$.

In the rest of the section, we are going to fix a finite-dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, and an inner product $\langle -, - \rangle$ on V. This is called a finite-dimensional Hilbert space.

Definition 5.2.7. Fix $k = \mathbb{R}$ or $k = \mathbb{C}$. A **finite-dimensional Hilbert space** is a finite-dimensional vector space V over k along with a choice of an inner product $\langle -, - \rangle$.

In general, Hilbert spaces can be infinite-dimensinal—this is why we are adding the cumbersome adjective "finite-dimensional". The true story is that a Hilbert space needs to satisfy more properties suitable for analysis, but they are automatically satisfied for finite-dimensional vector spaces.

Let $k = \mathbb{R}$ or $k = \mathbb{C}$. Given a finite-dimensional Hilbert space V over k (where the inner product is implicit) and a vector $v \in V$, we have a k-linear functional

$$\langle v, - \rangle : V \to k; \quad w \mapsto \langle v, w \rangle.$$

Note that the functional $\langle -, v \rangle$ might not be $\mathbb{C}$-linear if $k = \mathbb{C}$, because we have $\langle cw, v \rangle = \bar{c} \langle w, v \rangle$. Functionals satisfying such identities are called **conjugate-linear** instead of linear.

Every linear functional can be uniquely represented by a vector

Theorem 5.2.8 (Riesz representation theorem)**.** *Let $k = \mathbb{R}$ or $k = \mathbb{C}$, and let V be a finite-dimensional Hilbert space over k. The map*

$$\Phi : V \to V^*; \quad v \mapsto \langle v, - \rangle$$

is a bijection.

Note that Φ is not k-linear in general; rather, it is conjugate-linear.

Proof. Even if Φ is conjugate-linear over k, it is $\mathbb{R}$-linear because conjugation does nothing on $\mathbb{R}$. That is, when we regard both V and V^* as $\mathbb{R}$-vector spaces, Φ is a linear map. Now let us compare dimension over $\mathbb{R}$ on both sides. If $k = \mathbb{R}$, then $\dim_{\mathbb{R}} V = \dim_{\mathbb{R}} V^*$. If $k = \mathbb{C}$, we only have $\dim_{\mathbb{C}} V = \dim_{\mathbb{C}} V^*$, but when take a $\mathbb{C}$-vector space of dimension n and consider it as a $\mathbb{R}$-vector space, we have $\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V$. Likewise we have $\dim_{\mathbb{R}} V^* = 2 \dim_{\mathbb{C}} V^*$, and this implies $\dim_{\mathbb{R}} V = \dim_{\mathbb{R}} V^*$. Therefore

$$\dim_{\mathbb{R}} V = \dim_{\mathbb{R}} V^*$$

in both cases.

To show that the linear map $\Phi : V \to V^*$ is an isomorphism, it now suffices to show that $\ker \Phi = 0$. Suppose $\langle v, - \rangle = 0 = \langle 0, - \rangle$. Evaluating this functional on v gives $\langle v, v \rangle = 0$, and this implies $v = 0$ because $\langle -, - \rangle$ is an inner product. This shows that $\ker \Phi = 0$, which immediately implies that Φ is an isomorphism. □

The original Riesz representation theorem holds for all Hilbert spaces, not only finite-dimensional ones.

Definition 5.2.9. Let V be a finite-dimensional Hilbert space over k, and let $W \subseteq V$ be a subspace over k. We define the **orthogonal complement** of W as

$$W^{\perp} = \{ v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W \}.$$

It is not hard to verify that $W^{\perp}$ again is a k-vector space.

Exercise 5.2.B. In the above situation, prove the following statements.

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

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

(c) For any $v \in V$ there exist unique $w \in W$ and $w' \in W^{\perp}$ such that $v = w + w'$.

In a Hilbert space, we can assign to each vector a length. Recall how Euclidean length was defined in $\mathbb{R}^n$. The length of $(a_1, \ldots, a_n)$ is given by $\sqrt{a_1^2 + \cdots + a_n^2}$. So maybe a good definition is to define length so that its square is equal to the inner product with itself. This make sense, because one of our conditions stated that $\langle v, v \rangle \geq 0$ for all v.

Definition 5.2.10. Let V be a finite-dimensional Hilbert space over $k = \mathbb{R}$ or $k = \mathbb{C}$. Then we define a **norm** map

$$\|-\| : V \to \mathbb{R}_{\geq 0}; \quad \|v\| = \sqrt{\langle v, v \rangle}.$$

The norm satisfies many of the properties we are used to from Euclidean geometry.

Cauchy–Schwartz for Hilbert spaces

Exercise 5.2.C (Cauchy–Schwartz)**.** Let V be a finite-dimensional Hilbert space, and let $v, w \in V$. Show that

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

(Hint: the quadratic polynomial $P(\lambda) = \|v - \lambda w\|^2$ is nonnegative for all λ)

Norm satisfies the triangle inequality

Exercise 5.2.D. Let V be a finite-dimensional Hilbert space, and let $v, w \in V$. Show that

$$\|v + w\| \leq \|v\| + \|w\|.$$

Norm is a continuous function

Exercise 5.2.E. Show that the norm map on a finite-dimensional Hilbert space, as a map $V \to \mathbb{R}$, is continuous.

Let V be a finite-dimensional Hilbert space over k, and let $W \subseteq V$ be a subspace. We have shown in Exercise 5.2.B that every $v \in V$ has a unique decomposition $v = \mathrm{proj}_W(v) + (v - \mathrm{proj}_W(v))$ such that $v - \mathrm{proj}_W(v) \in W^{\perp}$. We call $\mathrm{proj}_W(v)$ the **orthogonal projection** of v onto W.

Exercise 5.2.F. In the above setting, prove the following statements.

(a) The projection map $\mathrm{proj}_W : V \to W$ is k-linear.

(b) The vector proj_W is characterized by

$$\|v - \mathrm{proj}_W(v)\| = \min_{w \in W} \|v - w\|.$$

It is sometimes useful to choose and work with a basis. Because a Hilbert space has an additional structure of an inner product, we will impose a condition on the inner products between the basis vectors.

Definition 5.2.11. Let V be a finite-dimensional vector space over $k = \mathbb{R}$ or $k = \mathbb{C}$. An **orthonormal basis** is a basis $v_1, v_2, \dots, v_n$ of V such that

$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Having an orthonormal basis is a useful thing, because once we write down the vector as a linear combination of the basis, we can easily compute the inner product. Let $v_1, \dots, v_n$ be an orthonormal basis. Then

$$\langle a_1 v_1 + \cdots a_n v_n, b_1 v_1 + \cdots + b_n v_n \rangle = \sum_{i,j=1}^{n} \bar{a}_i b_j \langle v_i, v_j \rangle = \bar{a}_1 b_1 + \cdots + \bar{a}_n b_n.$$

So the Hilbert space V looks like the standard Hilbert space $\mathbb{R}^n$ or $\mathbb{C}^n$ when we choose an orthonormal basis.

The next question to ask is, does an orthonormal basis always exist? In fact, we can give a better answer. Given any basis, there is a procedure for producing an orthonormal basis. This is called the **Gram–Schmidt process**.

We start with an arbitrary basis $v_1, \dots, v_n$ of V. Because v_i might not be of length 1, we first normalize this as

$$w_1 = \frac{v_1}{\|v_1\|}$$

so that $\|w_1\| = 1$. Then, we want to make v_2 into a unit length vector orthogonal to v_1. This can be done by defining

$$w_2 = \frac{v_2 - \langle w_1, v_2 \rangle w_1}{\|v_2 - \langle w_1, v_2 \rangle w_1\|}.$$

Then clearly w_2 has length 1, and its inner product against w_1 is

$$\langle w_1, w_2 \rangle = \frac{1}{\|v_2 - \langle w_1, v_2 \rangle w_1\|} (\langle w_1, v_2 \rangle - \langle w_1, v_2 \rangle \langle w_1, w_1 \rangle) = 0.$$

Moreover, we have the property that $\operatorname{span}(v_1, v_2) = \operatorname{span}(w_1, w_2)$. We can similarly inductively define the vectors

$$w_k = \frac{v_k - \langle w_1, v_k \rangle w_1 - \langle w_2, v_k \rangle w_2 - \cdots - \langle w_{k-1}, v_k \rangle w_{k-1}}{\|v_k - \langle w_1, v_k \rangle w_1 - \langle w_2, v_k \rangle w_2 - \cdots - \langle w_{k-1}, v_k \rangle w_{k-1}\|}.$$

At the end, we get an orthonormal basis $w_1, \dots, w_n$ of V with the additional property $\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(w_1, \dots, w_k)$.

Exercise 5.2.G. Consider $\mathbb{R}^3$ equipped with the inner product given by

$$\langle (a_1, a_2, a_3), (b_1, b_2, b_3) \rangle = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Find an orthonormal basis for this Hilbert space.

Exercise 5.2.H. Let V be a finite-dimensional Hilbert space, and let $v_1, \ldots, v_n$ be a linear basis for V. Show that there uniquely exist one-dimensional subspaces $W_1, \ldots, W_n \subseteq V$ satisfying

(i) the subspaces W_i are orthogonal to each other,

(ii) $\operatorname{span}(v_1, \ldots, v_k) = W_1 \oplus W_2 \oplus \cdots \oplus W_k$ as subspaces of V, for all k (see Exercise 2.5.D).

An orthonormal basis of a subspace can be exteded to an orthonormal basis

Exercise 5.2.I. Let V be a finite-dimensional Hilbert space, and $W \subseteq V$ a linear subspace. Show that any orthonormal basis of W can be exteded to an orthonormal basis of V. More precisely, if $v_1, \ldots, v_k$ is a orthonormal basis of W, then show that there exist vectors $v_{k+1}, \ldots, v_n$ such that $w_1, \ldots, v_n$ is an orthonormal basis of V.

Lastly, we define a notion of an isomorphism of Hilbert spaces. This should be a map that preserves all inner products while being an isomorphism of vector spaces.

Definition 5.2.12. Let V, W be Hilbert spaces. A linear map $T : V \to W$ is called an **isometry** if it is an isomorphism of vector spaces, and

$$\langle Tv_1, Tv_2 \rangle_W = \langle v_1, v_2 \rangle_V$$

for all $v_1, v_2 \in V$.

Exercise 5.2.J. Let V be a finite-dimensional Hilbert space. Show that the linear map $T : k^{\oplus n} \to V$ induced by n vectors $v_1, \ldots, v_n \in V$ is an isometry if and only if $v_1, \ldots, v_n$ form an orthonormal basis for V.

Note that the condition $\langle Tv_1, Tv_2 \rangle_W = \langle v_1, v_2 \rangle_V$ is not enough to conclude that T is an isometry.

Exercise 5.2.K. Let V, W be Hilbert spaces and let $T : V \to W$ be a linear map satisfying $\langle Tv_1, Tv_2 \rangle = \langle v_1, v_2 \rangle$ for all $v_1, v_2 \in V$. Show that T is injective. Find an example when T is not surjective. In general, such T is called an **isometric embedding**.

Inner products are determined by the norm

Exercise 5.2.L. Show that if $T : V \to W$ is an isometric embedding if and only if

$$\|Tv\|_W = \|v\|_V$$

for all $v \in V$. (Hint: express the real part $\Re\langle v, w \rangle$ in terms of $\|v\|$, $\|w\|$, and $\|v + w\|$.)

When the isometry is from V to itself, we use several names.

Definition 5.2.13. Let V be a finite-dimensional Hilbert space over $k = \mathbb{R}$. An isometry $T : V \to V$ is called **orthogonal** if $k = \mathbb{R}$ and **unitary** if $k = \mathbb{C}$.

Exercise 5.2.M. Show that a $n \times n$ real matrix, considered as a linear map $\mathbb{R}^n \to \mathbb{R}^n$, is an isometry if and only if its columns form an orthonormal basis for $\mathbb{R}^n$.

Every matrix has a QR decomposition

Exercise 5.2.N (QR decomposition)**.** Let A be a $n \times n$ matrix with real entries. Show that there exists a decomposition

$$A = QR$$

where Q is an orthogonal matrix (considered as $\mathbb{R}^n \to \mathbb{R}^n$) and R is an upper triangular matrix (if $R = (r_{ij})$ then $r_{ij} = 0$ for $i > j$).

Similar statements to the previous two exercises hold also for $k = \mathbb{C}$.

5.3 Operators on Hilbert spaces

We are now going to discuss linear maps between Hilbert spaces. In most contexts, Hilbert spaces are spaces of functions. So a linear map between Hilbert spaces is usually called a **linear operator** instead of a linear map. The map $f \mapsto \frac{df}{dx}$ being called a differential operator explains the terminology.

Given an operator between Hilbert spaces, we want to define how "large" that operator is. More precisely, this measures how large Tv can be compared to v.

Definition 5.3.1. Let V, W be finite-dimensional vector spaces over k, where $k = \mathbb{R}$ or $k = \mathbb{C}$. Assume that $\dim V \geq 1$. For a linear operator $T : V \to W$, we define its **norm** as

$$\|T\| = \max_{\|v\|_V \leq 1} \|Tv\|_W = \max_{v \neq 0} \frac{\|Tv\|_W}{\|v\|_V}.$$

(Here, $\|Tv\|_W$ is the norm of Tv with respect to the inner product in W, and $\|v\|_V$ is the norm of v with respect to the inner product in V.)

But before we start using this definition, we need to ask if there really is a maximum. Why cannot $\|Tv\|_W$ be indefinitely large with $\|v\|_V \leq 1$? What if $\|Tv\|_W$ can take every value smaller than 1 but never 1, for instance? This is where we need the Heine–Borel theorem (Theorem 5.1.8). The unit ball

$$B = \{v \in V : \|v\|_V \leq 1\}$$

is clearly bounded. To see that B is closed, we use the fact that the norm is continuous. (See Exercise 5.2.E.) If $b_1, b_2, \ldots \in B$ and $\lim_{k\to\infty} b_k = b$ then $\|b\|_V = \lim_{k\to\infty} \|b_k\|_V \leq 1$ because $|b_k|_V \leq 1$ for all k. Therefore B is bounded and closed, and Heine–Borel implies that our definition makes sense. This means that $\|Tv\|_W \leq \|T\|\|v\|_V$ for all $v \in V$, and there exists a $v \neq 0$ such that $\|Tv\|_W = \|T\|\|v\|_V$.

Exercise 5.3.A. Consider the Hilbert space $\mathbb{R}^2$, with the standard inner product. Compute the norm of the linear operator

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Exercise 5.3.B. Let V, W, U be finite-dimensional Hilbert spaces, and let $T : V \to W$ and $S : W \to U$ be linear operators. Show that

$$\|ST\| \leq \|S\|\|T\|,$$

and find examples of S and T such that $\|ST\| < \|S\|\|T\|$.

Exercise 5.3.C. Let V, W be finite-dimensional Hilbert spaces, and let $T, S : V \to W$ be linear operators. Show that

$$\|T + S\| \leq \|T\| + \|S\|.$$

Let V, W be finite-dimensional Hilbert spaces, and fix $T : W \to V$ a linear operator. For each vector $v \in V$, there is a linear functional

$$\langle v, T(-)\rangle_V : W \to k; \quad w \mapsto \langle v, Tw\rangle_V.$$

By the Riesz representation theorem (Theorem 5.2.8), this linear functional is uniquely representable.

Definition 5.3.2. Let V, W be finite-dimensional Hilbert spaces, and let $T : W \to V$ be a linear operator. The **adjoint operator** $T^\dagger$ of T is defined to satisfy

$$\langle v, Tw\rangle_V = \langle T^\dagger v, w\rangle_W$$

for all $v \in V$ and $w \in W$.

Exercise 5.3.D. Show that the adjoint operator $T^\dagger : V \to W$ is linear. (The above definition is at the level of sets.)

Exercise 5.3.E. In the above setting, show that $(T^\dagger)^\dagger = T$.

The adjoint of the adjoint is itself

Exercise 5.3.F. Let V, W, U be finite-dimensional Hilbert spaces, and let $T : V \to W$ and $S : W \to U$ be linear operators. Show that

$$(ST)^\dagger = T^\dagger S^\dagger.$$

Take $V = \mathbb{C}^n$ and $W = \mathbb{C}^m$ (with standard inner products), and write $T : W \to V$ as a $n \times m$ matrix $T = (t_{ij})$ with complex entries. Its adjoint will be a linear map $T^\dagger : V \to W$, which will be a $m \times n$ matrix. What matrix is this going to be? The equation $\langle v, Tw\rangle_V = \langle T^\dagger v, w\rangle_W$ can be written as

$$\sum_{i=1}^{n} \bar{v}_i \sum_{j=1}^{m} t_{ij} w_j = \sum_{j=1}^{m} \overline{(T^\dagger v)}_j w_j,$$

and this shows that $(T^\dagger v)_j = \sum_{i=1}^n \bar{t}_{ij} v_i$. That is, $T^\dagger$ is the **conjugate transpose** of T defined so that the ith column jth row entry of $T^\dagger$ is the complex conjugate of the jth column ith row entry of T. We can also write $T^\dagger = \bar{T}^*$ if we wish, where $*$ means transpose and $\_$ means taking complex conjugate of all the entries.

Exercise 5.3.G. Let V, W be finite-dimensional Hilbert spaces over k. We can define an inner product structure on $V \oplus W$ by

$$\langle (v_1, w_1), (v_2, w_2) \rangle = \langle v_1, v_2 \rangle + \langle w_1, w_2 \rangle,$$

and $V \oplus W$ is a finite-dimensional Hilbert space with this inner product. Let $T : W \to V$ be a linear operator, and define subspaces $X, Y \subseteq V \oplus W$ as

$$X = \{(Tw, w) : w \in W\}, \quad Y = \{(v, -T^\dagger v) : v \in V\}.$$

Show that $X^\perp = Y$ in $V \oplus W$.

Exercise 5.3.H. Let V be a finite-dimensional Hilbert space and let $T : V \to V$ be a linear operator. Show that T is unitary (or orthogonal) if and only if $T^{-1} = T^\dagger$.

So a square matrix Q with real entries is orthogonal if and only if $QQ^* = Q^*Q = I$ and a sqaure matrix U with complex entries is unitary if and only if $UU^\dagger = U^\dagger U = I$.

We now consider the case when $V = W$. In this case, we can compare the two linear operators $T, T^\dagger : V \to V$.

Definition 5.3.3. Let V be a finite-dimensional Hilbert space over k and let $T : V \to V$ be a linear operator. We say that T is **self-adjoint** if $T = T^\dagger$. When $k = \mathbb{R}$, a synonym for self-adjoint is **symmetric** and when $k = \mathbb{C}$, a self-adjoint operator is also called a **Hermitian** operator.

The condition $T = T^\dagger$ just means

$$\langle Tv, w \rangle = \langle v, Tw \rangle$$

for all $v, w \in V$. If $V = \mathbb{C}^n$, the $n \times n$ matrix T is Hermitian if and only if its conjugate transpose is itself.

Why are we interested in self-adjoint operators? The main reason is that the spectral theorem, which we will learn in the next section, applies to self-adjoint operators. Let us get a glimpse of the theorem.

Exercise 5.3.I. Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a self-adjoint operator. Show that every eigenvalue of T is real, i.e., if $Tv = \lambda v$ for $v \neq 0$ and $v \in k$ then $\lambda \in \mathbb{R}$. (This statement is nontrivial only when $k = \mathbb{C}$.)

Exercise 5.3.J. Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a self-adjoint operator. If v_1 is an eigenvector with eigenvalue λ_1, v_2 is an eigenvector with eigenvalue λ_2, and $\lambda_1 \neq \lambda_2$, show that v_1 and v_2 are orthogonal, i.e., $\langle v_1, v_2 \rangle = 0$.

This is what it means. When V is a finite-dimensional Hilbert space with T an operator on it, for each $\lambda \in \mathbb{R}$ we can define

$$V_\lambda = \ker(T - \lambda I) = \{v \in V : Tv = \lambda v\}.$$

Then V_λ is nonzero only for finitely many λ, and all these vector spaces are orthogonal to each other.

Exercise 5.3.K. Let V be a finite-dimensional Hilbert space with T an operator on it. For a vector $v \in V$, show that there exists at most one representation

$$v = v_1 + v_2 + \cdots + v_n$$

with each $v_i \neq 0$ an eigenvector with eigenvalue λ_i, and $\lambda_1 < \lambda_2 < \cdots < \lambda_n$.

5.4 The spectral theorem

Given an abstract vector space V and a linear map $T : V \to V$, many of the important properties of T are determined by the $k[x]$-module structure on V. The module structure is then almost characterized by the set of eigenvalues of T (with multiplicity).

In the context of analysis, the set of eigenvalues is called the **spectrum** of a linear operator. Let us look at one example. Consider the Hilbert space

$$\begin{aligned}\mathcal{H} &= \{\text{complex-valued 1-periodic functions on } \mathbb{R}\} \\ &= \{(f : \mathbb{R} \to \mathbb{C}) : f(x) = f(x+1) \text{ for all } x \in \mathbb{R}\}.\end{aligned}$$

Well, this is only a vector space over a $\mathbb{C}$ and we need to specify the inner product. We can define the inner product as

$$\langle f, g \rangle = \int_0^1 \bar{f}(x) g(x) dx.$$

Consider the differential operator

$$T = \tfrac{d^2}{dx^2} : \mathcal{H} \to \mathcal{H}; \quad f \mapsto \frac{d^2 f}{dx^2}.$$

Technically, not all functions are differentiable, and hence the operator is not really defined. But let us overlook this issue. We have

$$\langle f'', g \rangle = \int_0^1 \bar{f}'' g dx = -\int_0^1 \bar{f}' g' dx = \int_0^1 \bar{f} g'' dx = \langle f, g'' \rangle$$

by integration by parts, and this shows that $T = \frac{d^2}{dx^2}$ is self-adjoint. What is the spectrum of this operator? To answer this question, we need to solve the differential equation

$$\frac{d^2 f}{dx^2} = \lambda f$$

where λ is a constant. It turns out that a solution exists only when $\lambda = 4\pi^2 n^2$ and $f(x) = ce^{2\pi i n}$, with $n \in \mathbb{Z}$. This set $\{0, 4\pi^2, 4\pi^2, 16\pi^2, 16\pi^2, \ldots\}$ is then the spectrum of T.

Before stating the spectral theorem, we state and prove a lemma that will be used in its proof.

Lemma 5.4.1. *Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a self-adjoint linear operator. If $\dim V \geq 1$, then there exists a nonzero $v \in V$ such that*

$$Tv = \pm\|T\|v.$$

Proof. By the definition of $\|T\|$ (and Heine–Borel), there exists a $v \neq 0$ such that $\|Tv\| = \|T\|\|v\|$. Note that $T^2 v = (T^2 v - \|T\|^2 v) + \|T\|^2 v$. Observe that the inner product of the two summands is

$$\langle T^2 v - \|T\|^2 v, \|T\|^2 v\rangle = \|T\|^2 \langle T^2 v, v\rangle - \|T\|^4 \langle v, v\rangle = \|T\|^2 \|Tv\|^2 - \|T\|^4 \|v\|^2 = 0$$

because $\langle T^2 v, v\rangle = \langle Tv, Tv\rangle$. This implies that the length of $T^2 v$ can be expressed as

$$\|T^2 v\|^2 = \|T^2 v - \|T\|^2 v\|^2 + \|T\|^4 \|v\|^2 \geq \|T\|^4 \|v\|^2.$$

On the other hand,

$$\|T^2 v\|^2 \leq \|T\|^2 \|Tv\|^2 \leq \|T\|^4 \|v\|^2.$$

Therefore all the inequalities are equalities. In particular, $T^2 v - \|T\|^2 v = 0$.

If $Tv = -\|T\|v$, we are done. Hence assume that

$$w = Tv + \|T\|v \neq 0.$$

Then

$$Tw = T^2 v + \|T\|Tv = \|T\|Tv + \|T\|^2 v = \|T\|w$$

finishes the proof. □

Exercise 5.4.A. Without the self-adjointness assumption, find a counterexample to the above lemma.

Exercise 5.4.B. Let V be a nonzero finite-dimensional Hilbert space and $T : V \to V$ be a self-adjoint operator. Show that

$$\|T\| = \max_{\|v\| \leq 1} |\langle Tv, v\rangle|.$$

Moreover, show that if $|\langle Tv, v\rangle| = \|T\|\|v\|^2$ then $Tv = \pm\|T\|v$.

Let us now state and prove the spectral theorem.

A self-adjont operator has an orthonormal set of eigenvectors

Theorem 5.4.2 (spectral theorem)**.** *Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a self-adjoint operator. Then there exist and orthonormal basis $v_1, v_2, \ldots, v_n \in V$ of eigenvectors, so that*

$$\langle v_i, v_j\rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases} \qquad Tv_i = \lambda_i v_i$$

where λ_i are (not necessarily distinct) real numbers.

Proof. The idea is to find one eigenvector at a time. We induct on $\dim V$. If $\dim V = 0$, the statement is trivial. If $\dim V \geq 1$, Lemma 5.4.1 finds a nonzero vector $v \in V$ such that $Tv = \lambda v$ where $\lambda = \pm\|T\|$.

Take the orthogonal complement of v,

$$W = v^{\perp} = \{w \in V : \langle w, v\rangle = 0\} \subseteq V.$$

Then for an arbitrary $w \in W$, we have

$$\langle Tw, v\rangle = \langle w, Tv\rangle = \langle w, \lambda v\rangle = \lambda\langle w, v\rangle = 0.$$

This shows that T sends a vector in W to W. Hence we may restrict T to a self-adjoint operator $T : W \to W$, where $\dim W = \dim V - 1$. By the induction hypothesis, we can find an orthonormal basis of W consisting of eigenvectors. Together with $v/\|v\|$, they form an orthonormal basis of V consisting of eigenvectors. □

Exercise 5.4.C (spectral decomposition)**.** Let A be a $n \times n$ Hermitian (resp. symmetric) matrix with complex (resp. real) entries. Show that there exists a unitary (resp. orthogonal) matrix U (resp. Q) and a diagonal matrix Λ such that

$$A = U\Lambda U^{-1} \text{ (resp. } Q\Lambda Q^{-1}).$$

Exercise 5.4.D. Consider the symmetric real matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

Find an orthogonal matrix Q and a diagonal matrix Λ such $A = Q\Lambda Q^{-1}$.

There is another way to state the spectral theorem. Consider the eigenspace

$$V_\lambda = \ker(T - \lambda I) = \{v \in V : Tv = \lambda v\} \subseteq V.$$

Then $\dim V_\lambda > 0$ if and only if λ is an eigenvalue of T.

Exercise 5.4.E. Show that

$$V = \bigoplus_{\lambda \in \mathbb{R}} V_\lambda,$$

where the spaces V_λ are orthogonal to each other. (Translation: for each $v \in V$ there uniquely exist vectors $v_\lambda \in V_\lambda$, zero except for finitely many λ, such that $v = \sum_\lambda v_\lambda$.)

Commuting self-adjoint operators can be simultaneously diagonalized

Exercise 5.4.F (simultaneous diagonalization)**.** Let V be a finite-dimensional Hilbert space, and let $X, Y : V \to V$ be self-adjoint operators. Suppose that X and Y commutes, i.e., $XY = YX$.

(a) Consider the eigenspaces $V_\lambda = \ker(X - \lambda I)$ with respect to X so that $V = \bigoplus_\lambda V_\lambda$. Show that Y sends everything in V_λ into V_λ, i.e., Y restricts to $Y : V_\lambda \to V_\lambda$ for each $\lambda \in \mathbb{R}$.

(b) Consider the simultaneous eigenspaces

$$V_{\lambda,\kappa} = \ker(X - \lambda I) \cap \ker(Y - \kappa I) = \{v \in V : Xv = \lambda v, Yv = \kappa v\}.$$

Show that $V = \bigoplus_{\lambda,\kappa} V_{\lambda,\kappa}$ where the spaces $V_{\lambda,\kappa}$ are all orthogonal to each other.

(c) Show that there exists an orthonormal basis $v_1, \dots, v_n$ of V such that each v_k is an eigenvector of both X and Y.

(d) If X, Y are $n \times n$ Hermitian matrices and $XY = YX$, show that there exists a unitary matrix U such that both $U^{-1}XU$ and $U^{-1}YU$ are diagonal.

The spectral theorem applies only to self-adjoint operators. If the base field is $k = \mathbb{R}$, diagonalization by an orthogonal operator exists if and only if the operator is symmetric.

Exercise 5.4.G. Let V be a finite-dimensional Hilbert space over $k = \mathbb{R}$. Show that for a linear operator $T : V \to V$, the following are equivalent:

(a) The operator T is symmetric.

(b) There exists an orthonormal basis $v_1, \dots, v_n$ consisting of eigenvectors.

(c) There exists an isometry $Q : \mathbb{R}^n \to V$ and an diagonal matrix $\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ such that $T = Q\Lambda Q^{-1}$.

However for $k = \mathbb{C}$, there is a larger class of matrices that can be diagonalized by a unitary matrix. The reason is that self-adjoint operators can only have real eigenvalues. If we take a unitary matrix U, a diagonal matrix Λ with non-real entries, and consider the linear operator $T = U\Lambda U^{-1}$, it will have an orthonormal set of eigenvectors but not necessarily self-adjoint.

Definition 5.4.3. Let V be a finite-dimensional Hilbert space over $k = \mathbb{C}$. We say that a linear operator $T : V \to V$ is **normal** if $T^\dagger T = TT^\dagger$.

Proposition 5.4.4. *Let V be a finite-dimensional Hilbert space over $k = \mathbb{C}$, and let $T : V \to V$ be a linear operator. The following are equivalent:*

(i) *T is normal.*

(ii) *The Hermitian part $X = \frac{1}{2}(T + T^\dagger)$ and skew-Hermitian part $Y = \frac{1}{2}(T - T^\dagger)$ commutes, i.e., $XY = YX$.*

(iii) *$\|Tv\| = \|T^\dagger v\|$ for all $v \in V$.*

(iv) *T is diagonalizable by a unitary matrix, i.e., there exists an isometry $U : k^n \to V$ and a diagonal matrix Λ such that $T = U\Lambda U^{-1}$.*

Proof. (i) ⇔ (ii) We compute

$$XY - YX = \frac{1}{4}[(T + T^\dagger)(T - T^\dagger) - (T - T^\dagger)(T + T^\dagger)] = \frac{1}{2}(T^\dagger T - TT^\dagger).$$

So $XY = YX$ if and only if $T^\dagger T = TT^\dagger$ if and only if T is normal.

(i) ⇔ (iii) Define $S = T^\dagger T - TT^\dagger$, which is easily seen to be a self-adjoint operator. Then

$$\|Tv\|^2 - \|T^\dagger v\|^2 = \langle Tv, Tv\rangle - \langle T^\dagger v, T^\dagger v\rangle = \langle v, (T^\dagger T - TT^\dagger)v\rangle = \langle v, Sv\rangle.$$

If T is normal so that $S = 0$, then this is always zero. Conversely, suppose that $\langle v, Sv\rangle = 0$ for all $v \in V$. By Exercise 5.4.B, this implies that $\|S\| = 0$, that is, $S = 0$.

(ii) ⇒ (iv) We note that $X^\dagger = \frac{1}{2}(T^\dagger + T) = X$, i.e., X is Hermitian. On the other hand, $Y^\dagger = -Y$ shows that iY is Hermitian. By Exercise 5.4.F, there exists a unitary matrix U such that $U^{-1}XU$ and $U^{-1}(iY)U$ are both diagonal matrices. Then

$$T = X + Y = U[U^{-1}XU + (-i)(U^{-1}(iY)U)]U^{-1},$$

where $U^{-1}XU + (-i)(U^{-1}(iY)U)$ is a sum of diagonal matrices, hence diagonal.

(iv) ⇒ (i) This can be verified directly. If $T = U\Lambda U^{-1}$ then

$$TT^\dagger = (U\Lambda U^{-1})(U\Lambda^\dagger U^{-1}) = U(\Lambda\Lambda^\dagger)U^{-1} = U(\Lambda^\dagger\Lambda)U^{-1} = T^\dagger T.$$

Here, $\Lambda^\dagger\Lambda = \Lambda\Lambda^\dagger$ because Λ is a diagonal matrix. □

Commuting normal matrices can be simultaneously diagonalized

Exercise 5.4.H (simultaneous diagonalization)**.** Let V be a finite-dimensional Hilbert space over $k = \mathbb{C}$. Suppose $T_1, T_2, \ldots, T_k$ are normal operators $V \to V$ such that $T_iT_j = T_jT_i$ for all $1 \le i, j \le k$. Show that there exists an isometry $U : \mathbb{C}^n \to V$ such that $U^{-1}T_iU$ is a diagonal matrix for all $1 \le i \le k$.

5.5 Positivity of operators

We now restrict our attention further to a smaller class of operators.

Definition 5.5.1. Let V be a finite-dimensional Hilbert space. An operator $T : V \to V$ is called

- **positive definite** if it is self-adjoint and $\langle v, Tv\rangle > 0$ for all nonzero $v \in V$,
- **positive semi-definite** if it is self-adjoint and $\langle v, Tv\rangle \geq 0$ for all $v \in V$.

Clearly, being positive definite is a stronger condition than being positive semi-definite, as the terminology suggests.

Positive definite operators correspond to inner products

Exercise 5.5.A. Let V be a finite-dimensional Hilbert space. Show that the sesquilinear form

$$V \times V \to k; \quad (x, y) \mapsto \langle x, Ty\rangle$$

defines an inner product on V if and only if $T : V \to V$ is positive definite.

Exercise 5.5.B. Let V, W be finite-dimensional Hilbert spaces, and let $T : V \to W$ be a linear operator. Show that $T^\dagger T : V \to V$ is positive semi-definite.

Exercise 5.5.C. Let V be a finite-dimensional Hilbert space and $T : V \to V$ be a self-adjoint operator.

(a) Show that T is positive definite if and only if all the eigenvalues of T are positive.

(b) Show that T is positive semi-definite if and only if all the eigenvalues of T are nonnegative.

Positive semi-definite plus invertible is positive definite

Exercise 5.5.D. Let V be a finite-dimensional Hilbert space, and $T : V \to V$ be a positive semi-definite operator. If T is invertible, show that T is positive definite. Conversely, show that a positive definite operator is necessarily invertible.

The good thing about positive definite or semi-definite operators is that there are many things we can do with positive numbers, e.g., taking the square root. Let $T : V \to V$ be a positive semi-definite operator. Apply the spectral theorem to get a decomposition

$$T = U\Lambda U^{-1}.$$

The condition that T is positive semi-definite is equivalent to all the diagonal entries of Λ being nonnegative. Then we can define $\sqrt{\Lambda}$ as the diagonal matrix with diagonal entries the square root of the diagonal entries of Λ. Then it is clear that $(\sqrt{\Lambda})^2 = \Lambda$. Moreover, if we define the **square root** of T as

$$\sqrt{T} = U\sqrt{\Lambda}U^{-1},$$

then $\sqrt{T}$ is positive definite and $(\sqrt{T})^2 = T$. An uncomfortable fact about this construction is that the decomposition $T = U\Lambda U^{-1}$ is not unique. However, the following proposition tells us that the square root is well-defined.

Positive square root of a positive semi-definite operator is unique

Proposition 5.5.2. *Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a positive semi-definite operator. Then there exists a unique positive semi-definite operator $S : V \to V$ such that $S^2 = T$.*

Proof. Existence of S is demonstrated by the above construction. Let us now consider a positive semi-definite S such that $S^2 = T$. Since both S and T are positive semi-definite, the spectral theorem gives decompositions

$$\begin{aligned} V &= \bigoplus_{\lambda \geq 0} V_\lambda^{(S)}, \quad V_\lambda^{(S)} = \ker(S - \lambda I), \\ V &= \bigoplus_{\lambda \geq 0} V_\lambda^{(T)}, \quad V_\lambda^{(T)} = \ker(T - \lambda I). \end{aligned}$$

Then for $v \in V_\lambda^{(S)}$ we have $S^2 v = S(\lambda v) = \lambda^2 v$, and hence $V_{\lambda^2}^{(T)} \subseteq V_\lambda^{(S)}$. Since $V = \bigoplus_{\lambda \geq 0} V_\lambda^{(S)} = \bigoplus_{\lambda \geq 0} V_{\lambda^2}^{(T)}$, we have $V_{\lambda^2}^{(T)} = V_\lambda^{(S)}$ for all $\lambda \geq 0$. Therefore $Sv = \sqrt{\lambda} v$ for $v \in V_\lambda^{(T)} = V_{\sqrt{\lambda}}^{(S)}$. This uniquely determines S. □

Obviously, if we do not require S to be positive semi-definite, there can by many self-adjoint S such that $S^2 = T$. For instance, take $S = \pm I$ and $T = I$. In this case, we will have $V_\lambda^{(T)} = V_{\sqrt{\lambda}}^{(S)} \oplus V_{-\sqrt{\lambda}}^{(S)}$ in the above proof.

One can take the cube root of a self-adjoint operator

Exercise 5.5.E. Let V be a finite-dimensional Hilbert space and $T : V \to V$ be a self-adjoint operator (not necessarily positive definite). Show that there exists a unique self-adjoint operator $S : V \to V$ such that $S^3 = T$. Moreover, show that S is positive semi-definite if T is positive semi-definite.

We have noted that $T^\dagger T$ is positive semi-definite for every linear operator T. There is a decomposition based on this fact.

Every linear operator has a polar decomposition

Theorem 5.5.3 (polar decomposition)**.** *Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a linear operator. There exists a unitary operator $U : V \to V$ and a positive semi-definite operator $P : V \to V$ such that*

$$T = UP.$$

Moreover, if T is invertible, such a decomposition is unique.

Proof. Note that if $T = UP$ then $T^\dagger T = (PU^{-1})(UP) = P^2$. So the only choice we have is $P = \sqrt{T^\dagger T}$, which makes sense since $T^\dagger T$ is positive semi-definite. It now suffices to show that there is a unitary matrix U such that

$$T = U\sqrt{T^\dagger T}.$$

By Lemma 5.5.4, we need only check that $\|Tv\| = \|\sqrt{T^\dagger T} v\|$. But

$$\|Tv\|^2 = \langle Tv, Tv \rangle = \langle v, T^\dagger T v \rangle = \langle v, (\sqrt{T^\dagger T})^2 v \rangle = \|\sqrt{T^\dagger T} v\|^2$$

because $\sqrt{T^\dagger T}$ is self-adjoint.

If T is invertible, U is uniquely determined by the formula $U = TP^{-1}$. □

Lemma 5.5.4. *Let V, W be finite-dimensional Hilbert spaces, and let $T, S : V \to W$ be linear operators. If $\|Tv\| = \|Sv\|$ for all $v \in V$, then there exists an isometry $U : W \to W$ such that $S = UT$.*

Proof. The condition implies that $Tv = 0$ if and only if $Sv = 0$, i.e., $\ker T = \ker S$. By the first isomorphism theorem, there exists a linear isomorphism $U_0 : \operatorname{im} T \to \operatorname{im} S$ such that $U_0 T = S$.

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[dr, "S"'] & \operatorname{im} T \arrow[d, dashed, "U_0", "\cong"'] \\
& \operatorname{im} S
\end{tikzcd}$$

The condition $\|Tv\| = \|Sv\|$ implies that U_0 is further an isometry.

Let $w_1, \dots, w_n$ be an orthonormal basis of W such that $w_1, \dots, w_k$ is an orthonormal basis of $\operatorname{im} T \subseteq W$. Since U_0 is an isometry, the images $U_0 w_1, \dots, U_0 w_k$ form an orthonormal basis of $\operatorname{im} S \subseteq W$. We now extend $U_0 w_1, \dots, U_0 w_k$ to an orthonormal basis $U_0 w_1, \dots, U_0 w_k, \tilde{w}_{k+1}, \dots, \tilde{w}_n$ of W. Now we can extend $U_0 : \operatorname{im} T \to \operatorname{im} S$ to $U : W \to W$ by defining

$$Uw_i = \begin{cases} U_0 w_i & \text{if } 1 \leq i \leq k \\ \tilde{w}_i & \text{if } k+1 \leq i \leq n. \end{cases}$$

Because U sends an orthonormal basis to an orthonormal basis, it is an isometry. Moreover, $S = UT$ by construction. □

Exercise 5.5.F. Find a polar decomposition for the matrix

$$T = \begin{bmatrix} 3 & 4 & 4 & 3 \\ 1 & -2 & -2 & 1 \\ 3 & 2 & -2 & -3 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Exercise 5.5.G. Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a linear operator. Show that there exist isometries $U : k^n \to V$ and $W : V \to k^n$ and a $n \times n$ diagonal matrix Λ with nonnegative diagonal entries such that

$$T = U \Lambda W.$$

Exercise 5.5.H. Let V be a finite-dimensional Hilbert space, and let $T : V \to V$ be a linear operator. Show that the eigenvalues of $T^\dagger T$ and of $TT^\dagger$ are identical.

The above decomposition $T = U\Lambda W$ is called **singular value decomposition**. But the interesting fact is that this decomposition holds for rectangular matrices as well.

Every linear operator has a singular value decomposition

Theorem 5.5.5 (singular value decomposition)**.** *Let H, G be finite-dimensional Hilbert spaces (possibly of different dimension) and let $T : H \to G$ be a linear operator. Then there exist isometries $U : k^m \to G$ and $V : H \to k^n$, and a diagonal matrix with nonnegative diagonal entries $\Sigma : k^n \to k^m$ such that*

$$T = U\Sigma V.$$

(Here, a diagonal rectangular matrix is a matrix $\Sigma = (\sigma_{ij})$ such that $\sigma_{ij} = 0$ if $i \neq j$.)

Proof. Note that if we have a singular decomposition for $T^\dagger$, say $T^\dagger = U\Sigma V$, then we can take the adjoint of both sides and get

$$T = V^\dagger \Sigma^\dagger U^\dagger.$$

Here, $V^\dagger$ and $U^\dagger$ are isometries as well, and hence we get a singular decomposition for T. Thus we may prove existence of a singular decomposition either for T or $T^\dagger$, and hence we may as well assume that $n = \dim H \leq \dim G = m$.

Because $\dim H \leq \dim G$, there exists a linear subspace $G_0 \subseteq G$ such that $\operatorname{im} T \subseteq G_0$ and $\dim G_0 = \dim H$. Let us factor $T : H \to G$ as $T_0 : H \to G_0$ composed with $i : G_0 \hookrightarrow G$. Sending an orthonormal basis of H to an orthonormal basis of G_0 defines an isometry $\Phi : G_0 \to H$.

$$\begin{tikzcd}
H \arrow[r, "T_0"] & G_0 \arrow[r, hook, "i"] \arrow[d, shift left, "\Phi"] & G \\
& H \arrow[u, shift left, "\Phi^{-1}"] &
\end{tikzcd}$$

We can now apply Exercise 5.5.G to the operator $\Phi T_0 : H \to H$. This gives a decomposition

$$\Phi T_0 = U\Sigma V, \quad T_0 = (\Phi^{-1}U)\Sigma V.$$

But this T_0 is an operator $H \to G_0$, and to get the original T, we need to compose with the inclusion $i : G_0 \hookrightarrow G$. So

$$T = i(\Phi^{-1}U)\Sigma V.$$

We need to take care of the embedding i. At this point, the operator $(\Phi^{-1}U) : k^n \to G_0$ is a composition of isometries, hence also isometry. Then we extend this isometry $\Phi^{-1}U$ to an isometry $\tilde{U} : k^m \to G$ so that the following diagram commutes.

$$\begin{tikzcd}
k^n \arrow[r, "\Sigma"] \arrow[dr, "\tilde{\Sigma}"'] & k^n \arrow[r, "\Phi^{-1}U"] \arrow[d, hook] & G_0 \arrow[d, hook, "i"] \\
& k^m \arrow[r, dashed, "\tilde{U}"] & G
\end{tikzcd}$$

(To see existence of $\tilde{U}$, we can use Lemma 5.5.4. It is not hard to verify that the two operators $k^n \to k^m$ and $i\Phi^{-1}U$ satisfy the assumptions.) Now the composition $\tilde{\Sigma} : k^n \xrightarrow{\Sigma} k^m \hookrightarrow k^m$ is a diagonal matrix, and $\tilde{U}$ is a unitary operator so that $i\Phi^{-1}U\Sigma = \tilde{U}\tilde{\Sigma}$. Then

$$T = \tilde{U}\tilde{\Sigma}V.$$

This finishes the proof. □

Exercise 5.5.I. Show that in the singular value decomposition of T, the diagonal entries of Σ (including multiplicities) does the depend on the decomposition. The set of diagonal entries is called the **singular values** of T.

For instance, the singular values of a self-adjoint operator is the same as its eigenvalues, because the decomposition $T = U\Lambda U^{-1}$ is a singular decomposition.

Exercise 5.5.J. Find the singular value decomposition of the matrix

$$T = \begin{bmatrix} 1 & 7 & -1 & -7 \\ 7 & -1 & -7 & 1 \\ 5 & 5 & 5 & 5 \end{bmatrix}.$$

Singular values measure the distance from the space of operators with small rank

Exercise 5.5.K. Let V, W be finite-dimensional Hilbert spaces, and let $T : V \to W$ be a linear operator. Let the singular values of T be

$$s_1 \geq s_2 \geq \cdots \geq s_n \geq 0,$$

where $n = \min(\dim V, \dim W)$. Show that

$$s_k = \min\{\|T - L\| : (L : V \to W) \text{ is a linear operator with } \operatorname{rank} L < k\},$$

for $1 \leq k \leq n$.

5.6 Duality in linear programming

Epilogue

Congratulations, and thank you for going through this book. As I have said in the preface, linear algebra is a universal prerequisite for most mathematics. This also means that once you become familiar with linear algebra, there are many directions in which you can proceed. I would like to discuss what advanced mathematics one can learn having digested linear algebra.

Calculus in higher dimensions

We will assume that the reader already knows calculus in a single variable. To do calculus in many variables, we can simply use $\mathbb{R}^n$, but let me introduce the concept of a manifold.

Imagine being a physicist. You are faced with the problem of modeling the universe. Everyone knows that there are three linearly independent directions in which we can move. So the natural candidate for the universe is $\mathbb{R}^3$. But you look back and think about the age when people thought the Earth is flat. Well, you can walk in two linearly independent directions at any point on Earth, but $\mathbb{R}^2$ turns out to be a terribly wrong model. Why can't this happen for the universe?

Definition 5.6.1. Let $n \geq 1$ be an integer. A **manifold** of dimension n is a space that locally looks like $\mathbb{R}^n$.

Of course this is a non-rigorous description, but it captures a important idea that we only want a space to *locally* look like Euclidean space. Let us look at a few examples. The circle

$$S^1 = \{(x, y) : x^2 + y^2 = 1\}$$

is a 1-manifold. The set

$$X = [-1, 1] \times \{0\} \cup \{0\} \times [-1, 1] \subseteq \mathbb{R}^2$$

is not, because near $(0, 0)$ it doesn't look like a line. The doughnut

$$T^2 = \{(x, y, z) : z^2 + (\sqrt{x^2 + y^2} - 1)^2 = 0.5\} \subseteq \mathbb{R}^3$$

is a 2-manifold.

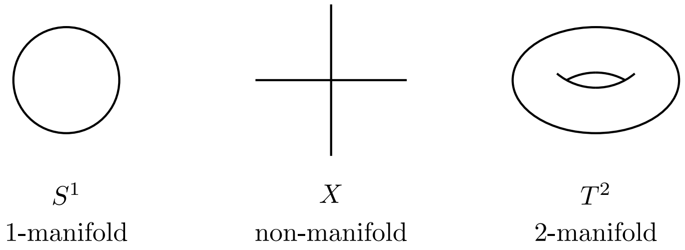

Figure 5.1: Manifolds and non-manifolds

Let us now fix an n-dimensional manifold M and consider an infinitely differentiable function $f : M \to \mathbb{R}$. Let us think about how we may define the derivative df of f. When $M = \mathbb{R}$, the derivative f' was defined as

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

But for a general manifold M, what does Δx mean? It should be an infinitesimal displacement of x, so that $x + \Delta x$ is a point very close to x. So we take Δx to be a tangent vector to the manifold at x.

It is difficult to precisely describe what a tangent vector is, but it should be considered as a "direction" with which a point can move, also remembering the "speed". A tangent vector defines a differential operator by taking the difference quotient,

$$D_v(f) = \lim_{\epsilon \to 0} \frac{f(p + \epsilon v) - f(p)}{\epsilon}.$$

This is sometimes called the directional derivative of f.

If we denote by $C_\infty(M)$ the vector space of infinitely differentiable functions on M, then the $\mathbb{R}$-linear functional $D_v : C^\infty(M) \to \mathbb{R}$ can be considered a differential operator. It is characterized by satisfying the Leibniz identity

$$D_v(fg) = D_v(f)g(p) + f(p)D_v(g).$$

The set of tangent vectors at a fixed point $p \in M$ naturally forms an n-dimensional $\mathbb{R}$-vector space (by $D_{v+w} = D_v + D_w$ and $D_{cv} = cD_v$), and this vector space T_pM is called the **tangent space** of M at p.

So given a function $f \in C_\infty(M)$ (i.e., a function $f : M \to \mathbb{R}$), we get a linear map

$$df|_p : T_pM \to \mathbb{R}; \quad v \mapsto D_v(f).$$

That is, for each point $p \in M$, we get an element of the dual tangent space $T_p^*M = (T_pM)^*$, also called the **cotangent space**. This assignment to each point $p \in M$ an element of T_pM is called a **1-form**.

Using the language of 1-forms, we can also formulate the fundamental theorem of calculus. Consider a curve γ in M, which is a smooth function γ :

$[a,b] \to M$ for some interval $[a,b] \subseteq \mathbb{R}$. Then for a partition $a = a_0 < a_1 < \cdots < a_{k-1} < a_k = b$, we can write

$$f(\gamma(b)) - f(\gamma(a)) = \sum_{i=0}^{k-1} \big(f(\gamma(a_{i+1})) - f(\gamma(a_i)) \big) \approx \sum_{i=0}^{k-1} (df)(\gamma(a_{i+1}) - \gamma(a_i)),$$

where the summand is evaluating the "tangent vector" $\gamma(a_{i+1}) - \gamma(a_i)$ at df. As $\max_i (a_{i+1} - a_i) \to 0$. The right hand side can be adequately interpreted as the integral $\int_\gamma df$. Then the equation can be rewritten as

$$f(\gamma(b)) - f(\gamma(a)) = \int_\gamma df.$$

There even is a higher-dimensional analogue of the fundamental theorem of calculus. This is now a bit complicated to explain, and hence we will only state the theorem without any explanation.

Definition 5.6.2. Let M be a manifold. A **k-form** ω is an assignment to each point $p \in M$ an element $\omega|_p \in \bigwedge^k T_p^* M$.

So a 0-form is just a function $f : M \to \mathbb{R}$, and a 1-form is an assignment to each $p \in M$ an element $\omega_p \in T_p^* M$. It turns out there is a way to take an arbitrary k-form ω and take its derivative to get a $(k+1)$-form $d\omega$.

Theorem 5.6.3 (Stokes's theorem). *Let M be a manifold of dimension m, and let N be a manifold with boundary of dimension n. Let ω be a $(n-1)$-form on M, and consider a smooth map $\varphi : M \to N$. Then*

$$\int_{\partial N} \omega = \int_N d\omega,$$

where ∂N is the boundary of N.

Consider the case when $n = 1$ and N is a interval. The differential form $\omega = f$ is a 0-form, which is a function, and the boundary of N is the two points $\gamma(a)$ and $\gamma(b)$. In this case, the integral of f over the two points is $f(\gamma(b)) - f(\gamma(a))$ and the right hand side is just $\int_\gamma df$. So we recover the fundamental theorem of calculus.

- For calculus with a single variable, there is Rudin's classic textbook *Principles of Mathematical Analysis*, which provides a rigorous and comprehensive introduction to analysis. It starts from scratch and also develops multivariable calculus near the end.
- There is a series of books on multivariable calculus written by Spivak. There is *Calculus on Manifolds*, which leads up to Stokes's theorem in the context of manifolds. There also is the more advanced book *A Comprehensive Introduction to Differential Geometry, Volume One* that discusses the topic in a geometric manner.
- Lee's *Introduction to Smooth Manifolds* contains almost the same material as in Spivak's *A Comprehensive Introduction*, but takes a differential topology perspective.

Representation theory

The Frobenius normal form, Jordan normal form, and related theorems convey information about the possible ways a linear map can act on a finite-dimensional vector space. More precisely, they tell us that we can find a suitable basis such that the linear map takes a certain form.

But what if there are many linear maps, and some of them has a certain relation they need to satisfy? To elucidate what I am trying to say, let us make a definition.

Definition 5.6.4. A **group** G is a set G along with a multiplication map $\mu : G \times G \to G$ satisfying the following axioms: (we denote $\mu(x, y) = xy$)

(G1) There exists an identity element $e \in G$ such that $eg = ge = g$ for all $g \in G$.

(G2) For all $g_1, g_2, g_3 \in G$, we have $(g_1 g_2)g_3 = g_1(g_2 g_3)$.

(G3) For every $g \in G$, there exists an inverse $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$.

Here is an example. Take the set $\mathbb{Z}$ with $\mu(x, y) = x + y$. One can readily check that it is a group. Here is a slightly more complicated example. Consider

$$S_n = \{(f : \{1, \dots, n\} \to \{1, \dots, n\}) \text{ bijective}\}$$

with composition as multiplication. For instance, $\tau \in S_3$ is defined as $\tau(1) = 2$, $\tau(2) = 1$, $\tau(3) = 3$, and $\sigma \in S_3$ is defined as $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 1$, then we have the identities

$$\tau^2 = 1, \quad \sigma^3 = 1, \quad \tau\sigma\tau = \sigma^2.$$

Once we have a group G, such as $G = S_3$, we might want to represent it as a matrix. For instance, if we associate

$$\tau = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix},$$

then the identities $\tau^2 = 1$, $\sigma^3 = 1$, $\tau\sigma\tau = \sigma^2$ are all satisfied.

Definition 5.6.5. Let G be a group, and let k be a field. A **representation** ρ of G is a finite-dimensional k-vector space V along with a function $\rho : G \to \mathrm{Hom}_k(V, V)$ satisfying

$$\rho(gh) = \rho(g)\rho(h)$$

for all $g, h \in G$.

If (V_1, ρ_1) and (V_2, ρ_2) are both representations of a group G, we can form the direct sum $V_1 \oplus V_2$, and the representation

$$(\rho_1 \oplus \rho_2)(g) = \rho_1(g) \oplus \rho_2(g).$$

It can be easily checked that this construction gives a representation of G.

Definition 5.6.6. Let G be a finite group. An **irreducible representation** of G is a representation V such that V cannot be expressed as $V = V_1 \oplus V_2$ as representations, where $\dim_k V_1, \dim_k V_2 > 0$.

Representations of finite groups over $k = \mathbb{C}$ has a nice characterization. First of all, all representations can be written as a direct sum of irreducible representations. This means that to study representations, it is enough to study irreducible representations. Here is the big theorem.

Theorem 5.6.7. *Let G be a finite group, and consider set of equivalence classes under the equivalence relation $g \sim hgh^{-1}$. The number of isomorphism classes of irreducible representations of G over $k = \mathbb{C}$ is equal to the number of equivalence classes of $g \sim hgh^{-1}$. Moreover, if we denote by $V_1, \dots, V_k$ all the irreducible representations, then*

$$(\dim_{\mathbb{C}} V_1)^2 + \cdots + (\dim_{\mathbb{C}} V_k)^2 = |G|.$$

Let me demonstrate the theorem with the group S_3. There are three equivalence classes for $g \sim hgh^{-1}$,

$$\{e\}, \quad \{\sigma, \sigma^2\}, \quad \{\tau, \sigma\tau, \sigma^2\tau\}.$$

Then there should be three irreducible representations. These are

$$\begin{aligned} \rho_1(\tau) &= \begin{bmatrix} 1 \end{bmatrix}, & \rho_1(\sigma) &= \begin{bmatrix} 1 \end{bmatrix}, \\ \rho_2(\tau) &= \begin{bmatrix} -1 \end{bmatrix}, & \rho_2(\sigma) &= \begin{bmatrix} 1 \end{bmatrix}, \\ \rho_3(\tau) &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & \rho_3(\sigma) &= \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}. \end{aligned}$$

And we have $1^2 + 1^2 + 2^2 = 6$, which is the size of the group.

This theorem pretty much tells us everything about finite-dimensional representations of finite groups over $\mathbb{C}$. A more interesting topic to study is representations of Lie groups or of Lie algebras.

Definition 5.6.8. A **Lie group** is a smooth manifold that is also a group.

Let us take S^1 for example. We may identify S^1 with the unit complex numbers.

$$S^1 \quad \leftrightarrow \quad \{z \in \mathbb{C} : |z| = 1\}$$

Then S^1 naturally has a multiplicative structure, and it can be easily verified that this is a group. Another example is

$$\mathrm{SU}(2) = \{2{\times}2 \text{ complex matrices } U \text{ such that } UU^\dagger = U^\dagger U = I \text{ and } \det U = 1\}.$$

Because the product of unitary matrices is unitary, and the determinant is multiplicative, this is a group. It is harder to show that SU(2) is manifold, but it turns out that $\mathrm{SU}(2) \cong S^3$.

We might ask what are the finite-dimensional irreducible representations of SU(2) over $\mathbb{C}$. It turns out that there is a nice answer to this.

Theorem 5.6.9 (represntation theory of SU(2))**.** *For each integer $n \geq 1$, there exists an n-dimensional representation (V_n, ρ_n) such that $V_1, V_2, \dots$ are all the irreducible representations up to isomorphism. Moreover, we have*

$$V_n \cong \operatorname{Sym}^{n-1} V_2.$$

Such a nice description exists when the Lie group G is compact and simply connected.

- For an introduction with many concrete examples, take a look at Fulton and Harris's *Representation Theory: A First Course*. The book starts with the representation theory of finite groups at the beginning, and later also discusses Lie groups and Lie algebras.
- There is also *Introduction to Representation Theory* written by Etingof et al., which is legally freely available on the internet. This book also begins with representation theory of finite groups, but delves into the combinatorial aspects of representations instead of Lie algebras.
- Kirillov's book *Introduction to Lie Groups and Lie Algebras* focuses on Lie algebras, Lie groups, and their representations. A preliminary version of the book is available for free from the author's website.

Functional analysis

In Chapter 5, we talked about finite-dimensional Hilbert spaces in detail. But we alluded that there is a notion of an infinite-dimensional Hilbert space.

Definition 5.6.10. Let $k = \mathbb{R}$ or $k = \mathbb{C}$. A **Hilbert space** is a vector space V over k is a vector space over k together with an inner product $\langle -, - \rangle$ such that

(H1) every Cauchy sequence in V (with respect to the norm) converges to a vector in V,

(H2) there exists a sequence of vectors $v_1, v_2, \dots \in V$ such that for any $v \in V$ and $\epsilon > 0$ there exists v_k such that $\|v - v_k\| < \epsilon$.

Here is an example. Consider the space

$$C^0(S^1) = \{(f : \mathbb{R} \to k) \text{ continuous} : f(x) = f(x+1) \text{ for all } x \in \mathbb{R}\},$$

and the inner product

$$\langle f, g \rangle = \int_0^1 \overline{f(x)} g(x) dx.$$

This is an inner product space that is infinite-dimensional, but it does not satisfy (H1). So we "complete" the vector space by adding in vectors if a Cauchy sequence does not converge. Define

$$L^2(S^1) = \text{completion of } C^0(S^1) \text{ with respect to the } L^2\text{-norm}.$$

This satisfies (H1) by construction, and it also satisfies (H2) because the set of piecewise linear functions with rational slope and bending points is both dense and countable.

For Hilbert spaces, many of the theorem we discuss holds.

Theorem 5.6.11 (Riesz represetation)**.** *Let V be a Hilbert space, and let $T : V \to k$ be a bounded linear operator. (Bounded means finite operator norm.) Then there exists a vector $v_0 \in V$ such that*

$$T = \langle v_0, - \rangle.$$

Moreover, $\|T\| = \|v_0\|$.

There is also a spectral theorem in this context. However, we need a somewhat restrictive condition to make the proof work.

Definition 5.6.12. Let V be a Hilbert space. A bounded linear operator $T; V \to V$ is called **compact** if for any sequence of $v_1, v_2, \ldots \in V$ with $\|v_i\| \leq 1$, there exists a subsequence $v_{n_1}, v_{n_2}, \ldots \in V$ with $n_1 < n_2 < \cdots$ such that $Tv_{n_1}, Tv_{n_2}, \ldots$ converges.

Theorem 5.6.13 (spectral theorem for compact operators)**.** *Let V be an infinite-dimensional Hilbert space, and let $T : V \to V$ be a compact self-adjoint operator. Then there exists an orthonormal basis of V consisting of eigenvectors for T. Moreover, if we denote the corresponding eigenvalues by $\lambda_1, \lambda_2, \ldots$ where*

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots$$

then $\lim_{n\to\infty} \lambda_n = 0$.

Although such a theorem seems satisfactory from a theoretic point of view, it is not very useful when doing analysis. We said that the differential operator $T : \frac{\partial^2}{\partial x^2}$ is roughly a self-adjoint operator because integration by parts gives

$$\int \bar{f}'' g = -\int \bar{f}' g' = \int \bar{f} g''.$$

In the attempt to apply the spectral theorem to this operator, the first problem we encounter is that $T : L^2(S^1) \to L^2(S^1)$ is not really defined on the entire domain. A function has to be twice-differentiable for its image under T to be defined. This is closely related to the fact that T is not a bounded operator. There is no universal constant $C > 0$ such that we have an inequality $\int_0^1 |f''|^2 < C \int_0^1 |f|^2$. Once the operator is not bounded, there is no hope for it to be compact.

One remedy to this issue is to try and apply the spectral theorem to $(\pi^2 + T)^{-1}$ instead of T. But what is $(\pi^2 + T)^{-1}$? If we define the operator

$$(Sf)(x) = \int_0^1 K(x,y) f(y) dy, \quad K(x,y) = \frac{1}{2\pi} \sin(\pi |x - y|),$$

then it is not difficult to check that $(\pi^2+T)Sf = f$ if f is continuous. Moreover, when we consider S as a operator $S : L^2(S^1) \to L^2(S^1)$ (which makes sense now), it turns out that it is a compact operator. Applying the spectral theorem to S gives the eigenvectors and eigenvalues

$$f_n(x) = e^{2\pi nix}, \quad \lambda_n = \frac{1}{\pi^2(1+4n^2)}.$$

Indeed, the functions $f_n(x)$ form an orthonormal basis of $L^2(S^1)$, and $\lambda_n \to 0$ as$n \to \pm\infty$. Because these are the eigenvalues for $S = (\pi^2+T)^{-1}$, we may take f_n to be eigenvectors for T with eigenvalues $4\pi^2 n^2$.

As you might notice, to make analysis rigorous takes a lot of care and energy. There are various complicated spaces and operators that are tailored to be useful in each situation. The entire theory was developed as an abstract foundation for solving differential equations, and hence I personally think it will be meaningless to study only topological vector spaces without context.

- Most constructions of Hilbert spaces, especially those used when solving differential equations, involve measure theory. There is a short introduction at the end of Rudin's *Principles of Mathematical Analysis*, and a neat formal development appears at the beginning of *Real and Complex Analysis* written by the same author.
- Stein and Shakarchi's books are great ways to learn real analysis. Assuming background in differentiation and Riemann integration, the third volume *Real Analysis* starts with a discussion of measure theory, develops the theory of Hilbert spaces, and discuss applications. The fourth volume *Functional Analysis* goes into studying Banach spaces, in particular L^p spaces, distributions, and applications in constant coefficient partial differential equations.
- Rudin has a third textbook *Functional Analysis*, which focuses more on the algebraic aspects of the theory with less applications to differential equations.
- If you are specifically interested in learning solving differential equations, there is the standard textbook *Partial Differential Equations* by Evans.

Homological algebra of modules

Recall that we had this theorem about exact sequences inducing exact sequences.

Theorem 5.6.14 (Exercise 2.8.F). *If $0 \to V_1 \to V_2 \to V_3 \to 0$ is a short exact sequence of vector spaces, then any other vector space W induces an exact sequence*

$$0 \leftarrow \mathrm{Hom}(V_1, W) \leftarrow \mathrm{Hom}(V_2, W) \leftarrow \mathrm{Hom}(V_3, W) \leftarrow 0.$$

This theorem uses the fact that W has a basis in an essential way. Hence in the context of modules, where bases need not exist, the theorem is false. For instance, let us look at the ring $R = \mathbb{Z}$ and the exact sequence

$$0 \to \mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \to 0.$$

If we apply $\mathrm{Hom}(-, \mathbb{Z}/2\mathbb{Z})$ to this sequence, we get

$$0 \leftarrow \mathbb{Z}/2\mathbb{Z} \xleftarrow{\times 2} \mathbb{Z}/2\mathbb{Z} \xleftarrow{\times 1} \mathbb{Z}/2\mathbb{Z} \leftarrow 0,$$

which is not exact at the left $\mathbb{Z}/2\mathbb{Z}$. (The kernel of $\mathbb{Z}/2\mathbb{Z} \to 0$ is $\mathbb{Z}/2\mathbb{Z}$, but the image of $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z}/2\mathbb{Z}$ is 0.) In fact, we only have the following.

Theorem 5.6.15 (Similar to Exercise 2.6.S)**.** *If $0 \to M_1 \to M_2 \to M_3 \to 0$ is a short exact sequence of modules, then any other module N induces an exact sequence*

$$\mathrm{Hom}(M_1, N) \leftarrow \mathrm{Hom}(M_2, N) \leftarrow \mathrm{Hom}(M_3, N) \leftarrow 0.$$

This asymmetry between the left and the right side is a bit annoying. Even if we want to know something about the cokernel of $\mathrm{Hom}(M_2, N) \to \mathrm{Hom}(M_1, N)$, we can't say much about it.

But it turns out that there is a way to extend this exact sequence further. Given two R-modules A and B, and an integer $n \geq 1$, there is a way to construct an R-module $\mathrm{Ext}^n(A, B)$. These collection of modules satisfy the following pleasant properties.

(1) If $0 \to M_1 \to M_2 \to M_3 \to 0$ is a short exact sequence, then any module N induces a long exact sequence

$$\begin{aligned} \cdots \leftarrow \mathrm{Ext}^2(M_2, N) \leftarrow \mathrm{Ext}^2(M_3, N) \leftarrow \\ \leftarrow \mathrm{Ext}^1(M_1, N) \leftarrow \mathrm{Ext}^1(M_2, N) \leftarrow \mathrm{Ext}^1(M_3, N) \leftarrow \\ \leftarrow \mathrm{Hom}(M_1, N) \leftarrow \mathrm{Hom}(M_2, N) \leftarrow \mathrm{Hom}(M_3, N) \leftarrow 0. \end{aligned}$$

(2) If $0 \to M_1 \to M_2 \to M_3 \to 0$ is a short exact sequence, then any module N induces a long exact sequence

$$\begin{aligned} 0 \to \mathrm{Hom}(N, M_1) \to \mathrm{Hom}(N, M_2) \to \mathrm{Hom}(N, M_3) \to \\ \to \mathrm{Ext}^1(N, M_1) \to \mathrm{Ext}^1(N, M_2) \to \mathrm{Ext}^1(N, M_3) \to \\ \to \mathrm{Ext}^2(N, M_1) \to \mathrm{Ext}^2(N, M_2) \to \cdots. \end{aligned}$$

(3) If A or B is a free R-module, then $\mathrm{Ext}^n(A, B) = 0$ for all $n \geq 1$.

The construction of these $\mathrm{Ext}^n(A, B)$ are quite complicated. You first find an exact sequence

$$0 \to B \to M_0 \to M_1 \to \cdots$$

of R-modules, where all M_i are free R-modules, and then define

$$\operatorname{Ext}^n(A, B) = \frac{\ker(\operatorname{Hom}(A, M_n) \to \operatorname{Hom}(A, M_{n+1}))}{\operatorname{im}(\operatorname{Hom}(A, M_{n-1}) \to \operatorname{Hom}(A, M_n))}.$$

It turns out that this does not depend on the choice of M_i, and satisfies all the properties above. Moreover, we have $\operatorname{Ext}^n(A, B) = 0$ for all $n \geq 2$, if the ring R is a principal ideal domain. So for instance, for $R = \mathbb{Z}$ we only have

$$\begin{aligned} 0 &\to \operatorname{Hom}_{\mathbb{Z}}(N, M_1) \to \operatorname{Hom}_{\mathbb{Z}}(N, M_2) \to \operatorname{Hom}_{\mathbb{Z}}(N, M_3) \\ &\to \operatorname{Ext}^1_{\mathbb{Z}}(N, M_1) \to \operatorname{Ext}^1_{\mathbb{Z}}(N, M_2) \to \operatorname{Ext}^1_{\mathbb{Z}}(N, M_3) \to 0. \end{aligned}$$

The construction and yoga of these modules Ext^n (there is also something called Tor_n) is called homological algebra. It might seem that the subject is quite dry and unmotivated, but homological algebra provides a lot of tools for finding invariants of geometric objects. In algebraic topology, there is something called singular (co)homology, that associates to each topological space an R-module, and in algebraic geometry, there is something called sheaf cohomology that associates to each sheaf on a variety a vector space. These invariants contain a lot of information about the spaces.

- It might be useful to get familiar with the language of categories. In fact, you might already be familiar with some concepts, if you went through this book. The first half of Mac Lane's *Categories for the Working Mathematician* is a great introduction.
- Weibel's *An introduction to homological algebra* goes straight into developing the formal framework of homological algebra.
- While it is possible to learn homological algebra purely algebraically, it is useful to know the applications (or motivations) for the subject. The standard introductory textbook in algebraic topology is Hatcher's *Algebraic Topology*, freely available on the author's website.
- Maybe you might want to know more about commutative rings before learning homological algebra. The study of commutative rings is called commutative algebra, and Atiyah and MacDonald's *Introduction to Commutative Algebra* is a concise and dense introduction.
- On the algebraic geometry side, Vakil's *Foundations of Algebraic Geometry* is a friendly and comprehensive introduction to algebraic geometry. The first chapter is a great reference for basic category theory and homological algebra language.

Index