MATH266 Linear Algebra II

ver. April 2021

A note on the text (April 2021)

In Winter 2020 the Department of Mathematics and Statistics completed is transition from having a pair of parallel second-year linear algebra courses (MATH264 and MATH266) to a single stream of linear algebra (MATH164 and MATH266). Originally, MATH264 was a terminal course designed for students who may not *need* linear algebra as part of their further studies. Whereas MATH266 was the first course in a two course sequence of linear algebra aimed at students in mathematics, computer science and physics. Following the reorganization of these courses, MATH264 was renamed as MATH164 and became a pre-requisite for MATH266. By making MATH164 a pre-requisite for MATH266, this latter course could immediately proceed with vector spaces as the object of interest, rather than spending a few weeks getting comfortable in $\mathbf{R}^n$.

In Winter 2021, MATH266 was offered remotely due to COVID-19. For students in MATH266 in that semester, the first draft of these notes comprised of their weekly reading for the course. Prior to each section of each chapter there was a 5-7 minute video introducing the broad ideas in the readings. Additionally, there were weekly drop-in office hours with the instructor to answer questions about course material. Each of office hours began with a mini-lecture on the previous week's topic. This approach proved successful and the hope is that students taking the course in the classroom can also find use for these notes. Doing so would require further editing to add material and remove commentary related to the specific offering for which these notes were authored.

These notes are written for students who are interested in mathematics but have little mathematics experience beyond secondary school mathematics and an introduction to computational techniques in linear algebra. Notably, the implied reader of these notes has no experience with mathematical proof, writing and formalism. Great effort is made to be explicit about notation, terminology and that choices in these regards are cultural; they are not inherently right or wrong, however best practice for communicating to others dictates that we adhere to standard usage.

Each section of the notes corresponds to roughly one week of material. However, due to the nature of remote learning, some material that should be presented in MATH266 does not appear in these notes. In particular, culminating material on trace, determinant and Jordan Canonical form is absent. Additionally, material on products and direct sums does not appear. This has the ripple of effect of a missing classification of diagonalizable finite-dimensional vector spaces using eigenspaces. (See Linear Algebra Done Right Thm 5.41).

Each section of the notes begins with a list of learning incomes and outcomes. The former list tells the reader what material they need to be familiar with in order to understand the upcoming material. Readers should return to the latter list once they have finished their work in section to be sure they have attained the learning outcomes.

Each section of the notes ends with a part titled *Test Your Understanding*. In general, these questions are not designed to challenge a reader. Instead they are meant to be a check of

broad understanding of the key ideas in the section. Challenging the learner to engage more fully with the material is a task left to assessed course work.

Throughout the text the reader will find short diversions under the heading **aside**. This fragments of text often present ideas beyond the stated learning outcomes and sometimes require students to have a more mature mathematical background than the stated learning incomes. These parts of the text can be fully ignored without detriment.

These notes are meant to be a companion to Axler's text *Linear Algebra Done Right*. In many places, these notes ask the reader to read corresponding text in Linear Algebra Done Right. These notes are mostly consistent with the approaches taken in this text, however these notes significantly expand the discussion of algebraic isomorphism.

The many errors herein are the sole responsibility of the original author. Note, however, that in some places this text intentionally opts for small fibs in pursuing a broader learning objective.

-cd

Contents

---

0 Things to Remember from Previous MATH Courses

This material is meant to be a reminder of some of the things you have seen in prior mathematics courses. Some of these are things you will have been taught explicitly, and others are things you would have been expected to pick up by academic osmosis. The notation below is chosen to be consistent with the course text for MATH 266, *Linear Algebra Done Right*.

Undoubtedly this will be the least interesting reading you do this semester. There are no companion videos for this reading because this module is mostly about notation – there isn't really anything for me to *teach*.

0.1 Sets

Intuitively, a mathematical set is a collection of unique objects. Let us begin by considering some sets that we may already familiar with.

Definition 0.1. *The set of* ***integers*** *is the set* $\{, \cdots -2-1, 0, 1, 2 \dots\}$. *We denote this set by* $\mathbf{Z}$

The ***rational numbers*** *are the elements of the set*

$$\left\{ \frac{p}{q} \mid p, q \in \mathbf{Z} \text{ and } q \neq 0 \right\}.$$

This set is denoted by $\mathbf{Q}$.

The set of ***real numbers*** *is the set of all numbers on the number line. We denote this set by* $\mathbf{R}$

In some texts these sets are denoted as $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$, respectively. We use the letter Z for integers because of the German word *zhalen*, which means numbers.

It is not an overstatement to say that sets underpin just about every mathematical concept that you have ever encountered. Thus your background mathematics knowledge already gives you some intuition about sets! And so likely you are already familiar with the following definition.

Definition 0.2. *When X is a set and x is a member of X, we say that x is an* ***element*** *of X. When x is an element of X we write $x \in X$. We write $x \notin X$ when x is not an element of X.*

Relying on our familiarity with the integers we may write $4 \in \mathbf{Z}$ since 4 is an integer and $\pi \notin \mathbf{Z}$ since π is not an integer. These notations are just shorthand for the sentences "4 is in the set of integers " and "π is not in the set of integers". The only reason to prefer the notation over the sentence is for brevity.

Our notation for set membership extends nicely to talk about multiple things being elements of the same set. For example, we can write "$2, 4 \in \mathbf{Z}$" to mean "$2 \in \mathbf{Z}$ and $4 \in \mathbf{Z}$".

In this class we are going to define a lot of terms. We do this so that we can develop a common vocabulary to communicate precise ideas. In every definition the newly defined term will be emboldened. From then on when we use that term we all have the same understanding of what the term means.

There are numerous ways to tell a reader what elements are contained in a particular set. If a set has few elements, then it may be defined by listing out the elements explicitly. For instance, $\{1, 3, 5, 7\}$ is the set of the first four odd numbers. The order in which we write the elements does not matter. The set $\{1, 3, 5, 7\}$ is the same set as $\{1, 5, 3, 7\}$. Repeating elements also does not matter. The set $\{1, 3, 5, 7\}$ is the same set as $\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 5, 7\}$

For sets with too many elements to list, we must provide the reader with a means to determine membership in the set. We can inform our reader that not all elements of the set have been listed, but that enough information has been provided for the reader to identify a pattern for determining membership in the set. For example, let X denote the set $\{2, 4, 6, 8, \ldots, 96, 98\}$. Here X is the set of positive even integers less than 100. However, using an ellipsis to define a set may not always work: it assumes that the reader will identify the pattern we wish to characterize.

The definitions of $\mathbf{Q}$ and $\mathbf{R}$ express sets in a different way than just listing out the elements. In defining $\mathbf{R}$ we have just written a sentence. This is a perfectly acceptable way to define a set. For example, it is perfectly reasonable to say *let S denote the set of students registered in MATH266.*

Let us look more closely at our definition of the rational numbers, $\mathbf{Q}$.

$$\mathbf{Q} = \left\{ \frac{p}{q} \mid \ p, q \in \mathbf{Z} \text{ and } q \neq 0 \right\}.$$

In fact this jumble of symbols is a sentence.

"$\mathbf{Q} =$ "	translates to	"The set $\mathbf{Q}$ is defined to be"
"$\frac{p}{q}$"	translates to	" those elements with the form $\frac{p}{q}$"
"$\mid$"	translates to	" where"
"$p, q \in \mathbf{Z}$ and $q \neq 0$"	translates to	" p and q are integers and q is not equal to 0"

And so when we see

$$\mathbf{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbf{Z} \text{ and } q \neq 0 \right\}$$

we think (and say and read) "The set $\mathbf{Q}$ is defined to be those elements with the form $\frac{p}{q}$ where p and q are integers and q is not equal to 0".

Our method for describing $\mathbf{Q}$ is easily adapted to define other sets. This method of describing a set is called *set-builder notation*. We will use this method quite often.

For example, consider the set

$$P_2 = \{a_2x^2 + a_1x + a_0 \mid a_0, a_1, a_2 \in \mathbf{R}\}$$

This is the set of all expressions of the form

$$a_2x^2 + a_1x + a_0$$

where each of a_0, a_1 and a_2 are permitted to take any real-number value. And so, P_2 is the set of all real polynomials of degree 2. For example, $2x^2 + 3 \in P_2$ (here $a_2 = 2, a_1 = 0$ and $a_0 = 3$)

For convenience we permit the existence of a set containing no objects. We call this set the **empty set** and we denote it as $\{\}$. Sometimes the empty set is denoted as $\emptyset$. This choice of notation (as well as every choice of notation!) is a matter of preference.

Just as we can relate a pair of real numbers using $=, \leq, \geq, <$ or $>$, we have analogue relations for sets.

Definition 0.3. *Let A and B be sets. We say A **equals** B when every element of A is an element of B and every element of B is an element of A. When A* equals *B we write $A = B$. When A does not equal B we write $A \neq B$.*

Definition 0.4. *Let A and B be sets.*

*When every element of A is also an element of B we say A **is a subset of** B. When A is a subset of B we write $A \subseteq B$.*

*When A is a subset of B and $A \neq B$ we say A **is a proper subset of** B and we write $A \subsetneq B$.*

Just as we an relate sets in a similar way to how we relate numbers, we can define operations for sets just like we have operations for numbers.

Definition 0.5. *Let X and Y be sets. The **union of** X **and** Y, denoted as $X \cup Y$, is the set*

$$X \cup Y = \{x, \mid x \in X \text{ or } x \in Y\}.$$

*The **intersection of** X **and** Y, written $X \cap Y$, is the set*

$$X \cap Y = \{x \mid x \in X \text{ and } x \in Y\}.$$

*The **cartesian product of** X **and** Y written $X \times Y$, is the set of all ordered pairs whose first element is from X and second element is from Y. That is*

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}.$$

For example, if $A = \{1, 2\}$ and $B = \{a, b\}$, then $A \times B = \{(1, a), (1, b), (2, a), (2, b)\}$.

Points in three dimensions are often specified with a triple of real numbers (x, y, z), where x is the coordinate on the x-axis, y is the coordinate on the y-axis and z is the coordinate on the z-axis. We often refer to the set of all points as $\mathbf{R}^3$. Much as we can use the cartesian product of sets to think about the cartesian plane, we can use the cartesian product to think about 3D space. Consider $\mathbf{R} \times \mathbf{R} \times \mathbf{R}$. We didn't explicitly define this, our definition of cartesian product was the product of two sets. But our choice of notation is flexible enough that we would likely understand $\mathbf{R} \times \mathbf{R} \times \mathbf{R}$ to denote the set of ordered triples where each element is an element of $\mathbf{R}$. And so even without providing a formal definition, we can reasonably consider the cartesian product of any number of sets.

0.2 Functions

Our intuition around function is likely some sense of "assigning". That is, a function assigns a value (an output) to each input. A function, then, in some sense, can be considered to be a set of pairs: the input and the assigned output.

Consider the following set

$$A = \{(z, z^2) \mid z \in \mathbf{R}\}$$

Using our notation from the previous section we notice $A \subseteq \mathbf{R} \times \mathbf{R}$. (recalling our definition of the notation $\subseteq$ and the notation $\times$ we should be convinced that every element of A is an ordered pair whose first entry is from $\mathbf{R}$ and whose second entry is from $\mathbf{R}$)

The set A has many elements, let us examine some of them.

- $(0, 0) \in A$ as $0 \in R$ and $0 = 0^2$
- $(-3, 9) \in A$ as $-3 \in \mathbf{R}$ and $9 = (-3)^2$
- $\left(\sqrt{2.1}, 2.1\right) \in A$ as $\sqrt{2.1} \in \mathbf{R}$ and $2.1 = \left(\sqrt{2.1}\right)^2$

With not a lot of thought we should be able to convince ourselves that the elements of A are exactly the points of the parabola $f(x) = x^2$. Similarly, we can construct such a set for any function. For example consider $g(x) = \sin(x)$. The set of points of this curve are exactly the points in the set

$$\{(x, \sin(x)) \mid x \in \mathbf{R}\}$$

Much like our set representation for $f(x) = x^2$ we have a set of ordered pairs in this set. We can consider the the second entry in the ordered pair to be "assigned" to the first entry.

With these examples in mind, we define a function as follows.

Definition 0.6. *Let A, B and f be sets. We say f **is a function from** A **to** B when f is a subset of $A \times B$ in which each element of A appears as the first entry of an ordered pair exactly once. When f is a function from A to B we write $f : A \to B$. We say that A is the **domain of** f and that B is the **codomain of** f*

Aside 0.7. *Let us parse the structure of this definition one part at a time:*

- Let A, B and f be sets.
  - *This sentence tells us the names of the objects we will need to define our new term.*

- We say f **is a function from** A **to** B
  - *The underlined part is the new term we are defining.*
- when f is a subset of $A \times B$ in which each element of A appears as the first entry of an ordered pair exactly once.
  - *This is what we mean we use our newly defined term. Our new term is shorthand to express this idea. In other words, this is the criteria that the set f must satisfy so that we may call it a function.*
- When f is a function from A to B we write $f : A \to B$.
  - *This sentence gives us a piece of notation to go along with our new definition. The piece of notation is shorthand for the underlined term in the second part.*
- We say that A is the **domain of** f and that B is the **codomain of** f
  - *Our new definition has a second definition hiding inside!*

The ordered pairs in f tell us which element of the codomain is assigned to each element of domain. To ensure that each element of A is assigned exactly one element of the codomain, we required that every element of A appears as the first entry of an ordered pair exactly once. Notice that we have no restriction on how many times an element of the codomain can appear? For $f(x) = x^2$ we are okay with having $(2, 4) \in f$ and $(-2, 4) \in f$.

Let us consider some examples.

Example 0.8. *First we consider the notion of functions with which we are likely intimately familiar: functions from* $\mathbf{R}$ *to* $\mathbf{R}$*. Consider the line* $y = 2x$*. Every point on this line is of the form* $(x, 2x)$ *and all points of the form* $(x, 2x)$ *appear on the line.*

Consider the set $f = \{(x, 2x) \mid x \in \mathbf{R}\}$*. We notice that for any particular* $z \in \mathbf{R}$*, the number* z *only appears in a single ordered pair: the ordered pair* $(z, 2z)$*. For example, the real number* 11 *only appears as the first entry of the ordered pair* $(11, 22)$*.* 11 *appears as the first entry of no other element of* f*. Thus this set fulfils our definition of* f is a function from $\mathbf{R}$ to $\mathbf{R}$.

Example 0.9. *Let* $A = \{1, 7, 9\}$*,* $B = \{1, 7, 9, 10\}$ *and*

$$f = \{(1, 7), (7, 7), (7, 9), (9, 9)\}$$

Using our new definition of function we can ask: is f *a function from* A *to* B*?*

Let us read the definition of function one part at a time:

- Let A, B and f be sets.

- *This is true. Certainly A, B and f are sets.*

- f is a subset of $A \times B$
  - *This is true; every element of f is an element of $A \times B$ as each element is an ordered pair which as its first entry from A and its second entry from B.*

- each element of A appears as the first entry of an ordered pair exactly once.
  - *This is false; 7 appears as the first entry of an ordered pair twice. Both of $(7, 7)$ are $(7, 9)$ elements of f.*

*The set f does not satisfy the definition of **function from** A **to** B and so we cannot say that f is a function from A to B.*

Let us return back to familiar territory with the function $f(x) = x^2$. Quite likely we understand the notation $f(7) = 49$. Put in terms of our new definition of a function, this is equivalent to saying $(7, 49) \in f$. As every function can be represented as set, we can extend our use of this piece of notation for any function.

Definition 0.10. *Let A and B be sets and let f be a function from A to B. For $(a, b) \in f$ we say b **is the image of** a and a **is a pre-image of** b. When b is the image of a we write $f(a) = b$.*

Our definition of function requires that each element of A appear as the first entry in exactly one ordered pair of f. Thus for each $a \in A$ the notation $f(a)$ has unambiguous meaning; $f(a)$ is the element of B so that the pair $(a, f(a))$ is an element of f.

We can give all of the information about a function by stating its domain, its codomain and the image of each element of the domain. This is what we are doing when we write something like: *let $f : \mathbf{R} \to \mathbf{R}$ so that $f(x) = 2x + 3$*. By convention we understand this statement of notation to mean: *let f be the function from $\mathbf{R}$ to $\mathbf{R}$ so that the image of an element $x \in \mathbf{R}$ is $2x + 3$*. This is equivalent to writing *let f be the function from $\mathbf{R}$ to $\mathbf{R}$ so that*

$$f = \{(x, 2x + 3) \,|x \in \mathbf{R}\}$$

Let us return back to the familiar ground of $f(x) = x^2$. How do we interpret the notation $f^{-1}(7)$? Perhaps the word *inverse* or *pre-image* slipped past our mental lips as we read that last sentence? (It is okay if it didn't!)

Definition 0.11. *Let A and B be sets, let f be a function from A to B and let b be an element of B. The **pre-image of** b is the set $\{a \in A | f(a) = b\}$. We denote this set as $f^{-1}(b)$.*

Let us pick apart this definition one sentence at a time:

- *Let A and B be sets, let f be a function from A to B and let b be an element of B.*
  - This sentence sets up the rest of the definition. It tells us about the mathematical objects we will need to define our new term.
- *The **pre-image of** b is the set $\{a \in A | f(a) = b\}$.*
  - This sentence tells us the term that we are defining. For some particular $b \in B$ the pre-image of b is defined to be the set of all elements of A that have b as their image.
- *We denote this set as $f^{-1}(b)$*
  - This sentence gives us a piece of notation to use for our new definition. When we see the symbols $f^{-1}(b)$ we should say "the pre-image of b"

There is nothing in the definition of function that requires that every element of the codomain be the image of some element of the domain. Thus we can consider the set of elements of the codomain that appear as the image of some element of the domain.

Definition 0.12. *Let A and B be sets and let f be a function from A to B. The **range of** f, denoted $range(f)$, is the set of elements of B that are the image of some element of A. That is, we have*

$$range(f) = \{f(a) \in B | a \in A\}$$

We should understand the range of a function to be the set of all elements of the codomain that appear as an image of an element in domain. In the context of linear algebra, sometimes range is referred to as **image**.

Consider all of the numbers on the number line. We call this set $\mathbf{R}$. Some of these numbers have interesting properties. For example, the set of integers , $\mathbf{Z}$, is the subset of the set of real numbers consisting of those real numbers that have no decimal part. The set of rational numbers, $\mathbf{Q}$, is the subset of the real numbers that can be represented as a ratio of two numbers in $\mathbf{Z}$. Much as we can consider sets of real numbers that have notable properties, so too can be think about particular functions that satisfy notable properties.

Let $A = \{1, 4, 9\}$ and $B = \{1, 7\}$. There are many different functions from A to B. Some of these functions have their range equal to their codomain. For example, the function from A to B given by

$$\{(1, 7), (4, 1), (9, 1)\}$$

has codomain equal to $\{1, 7\}$ and range equal to $\{1, 7\}$.

Whereas the function

$$\{(1,7),(4,7),(9,7)\}$$

does not have this property.

Definition 0.13. *Let A and B be sets and let f be a function from A to B. We say f* ***is a surjection*** *when $range(f) = B$. When f is an surjection we say that f* ***is surjective***

Intuitively, we should understand that a function to be a surjection when every element of the codomain has a pre-image.

Consider the function $f : \mathbf{R} \to \mathbf{R}$ so that $f(x) = 2x+1$. This function has the extra property that the image of each element of the domain is unique; no two elements of the domain have the same image. That is, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$.

On the other hand, the function $f : \mathbf{R} \to \mathbf{R}$ so that $f(x) = x^2$ does not have this property; we can notice $f(-7) = f(7)$.

Definition 0.14. *Let A and B be sets and let f be a function from A to B. We say f* ***is an injection*** *when each element of the domain has a unique image. When f is an injection we say that f* ***is injective***

Intuitively, we should understand a function to be an injection when for $x_1 \neq x_2$ we have $f(x_1) \neq f(x_2)$. Notice that if a function is an injection, then the preimage of every element of the codomain contains at most one element.

Functions that are both injective and surjective are particularly interesting:

Definition 0.15. *Let A and B be sets and let f be a function from A to B. We say f* ***is a bijection*** *when f is both an injection and a surjection. When f is an bijection we say that f* ***is bijective***

When a function is a surjection, every element of the codomain has at least one pre-image When a function is a injection, every element of the codomain has at most of one pre-image. Therefore bijections are those functions for which every element of the codomain has exactly one pre-image.

Aside 0.16. *You might recognize the terms* "one-to-one" *and* "onto" *rather than the terms* "injection" *and* "surjection". *These terms are not incorrect to use, but they can cause confusion. When students learn the terms* "one-to-one" *and* "onto" *for* "injection" *and* "surjection" *they are usually taught the term* "one-to-one correspondence" *rather than the term* "bijection". *This choice of terminology can lead to confusion as it is easy to confuse the terms* "one-to-one" *and* "one-to-one correspondence".

The word "surjection" *is particularly interesting. In French the word* "sur" *means* "on top" *or* "above". *In a surjection we can imagine the elements of the domain sitting on top and covering up all of the elements of the codomain.*

0.3 Complex Numbers

Recall our definition of the rational numbers from above

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbf{Z}, b \neq 0 \right\}$$

When we write a rational number, say $\frac{1}{2}$, we are expressing a single quantity. Though the horizontal line in $\frac{1}{2}$ reminds us of division, writing $\frac{1}{2}$ is denoting a single quantity – the value on the number line that is halfway between 0 and 1.

With this thought in our minds, we define the set of complex numbers

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbf{R}\}$$

For example, $3 + 7i$ is a complex number. So is $\frac{1}{2} + (-6)i$. As is $\pi + \pi i$.

Definition 0.17. *For a complex number $z = a + bi$ we say a is the real part of z and b is the imaginary part of z. We denote these parts respectively as $Re(z)$ and $Im(z)$.*

For $z = 3 + 7i$ we have $Re(z) = 3$ and $Im(z) = 7$.

When we write a complex number, say $1 + 2i$, we are expressing a single quantity. Though the plus sign in $1 + 2i$ reminds us of addition, writing $1 + 2i$ is denoting a single quantity.

We define a pair of operations, addition and multiplication for elements of $\mathbf{C}$.

$$(a + bi) + (c + di) = (a + c) + (b + d)i \tag{1}$$
$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i \tag{2}$$

For example

$$(1 + 2i) + (3 + 4i) = (1 + 3) + (2 + 4)i = 4 + 6i$$
$$(1 + 2i) \cdot (3 + 4i) + (1(3) - 2(4)) + (1(4) + 2(3))i = -5 + 10i$$

and

$$(0 + 1i) \cdot (0 + 1i) = (0 - (1)) + (0 + 0)i = 1 + 0i$$

There are lots of + signs hanging out in (1). Let us label them and talk about them one at a time:

$$(1 +_1 2i) +_2 (3 +_3 4i) = (1 +_4 3) +_5 (2 +_6 4)i$$

- $+_1$ is the + we use when we are denoting an element of $\mathbf{C}$.

- $+_2$ is the $+$ we are defining so that we have an agreed upon definition of what it means to add elements of $\mathbf{C}$.
- $+_3$ is the $+$ we use when we are denoting an element of $\mathbf{C}$.
- $+_4$ is the $+$ we use to denote addition of elements of $\mathbf{R}$
- $+_5$ is the $+$ we use when we are denoting an element of $\mathbf{C}$.
- $+_6$ is the $+$ we use to denote addition of elements of $\mathbf{R}$.

This is awful! Why would we use the symbol $+$ to refer to three different things?

Imagine you were teaching someone to add complex numbers without needing for them to really understand what they were doing. (... I suspect such a thing is easy for us to imagine.)

If you showed them

$$(1+2i)+(3+4i)$$

and then said

Collect like terms.

they would dutifully do so

$$\begin{aligned}(1+2i)+(3+4i) &= 1+3+2i+4i\\ &= 4+6i\end{aligned}$$

and arrive at the right answer. Though adding complex numbers looks like we are *collecting like terms*, we aren't in fact doing so. We use the phrase *collecting like terms* to talk about the process of simplifying an expression that contains variables. There are no variables in sight here!

Similarly if you gave them the product

$$(1+2i)\cdot(3+4i)$$

and told them to use the distributive law, collect like terms and set $i^2=-1$, they would find:

$$\begin{aligned}(1+2i)\cdot(3+4i) &= 1(3+4i)+2i(3+4i)\\ &= 3+4i+6i+8i^2\\ &= 3+4i+6i-8\\ &= -5+10i\end{aligned}$$

and arrive at the right answer. Though multiplying complex numbers looks like we are using the distributive law, we aren't in fact doing so. Even though we see a $+$, as discussed above this does not denote addition, this plus sign is part of the way we express a complex number.

It is exactly for this convenience that that the mathematics community collectively uses the notation $a + bi$ to denote a complex number. We could just as easily denote a complex number as (a, b), but choosing to denote a complex number as $a + bi$ allows us to ignore the fact that $a + bi$ is an exotic number thing and just pretend that it is a number and that i^2 denotes -1.

Aside 0.18. *One of the myths of science education is that the things in the curriculum are objectively true. If you took chemistry in secondary school you probably learned about the Bohr-Rutherford model of an atom. A key feature of this model is that electrons are these tiny little particles that orbit around a nucleus. Once we get to upper-year chemistry in university we learn, in fact, that electrons aren't really particles that can be said to be in any particular place. Instead they exist as a probability cloud.*

The Bohr-Rutherford model of the atom is just that – a model. It gives us the information we need to so that we can more easily learn about other concepts in chemistry without being bogged down by the details. Teaching a student about reduction and oxidation would be very difficult without the Bohr-Rutherford Model!

I suspect the same phenomenon occurs when we teach secondary school students about complex numbers. When we teach students about complex numbers for the first time, they are a tool to solve quadratic equations. And so students are taught that they are collecting like terms and using the distributive law when they manipulate complex numbers. Though this isn't completely true, it still leads to the correct outcome.

When we defined sets we took some time to define carefully what it means for a pair of sets to be equal. As we are using the equals sign in our work with complex numbers we should probably take a moment to think about what equality means for elements of $\mathbf{C}$.

Definition 0.19. *Let $z, z' \in \mathbf{C}$ so that $z = a + bi$ and $z' = c + di$. We write $z = z'$ when $a = c$ and $b = d$.*

Aside 0.20. *How does this differ to the meaning of equality for elements of $\mathbf{Q}$? It is certainly possible to have $\frac{a}{b} = \frac{c}{d}$, but have $a \neq b$ and $c \neq d$.*

Let us take some time to get more familiar with our definitions for multiplication and addition for elements of $\mathbf{C}$.

For $a + bi \in \mathbf{C}$ consider the product $(a + bi) \cdot (1 + 0i)$.

$$(a + bi) \cdot (1 + 0i) = (a - 0) + (0 - (-b)i = a + bi$$

Multiplication by $1 + 0i$ doesn't change anything! Let $e = 1 + 0i$. For any $z \in \mathbf{C}$ we have

$$z \cdot e = z$$

Looking at our definition for addition we see the same sort of behaviour when we consider the element $0 + 0i$

$$(a + bi) + (0 + 0i) = (a + 0) + (b + 0)i = a + bi$$

Addition by $0 + 0i$ doesn't change anything! Let $f = 0 + 0i$. For any $z \in \mathbf{C}$ we have

$$z + f = z$$

Our new definition of multiplication and addition for elements of $\mathbf{C}$ has some more things in common with our intuition for addition and multiplication of elements in $\mathbf{R}$.

Consider elements of the form $(a + 0i)$. For example we have

$$(6 + 0i) + (4 + 0i) = (6 + 4) + (0 + 0)i = 10 + 0i$$

and

$$(6 + 0i) \cdot (4 + 0i) = (6(4) - 0) + (0 + 0)i = 24 + 0i$$

In general we have

$$(a + 0i) + (b + 0i) = (a + b) + 0i$$

and

$$(a + 0i) \cdot (b + 0i) = ab + 0i$$

Addition and multiplication on elements $a+0i$ and $b+0i$ performs exactly the same function as addition and multiplication of a and b. And so we see that when we restrict to complex numbers z with $Im(z) = 0$, our version of addition and multiplication for elements of $\mathbf{C}$ has the same effect as addition and multiplication in $\mathbf{R}$. We take advantage of this similarity by not writing a real or imaginary part when it is equal to 0. For example, rather than writing $6 + 0i$ we can just write 6. In this same vein, rather than writing $0 + 7i$ we can just write $7i$.

Looking back to one of our examples from above we had:

$$(0 + 1i) \cdot (0 + 1i) = (0 - 1) + (0 + 0)i = -1 + 0i$$

Using our convention of dropping the zero parts of the imaginary number we write:

$$(1i) \cdot (1i) = -1$$

For variables with coefficient 1 we write x in place of $1x$. Adopting the same convention here yields

$$i \cdot i = -1$$

When we multiply a real number by itself we write r^2 rather than $r \cdot r$. Applying this same convention yields

$$i^2 = -1$$

Having $i^2 = -1$ was not part of our definition of complex numbers. It arose as a consequence of our definition of multiplication for elements of $\mathbf{C}$ and adopting conventions for expressing elements of $\mathbf{C}$ that we use in other domains of mathematics.

With our convention that we write elements of the form $a + 0i$ just as a, we then realize we may in fact interpret every real number as a complex number whose imaginary part is 0. And so we have $\mathbf{R} \subsetneq \mathbf{C}$.

Just like elements of $\mathbf{R}$, one can check that elements of $\mathbf{C}$ satisfy the following properties:

- **commutativity**: $\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{C}$.
- **associativity**: $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.
- **identities**: $\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbf{C}$.
- **additive inverse**: for every $\alpha \in \mathbf{C}$ there exists $\beta \in \mathbf{C}$ so that $\alpha + \beta = 0$. We denote β as $-\alpha$
- **multiplicative inverse**: for every $\alpha \in \mathbf{C}$ there exists $\beta \in \mathbf{C}$ so that $\alpha\beta = 1$. We denote β as $\frac{1}{\alpha}$ or α^{-1}.
- **distributive property**: $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

0.4 Vectors and $\mathbf{R}^n$

Above we defined cartesian products of sets. And so without thinking further for every $n \geq 1$ we have:

$$\mathbf{R}^n = \{(r_1, r_2, \ldots, r_n) \mid r_1, r_2, \ldots, r_n \in \mathbf{R}\}$$

However, by equipping elements of $\mathbf{R}^n$ with a notation of addition and scalar multiplication we get a rich algebraic structure. For $x, y \in \mathbf{R}^n$ and $\lambda \in \mathbf{R}$ we define:

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$
$$\lambda x = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$$

When we give elements of $\mathbf{R}^n$ an interpretation as vectors, these two operations correspond to vector addition and scaling. Expressed as a vector, we denote $x \in \mathbf{R}^n$ as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

And so

$$x + y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

and

$$\lambda x = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

Throughout the term we will recall lots of different facts from MATH164. Recalling them all now isn't much use, as these facts are much better recalled in context. And so, for now we will leave here our recollection of facts related to vectors.

---

1 An Introduction to Vector Spaces (LADR 1A, 1B)

Learning Incomes.

- *Familiarity with properties and operations of real and complex numbers from Module 0.*
- *Familiarity with addition and scalar multiplication in* $\mathbf{R}^n$.
- *Recall how one can specify a plane in* $\mathbf{R}^3$ *and check if a vector lies in the plane.*

Learning Outcomes.

- *Understand the meaning and usefulness of the notation* $\mathbf{F}^n$.
- *Understand the meaning of the six properties of a vector space.*
- *Be able to identify when a familiar set of mathematical objects does (or doesn't) satisfy these properties.*

Newly Defined Terms and Notation.

- *list of length* n*, coordinate,* $\mathbf{F}^n$*, addition (in* $\mathbf{F}^n$*), scalar multiplication (in* $\mathbf{F}^n$*), additive inverse (in* $\mathbf{F}^n$*), additive identity (in* $\mathbf{F}^n$*), addition (in* V*), scalar multiplication (in* V*), vector space, real vector space, complex vector space, additive inverse (in* V*), additive identity (in* V*)*

Every module will begin with these three sections. Return to the ***Learning Outcomes*** *and* ***Newly Defined Terms and Notation*** *when are you are done your work on the module. Be sure that your own notes (... you are making your own notes, right?) cover this material.*

Back in Module 0, we briefly reviewed addition and scalar multiplication for vectors from MATH164:

$$x + y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

and

$$\lambda x = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

With a little bit of work, we can verify that these operations satisfy the following properties:

1. **commutativity**: $x + y = y + x$ for all $x, y \in \mathbf{R}^n$
2. **associativity**: $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbf{R}^n$
3. **additive identity**: $x + 0 = x$ for all $x \in \mathbf{R}^n$, where 0 denotes the all zeros vector.
4. **additive inverse**: for all $x \in \mathbf{R}^n$ there exists x' such that $x + x' = 0$. We denote x' as $-x$
5. **multiplicative identity**: $1x = x$ for all $x \in \mathbf{R}^n$
6. **distributivity**: $a(x + y) = ax + by$ and $(a + b)x = ax + bx$ for all $x, y \in \mathbf{R}^n$ and all $a, b \in \mathbf{R}$.

For example, we can show vectors in $\mathbf{R}^n$ are associative as follows:

Proof. Let $x, y, z \in \mathbf{R}^n$. To show addition in $\mathbf{R}^n$ is associative we must show

$$(x + y) + z = x + (y + z)$$

Notice

$$(x+y)+z = \left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \right) + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \tag{3}$$

$$= \begin{bmatrix} (x_1 + y_1) + z_1 \\ (x_2 + y_2) + z_2 \\ \vdots \\ (x_n + y_n) + z_n \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} x_1 + (y_1 + z_1) \\ x_2 + (y_2 + z_2) \\ \vdots \\ x_n + (y_n + z_n) \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \left(\begin{bmatrix} y_1 + z_1 \\ y_2 + z_2 \\ \vdots \\ y_n + z_n \end{bmatrix} \right) \tag{6}$$

$$= x + (y + z) \tag{7}$$

Line (4) follows from line (3) by our definition of addition in $\mathbf{R}^n$. Line (5) follows from line (4) because addition is associative in $\mathbf{R}$. Line (6) follows from line (5) by our definition of addition in $\mathbf{R}^n$

Therefore $(x + y) + z = x + (y + z)$ and so it follows that addition in $\mathbf{R}^n$ is associative. $\square$

Aside. *For each of the six properties we should be able to write down an argument similar to the one above. This is the sort of thing we are talking about when we use the phrase* a little bit of work*.*

When we work with vectors in $\mathbf{R}^n$ we often appeal to these properties, even if we aren't being explicit in doing so. Let us take a moment to examine more closely the **additive inverse** property.

additive inverse: for all $x \in \mathbf{R}^n$ there exists x' such that $x + x' = 0$. We denote x' as $-x$

This property is telling us that for every vector x there exists a vector x' so that $x + x' = 0$. In practice we find the additive inverse of a vector by negating each of its entries. We denote the additive inverse as $-x$.

Additive inverses exist in a number of different contexts. For example, for every matrix A, there exists a matrix A' such that $A + A' = 0$. (Here, 0 denotes the matrix with all zeros.). We find the additive inverse of a matrix by negating all of the entries. We denote the additive inverse as $-A$. Curiously, our other properties (above) also are when we replace vectors in $\mathbf{R}^n$ with matrices!

To wit, let m and n be positive integers and let $\mathcal{M}_{m,n}(\mathbf{R})$ denote the set of $m \times n$ matrices with entries in $\mathbf{R}$. With a little bit of work, we can verify that matrix addition and scalar multiplication satisfy the following properties.

1. **commutativity**: $X + Y = Y + X$ for all $X, Y \in \mathcal{M}_{m,n}(\mathbf{R})$.
2. **associativity**: $(X + Y) + Z = X + (Y + Z)$ for all $X, Y, Z \in \mathcal{M}_{m,n}(\mathbf{R})$.
3. **additive identity**: There exists a matrix, denoted 0, so that $X + 0 = X$ for all $X \in \mathcal{M}_{m,n}(\mathbf{R})$.
4. **additive inverse**: for all $X \in \mathcal{M}_{m,n}(\mathbf{R})$ there exists X' such that $X + X' = 0$. We denote X' as $-X$.
5. **multiplicative identity**: $1X = X$ for all $X \in \mathcal{M}_{m,n}(\mathbf{R})$.
6. **distributivity**: $a(X+Y) = aX+bY$ and $(a+b)X = aX+bX$ for all $X, y \in \mathcal{M}_{m,n}(\mathbf{R})$ and all $a, b \in \mathbf{R}$.

Just like matrices and vectors, we can perform addition and scalar multiplication with polynomials. For $n \geq 0$, let P_n denote the set of all polynomials of degree n with real number coefficients. With a little bit of work, we can verify that polynomial addition and scalar multiplication satisfy the following properties.

1. **commutativity**: $f + g = g + f$ for all $f, g \in P_n$
2. **associativity**: $(f + g) + h = f + (g + z)$ for all $f, g, h \in P_n$
3. **additive identity**: there exists a function, denoted f so that $f + 0 = f$ for all $f \in P_n$

4. **additive inverse**: for all $f \in \mathbf{R}^n$ there exists f' such that $f + f' = 0$. We denote f' as $-f$.
5. **multiplicative identity**: $1f = f$ for all $f \in P_n$
6. **distributivity**: $a(f + g) = af + bg$ and $(a + b)f = af + bf$ for all $f, g \in P_n$ and all $a, b \in \mathbf{R}$

Note: In **additive inverse**, the notation f' is not referring to a derivative. Notation is contextual. In this case we should regard f' as nothing more than a label to refer to a function.

Why are these similarities worth pointing out? In MATH164 you spent an entire semester working with vectors and $\mathbf{R}^n$. Though the course was largely computational, there was some conceptual learning as well. Any conceptual learning you did in MATH164 that depended only on the properties above extends to concepts for matrices, functions, and any other set of mathematical objects that satisfy these six properties.

(That last sentence is rather important. Take a moment to read it again and dwell.)
.
.
.
.
.

In MATH266, rather than think only about vectors in $\mathbf{R}^n$ we will broaden our thinking to all sorts of mathematical objects that satisfy these six properties. We return to this thought in Section 1.2. Before we get to this work, we take a closer look at properties of $\mathbf{R}$ and $\mathbf{C}$.

1.1 Lists of Real Numbers and Complex Numbers (LADR: 1A)

We are so used to the fundamental properties of the real numbers, $\mathbf{R}$, that we invoke them without thinking. Let us take a moment review some properties of the real numbers:

- **commutativity**: $\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{R}$.
- **associativity**: $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{R}$.
- **identities**: $\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbf{R}$.
- **additive inverse**: for every $\alpha \in \mathbf{R}$ there exists $\beta \in \mathbf{R}$ so that $\alpha + \beta = 0$. We denote β as $-\alpha$
- **multiplicative inverse**: for every $\alpha \in \mathbf{R}$ there exists $\beta \in \mathbf{R}$ so that $\alpha\beta = 1$. We denote β as $\frac{1}{\alpha}$ or α^{-1}.
- **distributive property**: $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\alpha, \beta, \lambda \in \mathbf{R}$.

From our work in the previous module, we notice that the set of complex numbers, $\mathbf{C}$, satisfies these same properties. Many of the things we consider in this course will be true regardless of whether we consider $\mathbf{R}$ or $\mathbf{C}$ as our choice of scalar. And so throughout this course we will use the notation $\mathbf{F}$ to refer to $\mathbf{R}$ and $\mathbf{C}$. (This might make much sense right now, but it will by the time you are through with Module 1.)

Aside. *We use the letter* $\mathbf{F}$ *here for the word* field. *Informally, a field is a set in which we can add and multiply and these operations have properties similar to addition and multiplication of real numbers. If you are taking MATH 362 (Rings and Fields) you will learn a lot of about fields this semester!*

In the last module we briefly reviewed Cartesian products of sets. Here our text (LADR) takes a bit of an informal approach to such things. And so for sake of consistency we introduce the notion of a list.

Definition 1.1. *Let $n \geq 0$ be an integer. A* ***list of length*** *n is an ordered collection of n elements. We denote a list of length n as*

$$(x_1, x_2, \ldots, x_n)$$

As a list is a mathematical object, we may refer to it with a label. For example, we may write: let $x = (1, 1, 4)$. The second entry of x is 1 and x has length 3.

Notice that, as compared to a set, the order in which the elements of a list appear matter. The list $(1, 1, 4)$ is different to that of $(1, 4, 1)$. As a further comparison to sets, we notice that lists may have repeated elements.

Aside. *If all of the elements of a list are drawn from the same set, say, A, a list of length n is exactly an element of A^n.*

It is possible you have seen lists in other contexts (For example, as a data structure in Python). A list of length 2 is commonly called a **pair**. And a list of length 3 is commonly called a **triple**. A list of length k is sometimes called a **k-tuple**. For example, one may call a list of length 6 a 6-tuple.

Aside. *The use of the term list in mathematics mostly agrees with its use in common parlance. The main disagreement is when we use* list *as English speakers to talk about a collection of unordered objects – it would certainly be odd to use the phrase* shopping set *(as opposed to* shopping list*.)*

Though lists may contain elements of any type, we will usually concern ourselves with lists of complex or real numbers.

Definition 1.2. *Let $\mathbf{F}^n$ denote the set of all lists of length n of elements of $\mathbf{F}$. We refer the the elements of such a list as **coordinates.** That is, if $x = (x_1, x_2, \ldots, x_n)$ and $j \in \{1, 2, \ldots, n\}$, then we say x_j is the j^{th} **coordinate of** x*

This definition is actually two definitions in one. By letting $\mathbf{F} = \mathbf{R}$ we get a definition for the notation $\mathbf{R}^n$. By letting $\mathbf{F} = \mathbf{C}$ we get a definition for the notation $\mathbf{C}^n$. For example,

$$\mathbf{C}^3 = \{(z_1, z_2, z_3) \mid z_1, z_2, z_3 \in \mathbf{C}\}$$

Just as we did in MATH164 with elements of $\mathbf{R}^n$ we can definite addition for elements of $\mathbf{F}^n$

Definition 1.3. *For $x, y \in \mathbf{F}^n$ we denote by $x + y$ the list:*

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

*We refer to this operation as **addition**.*

For example, in $\mathbf{C}^3$ addition behaves as we expect:

$$\begin{aligned}(1,\ 1+i,\ 7+6i) + (0,\ 2+i,\ 2) &= ((1)+(0),\ (1+i)+(2+i),\ (7+6i)+(2)) \\ &= (1,\ 3+2i,\ 9+6i)\end{aligned}$$

When we think about elements of $\mathbf{R}^n$ as vectors, the definition of addition here corresponds to vector addition. We note, however, that for elements of $\mathbf{C}^n$ we don't have a *physical* interpretation of addition as manipulating the position of vectors in space. (In fact, this same fact is true for $\mathbf{R}^n$ whenever $n \geq 4$). Despite not having this physical interpretation for addition of elements of $\mathbf{C}^n$, our definitions (and not their physical interpretations) enable us to proceed in our exploration. When we want to emphase a geometric interpretation, we'll use vector notation (entries listed vertically with square brackets). Otherwise we will use list notation (entries listed horizontally with parentheses).

Before moving on to talk about some of the properties of addition of elements of $\mathbf{F}^n$, let us return to two familiar properties of $\mathbf{R}$ (and $\mathbf{C}$).

- **additive identity** there exists an element $0 \in \mathbf{F}^n$ so that $x + 0 = x$ for all $x \in \mathbf{F}^n$
- **additive inverse** for all $x \in \mathbf{F}^n$ there exists x' such that $x + x' = 0$. We denote x' as $-x$.

If someone asked us to give a definition of the meaning of the notation -7, an answer likely wouldn't be immediate; we know what -7 means through our experience of working with negative numbers. But, in fact, without us realizing it, this notation has a precise meaning:

Definition. *The notation* $\mathbf{-7}$ *refers to the number* x' *so that* $7 + x' = 0$

Yes, I agree that this is an awkward definition. But taking a moment to reckon with it holds some value. For example, we notice that we can only expect someone to understand what -7 means if they know the meaning of the symbols $+$ and 0. So, now we ask, what is the meaning of the symbol 0?

Definition. *The notation* 0 *refers to the number* z *so that* $y + z = 0$ *for all* $y \in \mathbf{R}$

Intuitively, the number 0 is the number so that we when we perform addition nothing changes.

Just like in $\mathbf{R}$ (and $\mathbf{C}$) we can meaningfully define the notations 0 and $-x$ as elements of $\mathbf{F}^n$.

Definition 1.4. *The* ***additive identity****, denoted* 0*, is the element of* $\mathbf{F}^n$ *so that* $y + 0 = y$ *for all* $y \in \mathbf{F}^n$*.*

For $x \in \mathbf{F}^n$*, the* ***additive inverse of*** x *is the element* $x' \in \mathbf{F}^n$ *so that* $x + x' = 0$*. We denote the additive inverse of* x *as* $-x$*.*

Aside. *Despite the reputation of being precise, much of mathematics communication depends on context. We use the same symbol for* 0 *whether we are referring to* $0 \in \mathbf{R}$*,* $0 \in \mathbf{C}$*,* $0 \in \mathbf{R}^n$

or $0 \in \mathbf{C}^n$*! As communicators using mathematics, the onus is on us to ensure that our reader knows which version of* 0 *we are invoking.*

As opposed to our time spent computing MATH164, we will spend most of our time in MATH 164 providing justification for mathematical facts. Depending on your mathematical journey, you may have had very little exposure to reading and writing mathematical proofs. Don't fret; the part of the learning outcomes for this course is developing tools for reading and writing mathematical proofs. We'll talk more about proofs at various points throughout the course. For now, we should take the following as the meaning of *proof* in this course.

A proof is a justification of a mathematical fact that will convince knowledgeable peers of the truth of the mathematical statement

(In this class, *peers* means students who are in second-year university and have taken MATH 164.)

And so with this in mind, we prove the following fact:

$$0 = (0, 0, \ldots, 0)$$

That is, we prove the the additive inverse in $\mathbf{F}^n$, denoted 0, is in fact the all zeroes vector.

On the left side of this equality is the 0 that is an element of $\mathbf{F}^n$. This piece of notation has a definition (see above.) On the right side of this equality is the element of $\mathbf{F}^n$ where every coordinate is the additive identity of $\mathbf{F}$. To prove that these two things are equal we will show that $(0, 0, \ldots, 0)$ satisfies the definition of the notation 0. That is, we show $(y_1, y_2, \ldots, y_n) + (0, 0, \ldots, 0) = (y_1, y_2, \ldots, y_n)$ for all $y \in F^n$.

Using our definition of addition we have

$$(y_1, y_2, \ldots, y_n) + (0, 0, \ldots, 0) = (y_1 + 0, y_2 + 0, \ldots, y_n + 0)$$

In our list on the far right, the addition is with elements of $\mathbf{F}$. And so $y_i + 0 = y_i$ for all $i \in \{1, 2, \ldots, n\}$. Therefore

$$(y_1, y_2, \ldots, y_n) + (0, 0, \ldots, 0) = (y_1 + 0, y_2 + 0, \ldots, y_n + 0) = (y_1, y_2, \ldots, y_n)$$

Therefore the element $(0, 0, \ldots, 0) \in F^n$ satisfies the definition of 0 and so we may write

$$0 = (0, 0, \ldots, 0)$$

We take these ideas and write them as a theorem with a proof.

Theorem 1.5. *In* $\mathbf{F}^n$ *we have* $0 = (0, 0 \ldots, 0)$.

Proof. We use the definition of $0 \in \mathbf{F}^n$ to show $0 = (0, 0 \ldots, 0)$. That is, we show $y + (0, 0, \ldots 0) = y$ for all $y \in \mathbf{F}^n$. Consider $y \in \mathbf{F}^n$ so that $y = (y_1, y_2, \ldots, y_n)$. We have

$$\begin{aligned} y + (0, 0, \ldots 0) &= (y_1, y_2, \ldots, y_n) + (0, 0, \ldots 0) \\ &= (y_1 + 0, y_2 + 0, \ldots, y_n + 0) \\ &= (y_1, y_2, \ldots, y_n) \end{aligned}$$

Therefore the element $(0, 0, \ldots, 0) \in F^n$ satisfies the definition of 0 and so we may write

$$0 = (0, 0, \ldots, 0)$$

□

Aside. *Take a moment to convince yourself that this proof satisfies the meaning of the word proof above. Are you convinced by it? This mathematical fact isn't terribly deep. But "simple" facts are a good proving ground (ha!) for us to get comfortable with the idea of mathematical proof.*

If you are following along with the textbook, you may notice a slightly different approach. In the text, the author defines the notation 0 *to mean* $(0, 0, \ldots 0)$ *and then shows that it is the additive identity. Here we take the opposite approach and define* 0 *to be the additive identity, and then show that* $(0, 0, \ldots 0)$ *satisfies the additive identity. The difference in approach between the notes and the text is a matter of preference; neither is wrong.*

We finish our work in this section by defining scalar multiplication for elements of $\mathbf{F}^n$. This definition is the same one that we used in MATH 164 for vectors.

Definition 1.6. *Let* $\lambda \in F$ *and let* $x \in F^n$. ***Scalar multiplication****, denoted* λx*, is given by*

$$\lambda x = (\lambda x_1, \lambda x_2, \ldots \lambda x_n)$$

1.1 Test Your Understanding

1. For each element of $\mathbf{C}^3$ below, find the additive inverse.

   (a) $(1, 1+i, 2+2i)$

   (b) $(0, 0, i)$

   (c) $(4i, 1+2i, 1+i)$

2. Fill in the blanks in the following proof:

   Theorem. *Let $n \geq 1$ be an integer. For every $v \in \mathbf{C}^n$ we have*

   $$-v = (-v_1, -v_2, \ldots, -v_n)$$

   Proof. Let $n \geq 1$ be an integer and let $v \in \mathbf{C}^n$. To show $-v = (-v_1, -v_2, \ldots, -v_n)$ we must show

   $$(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = (0, 0, \ldots, 0)$$

   By the definition of addition in $\mathbf{C}^n$ we have:

   $$(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = (v_1 + (-v_1), v_2 + (-v_2), \ldots, v_n + (-v_n))$$

   By the definition of additive inverse in $\mathbf{C}$ we have $v_i + (-v_i) =$ ____ for each $i \in \{1, 2, 3, \ldots, n\}$. Therefore

   $$(v_1 + (-v_1), v_2 + (-v_2), \ldots, v_n + -(v_n)) = (\_\_\_\_, \_\_\_\_, \ldots, \_\_\_\_)$$

   And so, it follows that$(v_1, v_2, \ldots, v_n)+(-v_1, -v_2, \ldots, -v_n) = 0$. Therefore $(-v_1, -v_2, \ldots, -v_n)$ is the additive inverse of v. And so we have

   ____ = _____

   □

1.1 Test Your Understanding Solution

1. (a) $(-1, -1-i, -2-2i)$

   (b) $(0, 0, -i)$

   (c) $(-4i, -1-2i, -1-i)$

2. **Theorem.** *Let $n \geq 1$ be an integer. For every $v \in C^n$ we have*

$$-v = (-v_1, -v_2, \ldots, -v_n)$$

Proof. Let $n \geq 1$ be an integer and let $v \in C^n$. To show $-v = (-v_1, -v_2, \ldots, -v_n)$ we must show

$$(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = (0, 0, \ldots, 0)$$

By the definition of addition in $\mathbf{C}^n$ we have:

$$(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = (v_1 + (-v_1), v_2 + (-v_2), \ldots, v_n + (-v_n))$$

By the definition of additive inverse in $\mathbf{C}$ we have $v_i + (-v_i) = \underline{0}$ for each $i \in \{1, 2, 3, \ldots, n\}$. Therefore

$$(v_1 + (-v_1), v_2 + (-v_2), \ldots, v_n + -(v_n)) = (\underline{0}, \underline{0}, \ldots, \underline{0})$$

And so, it follows that $(v_1, v_2, \ldots, v_n) + (-v_1, -v_2, \ldots, -v_n) = 0$. Therefore $(-v_1, -v_2, \ldots, -v_n)$ is the additive inverse of v. And so we have

$$\underline{-v} = \underline{(-v_1, -v_2, \ldots, -v_n)}$$

□

1.2 Defining Vector Spaces (LADR:1B)

On to the main matter of this course – vector spaces!

In the previous section we introduced a piece of notation, $\mathbf{F}$, that let us talk about both $\mathbf{R}$ and $\mathbf{C}$ at the same time. In the spirit of our discussion from the introduction to this module, let us do the same for mathematical objects that satisfy our six properties. Before we do so, we need to be a little bit careful about what we mean by **addition** and **scalar multiplication**.

Definition 1.7. *Let V be a set.*

- *An **addition** on V is a function that assigns an element $u + v \in V$ to each pair of elements in V.*
- *A **scalar multiplication** on V is a function that assigns an element $\lambda v \in F$ to each $v \in V$*

For example, let $M_{2,2}(\mathbf{F})$ be the set of 2×2 matrices with entires in $\mathbf{F}$. We add a pair of matrices as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} u & v \\ x & y \end{bmatrix} = \begin{bmatrix} a+u & b+v \\ c+x & d+y \end{bmatrix}$$

Notice that we have $\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} u & v \\ x & y \end{bmatrix} \in M_{2,2}(\mathbf{F})$ since $\begin{bmatrix} a+u & b+v \\ c+x & d+y \end{bmatrix} \in M_{2,2}(\mathbf{F})$.

Having the result of addition of a pair of elements in V be an element of V is an important feature of addition (and scalar multiplication). In this context, we say that the operation is **closed**. Otherwise, we say the operation is **not closed.**

For example, the set of vectors in $\mathbf{R}^2$ whose first entry is 0 is closed with respect to addition; adding together any such pair of vectors results in a vector that again has 0 as its first entry.

On the other hand, the set of vectors in $\mathbf{R}^2$ whose first entry is 1 is not closed with respect to addition. Let A be the set of set of vectors in $\mathbf{R}^2$ whose first entry is 1. We have $(1, 1) \in A$ and $(1, 2) \in A$, but $(1, 1) + (1, 2) \notin A$

With these two definitions and our intuition from the introduction we are ready to define a **vector space**.

Definition 1.8. *A **vector space** is a set V together with an addition and a scalar multiplication, both of which are closed, so that the following properties hold:*

1. ***commutativity**: $x + y = y + x$ for all $x, y \in V$*
2. ***associativity**: $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$*

3. ***additive identity***: *there exists an element* $0 \in V$ *so that* $x + 0 = x$ *for all* $x \in V$
4. ***additive inverse***: *for all* $x \in V$ *there exists* x' *such that* $x + x' = 0$. *We denote* x' *as* $-x$.
5. ***multiplicative identity***: $1x = x$ *for all* $x \in V$
6. ***distributivity***: $a(x + y) = ax + by$ *and* $(a + b)x = ax + bx$ *for all* $x, y \in V$ *and all* $a, b \in \mathbf{F}$

We call the elements of a vector space ***vectors*** *or* ***points***.

Even though we are seeing vector spaces for the first time, we already know lots of examples of vector spaces! From our discussion in Section 1, we have

- $\mathbf{F}^n$ is a vector space for every $n \geq 1$
- $\mathcal{M}_{m,n}(\mathbf{F})$ is a vector space of every $m, n \geq 1$
- P_n is a vector space for every $n \geq 1$

Let us consider another example of a vector space – a plane in $\mathbf{R}^3$ through the origin.

Example 1.9. *Let* P *be the set of solutions to*

$$x_1 + 3x_2 + x_3 = 0$$

The set P *defines a plane in* $\mathbf{R}^3$. *We claim that* P *is a vector space with respect to vector addition and scalar multiplication in* $\mathbf{R}^3$.

To show P *is a vector space we must show:*

- *vector addition and scalar multiplication are closed on* P.
- P *satisfies each of the six parts of the definition of vector space.*

To show vector addition is closed on P *we must show that the sum of any pair of vectors in* P *results in a vector that is also in* P. *Consider* $u, u' \in P$. *Since* $u, u' \in P$, *we have*

$$u_1 + 3u_2 + u_3 = 0$$
$$u'_1 + 3u'_2 + u'_3 = 0$$

Consider the vector $u + u' = (u_1 + u'_1, u_2 + u'_2, u_3 + u'_3)$. *Notice*

$$\begin{aligned}(u_1 + u'_1) + 3(u_2 + u'_2) + (u_3 + u'_3) &= (u_1 + 3u_2 + u_3) + (u'_1 + 3u'_2 + u'_3) \\ &= 0 + 0 \\ &= 0\end{aligned}$$

Therefore $u + u' \in P$. Therefore vector addition is closed on P.

To show scalar multiplication closed on P we must show that the $\lambda v \in P$ for each $\lambda \in \mathbf{R}$ and each $v \in P$. If $v \in P$, then

$$v_1 + 3v_2 + v_3 = 0$$

Consider $\lambda v = (\lambda v_1, \lambda v_2, \lambda v_3)$. We have

$$\lambda v_1 + 3(\lambda v_2) + \lambda v_3 = \lambda(v_1 + 3v_2 + v_3) = \lambda(0) = 0$$

Therefore $\lambda v \in P$. Therefore scalar multiplication is closed on P.

We now turn to show that P satisfies each of the six properties of a vector space. Notice that addition and scalar multiplication in P are the same as addition and scalar multiplication in $\mathbf{R}^3$. Let $u, v, w \in P$ and let $a, b \in \mathbf{R}$

1. ***commutativity:*** *Since $u, v \in \mathbf{R}^3$ and $\mathbf{R}^3$ is a vector space we have $u + v = u + v$.*
2. ***associativity:*** *Since $u, v, w \in \mathbf{R}^3$ and $\mathbf{R}^3$ is a vector space we have $(u + v) + w = u + (v + w)$.*
3. ***additive identity:*** *Since $0 + 3(0) + 0 = 0$ we have $0 = (0, 0, 0) \in P$.*
4. ***additive inverse:*** *For $u \in P$ we have $u_1 + 3u_2 + u_3 = 0$. Therefore $-u_1 + 3(-u_2) + (-u_3) = -(u_1 + 3u_2 + u_3) = -(0) = 0$. And so $-u = (-u_1, -u_2, -u_3) \in P$.*
5. ***multiplicative identity:*** *Since $u \in \mathbf{R}^3$ and $1 \in \mathbf{R}$, we have $1u = u$.*
6. ***distributivity:*** *Since $u, v, w \in \mathbf{R}^3$, $a, b \in \mathbf{R}$ and $\mathbf{R}^3$ is a vector space, we have $a(u + v) = au + bv$ and $(a + b)u = au + bu$.*

Since P satisfies each of the six properties of a vector space, P is a vector space.

Aside 1.10. *That was a lot of work! But, this is what is required to show that a particular set of mathematical objects is a vector space. A good question to ask is* how did we know this was the right thing to do to *show* that P is a vector space.*. The answer to this question, unfortunately, is experience. However, every time we want to* show *something, our broad strategy will be the same: write an explanation that convinces the reader that the object we are considering satisfies the desired property*

From our work above, it seems that every time we have addition and scalar multiplication, we end up having a vector space. Though this will often be the case in this class, it is not always true. For example, let $\mathcal{I}_2$ be the set of invertible 2×2 matrices. Though we have a notion of addition and scalar multiplication for these objects, our standard matrix addition

is not closed on $\mathcal{I}_2$ – it is possible to find a $A, B \in \mathcal{I}_2$ so that $A + B \notin \mathcal{I}_2$. (We'll return to this on assignment 1.) Therefore $\mathcal{I}_2$ is not a vector space.

Our choice of scalars can also lead to situations where we don't have a vector space. For example, let P_2 be the set of degree two polynomials with real coefficients. If choose our scalars to be elements of $\mathbf{C}$, then there is no guarantee that $\lambda f \in P_2$ for all $\lambda \in \mathbf{C}$. For example

$$2 + 2i(x^2 + x + 1) = (2 + 2i)x^2 + (2 + 2i)x + (2 + 2i)$$

This quadratic polynomial does not have real coefficients and so is not an element of P_2. Therefore P_2 is not a vector space when we choose our scalars to be elements of $\mathbf{C}$. On the other hand, as we saw above P_2 is a vector space when we choose our scalars to be in $\mathbf{R}$. For this reason we define the following terms:

Definition 1.11. *A vector space with scalars in* $\mathbf{R}$ *is called a* ***real vector space****.*
A vector space with scalars in $\mathbf{C}$ *is called a* ***complex vector space****.*

From our work above, P_2 is not a complex vector space, but it is a real vector space.

As we did in the previous section, let us turn for a moment to examine the properties of **additive identity** and **additive inverse** more carefully.

- **additive identity** there exists an element $0 \in V$ so that $x + 0 = x$ for all $x \in V$
- **additive inverse** for all $x \in V$ there exists x' such that $x + x' = 0$. We denote x' as $-x$

Just as had with real numbers, the notation $\mathbf{0}$ refers to the element of V so that $x + 0 = x$ for each $x \in V$. Similarly, the notation $-\mathbf{x}$ refers to the element of $x' \in V$ so that $x + x' = 0$. The existence of an additive inverse, $-x$, is predicated upon the existence of the additive identity, 0.

For vectors in $\mathbf{R}^n$, we have $-x = (-x_1, -x_2, \ldots, -x_n)$. Similarly, to find the additive inverse of a matrix, we negate each of its entries. However, how do we know these are the only additive identities? That is, how do we know that the additive identity is unique.

Consider for a moment square roots in $\mathbf{R}$. We need to be a little but careful when we use the notation $\sqrt{x}$ as there may be two solutions – square roots are not unique! For example, there exists $r, s \in \mathbf{R}$ so that $r^2 = s^2 = 4$ and $r \neq s$. The number 4 has two square roots.

Our experience with additive inverses, tells us that this isn't the case with additive inverses – there is only one number we can add to 7 to get 0. Despite the fact that our experience tells us otherwise, our definition of vector space does not require that additive inverses are unique. (Take a moment to confirm this. Turn back to the page with the definition of vector space and notice that the word *unique* does not appear under **additive inverse**. Go ahead. We'll wait) .

.
.
.
.
.
.

That uniqueness does not appear as part of the definitions suggest that either there exist vector spaces in which additive inverse are not unique (like square roots in $\mathbf{R}$) or we may be able to prove that additive inverses are unique in every vector space. It turns out that the latter is true.

Let V be a vector space and let $v \in V$. We show that if v' and v'' both satisfy the definition of additive inverse of v, then $v' = v''$.

If v' and v'' both satisfy the definition of additive inverse of v, then $v + v' = 0$ and $v + v'' = 0$ Therefore

$$v' = v' + 0 = v' + (u + v'') = (v' + u) + v'' = 0 + v'' = v'' \tag{8}$$

Thus if v' and v'' both satisfy the definition of additive inverse of v, then $v' = v''$. Therefore additive inverses are unique in V.

Aside 1.12. *Notice how if we replace V with, say, $\mathbf{R}^4$ the proof is exactly the same. Similarly, the proof is the same if we replace V with, say, $\mathcal{M}_{2,3}(\mathbf{R})$. Rather than have to prove our fact for any individual vector space, our proof simultaneously works for every vector space.*

With our proof that additive inverse are unique in a vector space, we are assured that our notation $-v$ has precise meaning. For a vector space V and a point $v \in V$, the notation $-v$ refers to the unique element $v' \in V$ so that $v + v' = 0$.

Aside. *This argument tells us that additive inverses are unique in vector spaces. And so we wonder, why does it not appear as part of the definition? We do not include properties we can prove as part of the definition because doing so would be redundant. If we can use the six properties to prove that vector spaces have further properties, we do not need to include these further properties are part of the definition. We'll return to this idea later in the course.*

1.2 Test Your Understanding

1. For each equals sign in (8), determine which of the six vector space property implies that the left side is equal to the right side. For example, we have $v' = v' + 0$ by **additive identity** property.

2. Let **E** and **O** respectively be the set of even and odd integers. Explain (with a sentence or two) how you know **E** is closed with respect to addition. Give an example that shows **O** is not a closed with respect to addition.

3. Let P be the set of solutions to

$$x_1 + 3x_2 + x_3 = 1$$

   The set P defines a plane in $\mathbf{R}^3$. Explain how you know that P is not a vector space (hint: look at the six required properties of a vector space. Is there one that is not satisfied for P?)

4. Explain how you know that $\mathbf{R}^n$ is not a complex vector space.

1.2 Test Your Understanding Solutions

1.
   - *We have* $v' = v' + 0$ *by* ***additive identity****.*
   - *We have* $v' + 0 = v' + (u + v'')$ *by* ***additive inverse****.*
   - *We have* $v' + (u + v'') = (v' + u) + v''$ *by* ***associativity****.*
   - *We have* $(v' + u) + v'' = 0 + v''$ *by* ***additive inverse****.*
   - *We have* $0 + v'' = v''$ *by* ***additive inverse****.*
2. *The sum of two even numbers is even. Therefore* $m + n \in \mathbf{E}$ *for every* $m, n \in \mathbf{E}$*. However, since* $1 \in \mathbf{O}$ *and* $1 + 1 \notin \mathbf{O}$*, then* $\mathbf{O}$*, then* $\mathbf{O}$ *is not closed with respect to addition.*
3. *Since* $0 + 3(0) + 0 \neq 1$ *we have* $0 \notin P$*. Therefore* P *does not have an additive identity. And so* P *is not a vector space.*
4. *We can find* $\lambda \in \mathbf{C}$ *and* $v \in \mathbf{R}^n$ *so that* $\lambda v \notin \mathbf{R}^n$*. For example* $(1 + i)(1, 1, 1, \ldots, 1) = (1 + i, 1 + i, \ldots, 1 + i)$ *and* $(1 + i, 1 + i, \ldots, 1 + i) \notin \mathbf{R}^n$*. Therefore scalar multiplication is not closed on* $\mathbf{R}^n$ *when our scalars are in* $\mathbf{C}$*. And so* $\mathbf{R}^n$ *is not a complex vector space.*

1.3 The Mathematical Canon and Linear Algebra

A good conversation to have at the start of any class (in mathematics or otherwise) is one of motivation for the course material. Sometimes mathematics courses get away with a hand wave towards *real-world* application or *improving reasoning skills*, but I think we can do a little bit better here.

We will spend most of our time in MATH 266 muddling with abstract nonsense. We'll prove theorems about abstract objects without any clear or obvious application to industry or other sciences. This isn't to say that the work we do doesn't have applications in these areas – in fact the opposite is true. However, applying the material in this course to other domains requires background knowledge that we don't have. And so, at some point, you may find yourself thinking *why am I learning this?*. From the point of view of the modern higher education in the sciences (we go to school to train for a profession), it can be hard to understand why a course in abstract mathematical theory should appear as part of a degree, especially for those of you who are not pursuing a degree in mathematics.

The College of Arts and Sciences at the University of Saskatchewan espouses a model of *liberal arts education*. The idea is that by studying material in a variety of disciplines (including humanities classes!) we can become more culturally literate members of our society. A university education isn't the only means of attaining this literacy, and indeed one can argue that at times it isn't a particularly effective means to this goal.

For those of you who are unlikely to pursue further courses in mathematics, physics or computer science, your learning in this course is unlikely to provide you with additional training to *do* anything in your (eventual) profession. For you, this course is part of your broader education in becoming a more culturally literate member of society. By immersing yourself in the course material you are learning what it is like to be a mathematician – to grapple with abstract ideas that may only exist in our collective human consciousness.

The beauty in both application and abstraction gives linear algebra a special place in the mathematical canon – the collection of mathematical topics that the community deems necessary to be a fully literate mathematician. Like the concept of canon in literature, the mathematical canon is challenged to discard outdated work from a single perspective, and include more modern work from a variety of perspectives. The canon is constantly changing, being updated and the inclusion of particular topics is highly subjective. For example, courses in graph theory, combinatorics and number theory were rare in mathematics departments fifty years ago. Whereas now these topics are considered (by some) to be essential learning for mathematicians.

As part of your work in Module 0, you were asked to read the *Preface for the Student* in our course text, Linear Algebra Done Right (LADR). As we finish our work for the week, I want to take a moment to dwell on one passage:

... you are about to immerse yourself in serious[1] *mathematics, with an emphasis on attaining a deep understanding of the definitions, theorems, and proofs.*

Our primary goal in this course is understanding. More often than not, we will show our understanding though written explanations in complete sentences. As we progress though the course, we will talk more about what is means to *explain* as a mathematician. For now, the key thing to remember is that defined terms have meanings – they are synonyms for precise ideas. Most of time, our difficulty in understanding (and subsequently, being able to explain) stems from an incomplete understanding of definitions.

But this is enough talking *about* mathematics for now. And so we return to talking mathematics in Module 2.

---

[1] The word *serious* here should not be taken to be a synonym for *frightening*.

2 Subspaces, Spans and Linear Independence (LADR 1C,2A)

Learning Incomes.

- *Recall the ideas of span and linear independence from MATH164*
- *Be able to find the general solution to a system of linear equations*
- *Recall that a line or a plane through the origin can be expressed as the span of a set of vectors.*
- *Recall the meaning of the term* ***standard basis vectors*** *for* $\mathbf{R}^n$.

Learning Outcomes.

- *Understand the definition of* ***subspace, span, linear independence, linear dependence, finite dimensional vector space*** *and* ***infinite dimensional vector space***
- *Be able to give an example that satisfies each of new concepts listed in the previous point.*
- *Understand the steps to check if a subset is a subspace*
- *Understand the statement and proof of Theorem 2.5*
- *Use MATH164 techniques to be able to determine if a set of points in an arbitrary vector space is linearly independent.*

Newly Defined Terms and Notation.

- *subspace, span, linear independence, linear dependence, finite dimensional vector space, infinite dimension vector space*

Recall the example from the previous module where we showed that the set P of solutions to $x_1 + 3x_2 + x_3 = 0$ formed a vector space. In showing that the set of vectors formed a vector space, we leaned heavily on the fact that $\mathbf{R}^3$ was a vector space. With our physical interpretation of vectors in $\mathbf{R}^3$ we recall that the set of solutions to $x_1 + 3x_2 + x_3 = 0$ forms a plane that passes through the origin. And from our work in the previous module, we see that this plane is a vector space. In some sense, the vector space P is inside of the vector space $\mathbf{R}^3$. However, we also saw that the set of solutions to $x_1 + 3x_2 + x_3 = 1$ did not form a vector space.

Just like sets can be nested inside of other sets, so too can vector spaces be nested inside of other vector space. Every non-empty set necessarily has at least two subsets: the empty set and the set itself. However, right away we can see that not every subset of a vector space

forms another vector space: The empty set is a subset, but it cannot be a vector space – it does not contain an additive identity.

In Section 2.1 we begin to consider the question: for a vector space V which subsets of V are vector spaces? The work brings to a close our work in Chapter 1 in *Linear Algebra Done Right*.

In Chapter 2 of *Linear Algebra Done Right* we return to some familiar territory from MATH164: linear combinations, linear independence, span and bases. Recall the following two definitions from MATH164:

Definition 2.1. *Let $n, k \geq 1$ be an integer and let $S \subset \mathbf{R}^n$ so that*

$$S = \{v_1, v_2, \ldots, v_k\}$$

*The **span of** S, denoted $span(S)$ is the set of all linear combinations of elements of S. That is*

$$span(S) = \{a_1v_1 + a_2v_2 + \cdots + a_kv_k \mid a_1, a_2, \ldots, a_k \in \mathbf{R}\}$$

*We say S **spans** $\mathbf{R}^n$ when $span(S) = \mathbf{R}^n$.*

Definition 2.2. *Let $n, k \geq 1$ be an integer and let $S \subset \mathbf{R}^n$ so that*

$$S = \{v_1, v_2, \ldots, v_k\}$$

*We say S **is linearly independent** when the only solution to*

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$$

is $a_1 = a_2 = a_3 = \cdots = a_k = 0$.

Thinking about linear combinations in $\mathbf{R}^n$ (and subsequently, linear independence, spans and bases) requires us to be able to multiply a vector by a scalar and to be able to add together vectors. As we are able to do such operations in any vector space, it suggests then that we may study the linear combinations, linear independence, span and basis for arbitrary vector spaces. We begin to undertake this work in Section 2.2

2.1 Subspaces (LADR 1C)

To think about vector spaces contained inside other vector spaces we will need a piece of new vocabulary to refer to such vector spaces.

Definition 2.3. *Let V be a vector space and let U be a subset of V. We say U* ***is a subspace of*** *V when U is a vector space.*

Let $\mathcal{M}_{3\times 3}(\mathbf{R})$ be the set of 3×3 matrices with entries in $\mathbf{R}$. Let S be the set of upper triangular matrices. That is

$$S = \left\{ \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix} \mid a, b, c, d, e, f \in \mathbf{R} \right\}$$

Knowing that $\mathcal{M}_{3\times 3}(\mathbf{R})$ is a real vector space lets us see what work we need to do to prove that S is subspace of $\mathcal{M}_{3\times 3}(\mathbf{R})$. Looking at the definition above, S is a subspace of V when S is a vector space. And so, to show that S is a subspace of V we must show that S is a vector space. To do this, we return to the definition of vector space and justify that S satisfies each of the parts of the definition. Take a moment to recall the definition of a vector space.

.
.
.
.
.
.
.
.
.

(Perhaps you have copied this definition in to your own notes that you are making for each module.)

.
.
.
.
.

(This vertical space is here for a reason. If this definition isn't in your own notes, take a moment to look back at the Module 1 notes to find the definition of a vector space.)

.
.
.
.

To show S is a vector space we must show that S is closed with respect to addition and scalar multiplication, and S satisfies the six necessary properties.

Let A and B be elements of S. We add matrices of the same dimension by adding together corresponding entries. And so we see that $A + B$ is an upper-triangular matrix and so $A + B \in S$. Therefore S is closed with respect to addition.

Let $\lambda \in \mathbf{R}$. We scalar multiply by multiplying each entry by the scalar. And so we see that λA is an upper-triangular matrix and so $\lambda A \in S$. Therefore S is closed with respect to scalar multiplication.

We turn now to our six required properties. Let $X, Y, Z \in S$ and let $a, b \in \mathbf{R}$.

1. **commutativity:** Since $X, Y \in \mathbf{R}^3$ and $\mathcal{M}_{3\times 3}(\mathbf{R})$ is a vector space we have $X + Y = Y + X$.
2. **associativity:** Since $X, Y, Z \in \mathcal{M}_{3\times 3}(\mathbf{R})$ and $\mathcal{M}_{3\times 3}(\mathbf{R})$ is a vector space we have $(X + Y) + Z = X + (Y + Z)$.
3. **additive identity:** Since $0X = 0$ in $\mathcal{M}_{3\times 3}(\mathbf{R})$ and S is closed with respect to scalar multiplication, we have $0 \in S$.
4. **additive inverse:** Since $(-1)X = -X$ in $\mathcal{M}_{3\times 3}(\mathbf{R})$ and S is closed with respect to scalar multiplication, we have $-X \in S$.
5. **multiplicative identity:** Since $X \in \mathcal{M}_{3\times 3}(\mathbf{R})$ and $1 \in \mathbf{R}$, we have $1X = X$.
6. **distributivity:** Since $X, Y \in \mathcal{M}_{3\times 3}(\mathbf{R})$, $a, b \in \mathbf{R}$ and $\mathcal{M}_{3\times 3}(\mathbf{R})$ is a vector space, we have $a(X + Y) = aX + aY$ and $(a + b)X = aX + bX$.

Since S satisfies all of the necessary properties of a vector space and $S \subseteq \mathcal{M}_{3\times 3}(\mathbf{R})$ we have that S is a subspace of $\mathcal{M}_{3\times 3}(\mathbf{R})$.

Looking back at this work, our proofs for many of the six properties didn't actually depend on S being a set of matrices. Statements $1, 2, 5, 6$ were true because $\mathcal{M}_{3\times 3}(\mathbf{R})$ is a vector space and elements of S are elements of $\mathcal{M}_{3\times 3}(\mathbf{R})$. Where as statements $3, 4$ were true because we know what happens when we multiply a matrix by the scalar 0 or -1.

For an arbitrary vector space V and some $U \subset V$, properties $1, 2, 5, 6$ will always be true for U: in some sense U inherits these properties from V. However property 3 may not be satisfied for U – if we take a random subset of V, there is no guarantee we will have 0 in this subset. Similarly, property 4 may not be satisfied for U – if we take a random subset of V there is no guarantee that every element of this subsets will also have its additive inverse as part of the subset.

Looking at our arguments above for 3. and 4. we leaned on scalar multiplication being closed and the following two facts for matrices:

- $0X = 0$

- $(-1)X = -X$

With our familiarity with matrices, I expect we didn't notice anything particularly interesting about these facts. But when we think a little bit more carefully about how we have defined our terms so far, these innocuous statements become a little more pernicious.

$$0X = 0$$

On the left side we have scalar multiplication: $0X$. In this context, $0 \in \mathbf{R}$ is a scalar. On the right side we have 0. In this context, 0 is the matrix with the property that $Y + 0 = Y$ for all matrices Y. From our experience we know that multiplying a matrix by the scalar 0 results in the zero matrix. But, is this fact true for all vector spaces? Let V be a vector space and let $v \in V$. How do we know $0v = 0$?

Looking at the second equation, things become even more cruel:

$$(-1)X = -X$$

On the left side we have scalar multiplication $(-1)X$. On the right side we have an additive inverse $-X$. Recalling our definition for additive inverse, $-X$ is the unique matrix such that $X + (-X) = 0$. From our past experience, we know that multiplying a matrix by the scalar -1 results in the additive inverse for that matrix. But, is this fact true for all vector spaces? Let V be a vector space and let $v \in V$. How do we know $(-1)v = -v$?

Thankfully, both of these facts turn out to be true. We discuss the proofs of these facts on Assignment 1.

Lemma 2.4. *Let V be a vector space. For every $v \in V$ we have $0v = 0$ and $(-1)v = -v$.*

Returning to our main discussion: how can we tell whether or not a subset of elements of a vector space is a subspace? From our discussion above, it seems as if we do not need to check properties $1, 2, 5, 6$. These properties will be true for any subset of any vector space. Further, checking properties 3 and 4 depended on knowing that scalar multiplication is closed for our subset. And so we have:

Theorem 2.5. *Let V be a vector space and let U be a subset of V. We have that U is a subspace of V if and only if U is non-empty and U is closed with respect to addition and scalar multiplication.*

Before we prove this statement, let us take a moment to think about the structure of this theorem:

Let V be a vector space and let U be a subset of V.

This sentence introduces the pieces of notation we need to be able to state our mathematical truth.

U is a subspace of V if and only if U is non-empty and U is closed with respect to addition and scalar multiplication.

This sentence has two mathematical statements, separated by the phrase *if and only if*. The phrase *if and only if* means that the two statements are equivalent. That is, it means:

1. If U is a subspace of V, then U is non-empty and U is closed with respect to addition and scalar multiplication; and
2. if U is non-empty and U is closed with respect to addition and scalar multiplication, then U is a subspace of V.

If and only if theorems are statements about mathematical synonyms. If this theorem is true (and it is!), then the two statements:

U is a subspace of V

and

U non-empty and U is closed with respect to addition and scalar multiplication.

convey the same information; these two statements are equivalent. Thus, to know that a subset is a subspace it suffices to check that it is non-empty and that it is closed with respect to addition and scalar multiplication.

To justify that the theorem is true, we must justify both of 1. and 2. above are true.

Proof. Let V be a vector space and let U be a non-empty subset of V.

1. If U is a subspace of V, then U is a vector space. Since U is vector space it satisfies the definition of a vector space. Since U is a vector space, we have $0 \in U$. Therefore U is non-empty. Since U is vector space, U is closed with respect to addition and scalar multiplication.

2. Assume now U is non-empty and U is closed with respect to addition and scalar multiplication. Since U is closed with respect to addition and scalar multiplication, to show U is vector space we must verify that it satisfies the six parts of the definition of vector space.

Let $x, y, z \in U$ and let $a, b \in \mathbf{F}$. Since U is a subset of V, we have that $x, y, z \in V$

1. **commutativity**: Since $x, y \in V$ and V is a vector space, we have $x + y = y + x$.

2. **associativity**: Since $x, y, z \in V$ and V is a vector space, we have $(x+y)+z = x+(y+z)$.
3. **additive identity**: Since U is non-empty, there exists at least one element of U. By Lemma 2.4 scalar multiplying this element by 0 yields 0. Since U is closed with respect to scalar multiplication we have $0 \in U$.
4. **additive inverse**: Since U is closed with respect to scalar multiplication, we have $(-1)x \in U$. By Lemma 2.4, we have $(-1)x = -x$. And so $-x \in U$.
5. **multiplicative identity**: Since V is a vector space we have $1x = x$.
6. **distributivity**: Since $x, y \in V$ and V is a vector space we have $a(x + y) = ax + by$ and $(a + b)x = ax + bx$.

Since U satisfies the six properties of a vector space, U is a vector space. Therefore U is a subspace of V. □

Aside. *Our text takes a slightly different approach to this result. The corresponding result, 1.34 on page 18, is equivalent to the statement of Theorem 2.5.*

To see how this theorem simplifies the process of checking if a subset is a subspace, let us take a moment to recall a slightly more exotic subspace from our work in MATH164: the null space.

Let $n \geq 1$ be an integer and let A be an $n \times n$ matrix with entries in $\mathbf{R}$. Let

$$N = \{v \mid Av = 0\}$$

That is, N is the set of vectors $v \in \mathbf{R}^n$ so that $Av = 0$. By applying Theorem 2.5, we can check that N is a subspace of $\mathbf{R}^n$ To do so, we must show:

1. N is non-empty; and
2. N is closed with respect to addition and scalar multiplication.

For $v = 0$ we have $Av = 0$. Therefore $0 \in N$. Therefore N is non-empty.

Consider $u, v \in N$. Since $u, v \in N$ we have $Au = 0$ and $Av = 0$. And so

$$A(u + v) = Au + Av = 0 + 0 = 0$$

Therefore $u + v \in N$. And so N is closed with respect to addition.

Consider $\lambda \in \mathbf{R}$. Notice $A(\lambda u) = \lambda(Au) = \lambda 0 = 0$. Therefore $\lambda u \in N$ and so N is closed with respect to scalar multiplication.

Since N satisfies both (1) and (2), by Theorem 2.5 we have that N is a subspace of $\mathbf{R}^n$.

We conclude our first look subspaces by thinking about the subspaces of the real vector space $\mathbf{R}^3$. Trivially, $\mathbf{R}^3$ is a subspace of itself as $\mathbf{R}^3 \subseteq \mathbf{R}^3$ and $\mathbf{R}^3$ is a vector space. From

our work in the previous module we should be convinced that any plane in $\mathbf{R}^3$ that passes through the origin is a subspace. The only other *nice* set of points we might think about in $\mathbf{R}^3$ are ones that lay on a common line. From our work in thinking about planes, for a line in $\mathbf{R}^3$ to be a subspace it must pass through the origin. Every line through the origin in $\mathbf{R}^3$ can be represented as the span of a some vector.

$$L_v = span(v) = \{\lambda v \,|\lambda \in \mathbf{R}\}$$

Using Theorem 2.5 we can check that L_v is a subspace, for each $v \in \mathbf{R}^3$. To use Theorem 2.5 check that L_v is a subspace of $\mathbf{R}^n$, we must show:

1. L_v is non-empty and
2. L_v is closed with respect to addition and scalar multiplication.

Elements of L_v are of the form λv where $\lambda \in \mathbf{R}$. Choosing $\lambda = 1$ tells that that $1v = v \in L_v$. Therefore L_v is non-empty.

To check if L_v is closed with respect to addition we must check that for any pair $u, u' \in L_v$ we have $u + u' \in L_v$.

If $u, u' \in L_v$, then there exists scalars λ_u and $\lambda_{u'}$ so that $u = \lambda_u v$ and $u' = \lambda_{u'} v$. Therefore

$$u + u' = \lambda_u v + \lambda_{u'} v = (\lambda_u + \lambda_{u'})v$$

Since $\lambda_u + \lambda_{u'} \in \mathbf{R}$ we have that $u + u' \in L_v$.

To check if L_v is closed with respect to scalar multiplication, we must check that for any $u \in L_v$ and any $\beta \in \mathbf{R}$ we have $\beta u \in L_v$. Since $u \in L_v$, then there exists a scalar λ_u so that $u = \lambda_u v$. Therefore

$$\beta u = \beta(\lambda_u v) = (\beta \lambda_u)v$$

Since $\beta\lambda_u \in \mathbf{R}$ we have that $\beta u \in L_v$.

Since L_v satisfies both 1. and 2. above, by Theorem 2.5 we have that L_v is a subspace of $\mathbf{R}^3$.

Returning to our thought about subspaces of $\mathbf{R}^3$, we see that any plane and any line through the origin gives a subspace of $\mathbf{R}^3$.

Looking at our argument for L_v being a subspace, let us take a moment to consider letting $v = 0$ in this argument.

When we choose $v = 0$ we have

$$L_0 = span(0) = \{\lambda 0 \,|\lambda \in \mathbf{R}\} = \{0 \,|\lambda \in \mathbf{R}\} = \{0\}$$

Following through the rest of the argument, we see that $\{0\}$ is a subspace of $\mathbf{R}^3$.
.
.
.
.
.
.
(Are you sure that you believe this?)

We wonder then, are these all of the subspace of $\mathbf{R}^3$? Perhaps there are others out there that aren't one of the ones we have looked at (the zero vector, lines through the origin, planes through the origin and $\mathbf{R}^3$ itself).

We will return to this question in Module 3 when we think about the idea of basis for a vector space. For now, we notice that each of these objects: the zero vector, a line through the origin, a plane through the origin, and $\mathbf{R}^3$ can each be expressed as the span of some set of vectors in $\mathbf{R}^3$:

1. $\{0\} = span(0)$
2. $L_v = span(v)$
3. $P_{u,v} = span(u, v)$, where u and v are not colinear
4. $\mathbf{R}^3 = span(e_1, e_2, e_3)$, where e_1, e_2 and e_3 the standard basis vectors.

Since our definition of vector space allows us to scalar multiply and add, our notion of span in $\mathbf{R}^n$ can be extended to arbitrary vector spaces to produce subspaces. We consider this idea in the following section.

2.1 Test Your Understanding

1. In the statement of Theorem 2.5, why do we require U to be non-empty? That is, explain why the empty set is not a vector space (and thus not a subspace).
2. Let $\mathcal{P}_3(\mathbf{R})$ be the set of polynomials of degree at most 3 with coefficients in $\mathbf{R}$. Is the set
$$S = \{f \mid f(0) = 1\}$$
a subspace of $\mathcal{P}_3(\mathbf{R})$?

2.1 Test Your Understanding Solution

1. *Since the empty set does not contain any elements, it does not contain an element that can be an additive identity. Therefore the set* $\{\}$ *does not fulfill part 3. in the definition of a vector space.*

2. *In* $\mathcal{P}_3(\mathbf{R})$ *we have* $0 = 0x^3 + 0x^2 + 0x + 0$. *Evaluating this function at* 0 *does not yield* 1. *Therefore* $0 \notin S$. *Therefore* S *is not a vector space. Therefore* S *is not a subspace of* $\mathcal{P}_3(\mathbf{R})$.

2.2 Span, Linear Independence and Basis Part I (LADR 2A/2B)

Recall the following from definitions from MATH164.

Definition 2.6. *Let $n, k \geq 1$ be integers and let $S \subsetneq \mathbf{R}^n$ so that*

$$S = \{v_1, v_2, \ldots, v_k\}$$

*The **span of** S, denoted $span(S)$ is the set of all linear combinations of elements of S. That is*

$$span(S) = \{a_1v_1 + a_2v_2 + \cdots + a_kv_k \mid a_1, a_2, \ldots, a_k \in \mathbf{R}\}$$

*We say S **spans** $\mathbf{R}^n$ when $span(S) = \mathbf{R}^n$.*

Definition 2.7. *Let $n, k \geq 1$ be an integer and let $S \subsetneq \mathbf{R}^n$ so that*

$$S = \{v_1, v_2, \ldots, v_k\}$$

*We say S **is linearly independent** when the only solution to*

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$$

is $a_1 = a_2 = a_3 = \cdots = a_k = 0$.

Looking at these definitions, we notice that defining span and linear independence in $\mathbf{R}^n$ required us to know the meaning of addition and scalar multiplication in $\mathbf{R}^n$. Since these operations have meanings in an arbitrary vector space, we can make meaning of these terms for any vector space.

Aside. *In defining space and linear independence for vector spaces we will consider lists of elements as opposed to sets. This allows us to put our elements in order, which will simplify matters further on when we talk about adding and removing vectors from linearly dependent and independent lists.*

Definition 2.8. *Let V be a vector space and let $S = (v_1, v_2, \ldots, v_k)$ be a list of vectors of V.*

*The **span of** S, denoted $span(S)$ is the set of all linear combinations of elements of S.*

That is

$$span(S) = \{a_1v_1 + a_2v_2 + \cdots + a_kv_k \mid a_1, a_2, \ldots, a_k \in \mathbf{F}\}$$

*We say S **spans** V when $span(S) = V$.*

Why should we care about spans of elements in a vector space? In the previous section, we saw that (some?) subspaces of $\mathbf{R}^3$ can be expressed as a span of a set of vectors. In fact, the span of any set of elements of a vector space give rise to a subspace:

Theorem 2.9. *Let V be a vector space and let $S = (v_1, v_2, \ldots, v_k)$ be a list of elements of V. The set $span(S)$ is the smallest subspace of V that contains all of the elements of S.*

Wait... what does *smallest* mean in this context? We'll return to this theorem and its proof on Assignment 2.

Trivially, we saw that $\mathbf{R}^3$ is a subspace of $\mathbf{R}^3$. Without too much extra thinking, we can convince ourselves that every vector space is a subspace of itself. As we saw in MATH164, $\mathbf{R}^n$ can be expressed as the span of a set vectors (for example, the standard basis vectors). This turns out not to be true for all vector spaces. That is, there exists vector spaces that do not arise as the span of some set of objects.

Let $\mathcal{P}(\mathbf{F})$ be the set of all polynomials with coefficients in $\mathbf{F}$. Let S be a list of elements of $\mathcal{P}(\mathbf{F})$. We explain why S does not span $\mathcal{P}(\mathbf{F})$. This then implies that no list can span S.

Let $S = (f_1, f_2, f_3, \ldots, f_k)$. Consider a polynomial in S that has the largest degree among all polynomials in the list S. That is, let $g \in S$ so that $deg(g) \geq deg(f_i)$ for all $i \in \{1, 2, \ldots k\}$.

Since $g \in \mathcal{P}(\mathbf{F})$, we have

$$g(x) = a_t x^t + a_{t-1} x^{t-1} + \cdots + a_1 x + a_0$$

where $a_t \neq 0$ and $deg(g) = t$.

Consider the polynomial

$$h(x) = x^{t+1}.$$

Since g is a polynomial of maximum degree in S, when we take linear combinations of elements S we can only construct polynomials with degree at most t. Therefore $h \notin span(S)$. Since $h \in \mathcal{P}(\mathbf{F})$ and $h \notin span(S)$, by definition S does not span $\mathcal{P}(\mathbf{F})$. Therefore no list can span $\mathcal{P}(\mathbf{F})$.

Aside. *Thinking ahead to the next module when we discuss bases of vector spaces, what is this saying about a basis for $\mathcal{P}(\mathbf{F})$?*

Definition 2.10. *Let V be a vector space. We say V is **finite dimensional** when V can be expressed as the span of some list of vectors in V. Otherwise we say V is **infinite dimensional**.*

Our argument above tells us that $\mathcal{P}(\mathbf{F})$ is infinite dimensional. Surprisingly, however, $\mathcal{P}(\mathbf{F})$ contains finite dimensional subspaces. For example, the set of all polynomials of degree at most four is spanned by the list $(1, x, x^2, x^3, x^4)$.

Now that we are slightly comfortable thinking about spans in arbitrary vector spaces, we now turn to the idea of linear independence for a vector space. Just as we can define linear independence in $\mathbf{R}^n$, so too can we define the concept for an arbitrary vector space.

Definition 2.11. *Let V be a vector space and let $S = (v_1, v_2, \ldots, v_k)$ be a list of vectors of V. We say S* ***is linearly independent*** *when the only solution to*

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$$

is $a_1 = a_2 = a_3 = \cdots = a_k = 0$.

When S is not linearly independent, we say S ***is linearly dependent****.*

By way of example, let us think about the real vector space $\mathcal{P}_3(\mathbf{R})$, the set of polynomials of degree at most three with coefficients in $\mathbf{R}$. Consider the list $S = (x^2+2x+3, 2x+2, x^2+1)$.

Computing the span, we have:

$$span(S) = \{a_1(x^2 + 2x + 3) + a_2(2x + 2) + a_3(x^2 + 1) | a_1, a_2, a_3 \in \mathbf{R}\}$$

Rearranging this linear combination we find:

$$a_1(x^2 + 2x + 3) + a_2(2x + 2) + a_3(x^2 + 1) = (a_1 + a_3)x^2 + (2a_1 + 2a_2)x + (3a_1 + 2a_2 + a_3)$$

And so

$$span(S) = \{(a_1 + a_3)x^2 + (2a_1 + 2a_2)x + (3a_1 + 2a_2 + a_3) | a_1, a_2, a_3 \in \mathbf{R}\}$$

To determine if S is linearly independent, we recall that in $\mathcal{P}_3(\mathbf{R})$ we have

$$0 = 0x^3 + 0x^2 + 0x + 0.$$

And so to determine if S is linearly independent, we seek solutions, (a_1, a_2, a_3), for

$$(a_1 + a_3)x^2 + (2a_1 + 2a_2)x + (3a_1 + 2a_2 + a_3) = 0x^3 + 0x^2 + 0x + 0$$

Since polynomials are equal exactly when they have the same coefficients, we are looking for solutions to the following system of equations:

$$\begin{array}{rrrl} a_1 & & +a_3 & = 0 \\ 2a_1 & +2a_2 & & = 0 \\ 3a_1 & +2a_2 & +a_3 & = 0 \end{array}$$

(Take a moment to convince yourself that you know where these equations came from. The first equation is related to the coefficient of x^2.)

.
.
.
.

If the only solution is $a_1 = a_2 = a_3 = 0$, then our definition of linear independence tells us that S is linearly independent. Otherwise, if there exists non-trivial solutions, then S is linearly dependent. Using methods from MATH164, we can find the general solution for this system to be:

$$\begin{aligned} a_1 &= -t \\ a_2 &= t \\ a_3 &= t \end{aligned}$$

for $t \in \mathbf{R}$.

Therefore $a_1 = a_2 = a_3 = 0$ is not the only solution to this system. And so S is linearly dependent.

Just like for linear dependence in $\mathbf{R}^n$, when we have a linear dependent list, we can show that some vector on the list can be expressed as a linear combination of the others. For example, $x^2 + 1$ can be expressed as a linear combination of $x^2 + 2x + 3$ and $2x + 2$:

$$x^2 + 1 = (1)(x^2 + 2x + 3) + (-1)(2x + 2)$$

Aside. *Presumably, as part of your work in MATH164, you have had plenty of practice in solving systems of equations and manipulating matrices. Unless otherwise specified, you will not be required to show your work for these sorts of computations in MATH266. And so you are encouraged to use a computer for these sorts of computations in this course. If you do use a computer for computational work, please note what website (Desmos, Wolfram Alpha, etc...) or software (SageMath, Maple, etc...) you've used.*

Just like with linear dependent lists in $\mathbf{R}^n$, we can remove elements from a linearly dependent list to arrive at a linearly independent list. Similarly, maximal independent sets form a basis. In Module 3, we will use these ideas to show that every finite dimensional vector space has a basis and that the only subspaces of $\mathbf{R}^3$ are the zero vector, lines through the origin, planes through the origin and $\mathbf{R}^3$ itself.

2.2 Test your Understanding

1. Let $\mathcal{M}_{2\times 2}(\mathbf{R})$ be the set of 2×2 matrices with entries in $\mathbf{R}$. Using the definition of linearly independent, determine if the following list of elements of $\mathcal{M}_{2\times 2}(\mathbf{R})$ is linearly independent.

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

2. Let $\mathcal{T}_{2\times 2}(\mathbf{R})$ be the set of 2×2 upper triangular matrices with entries in $\mathbf{R}$. Find a linearly independent list S of length 3 so that $span(S) = \mathcal{T}_{2\times 2}(\mathbf{R})$.

2.2 Test your Understanding Solution

1. In $\mathcal{M}_{2\times 2}(\mathbf{R})$ we have
$$0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
And so to determine if the list is linearly independent we seek solutions (a_1, a_2, a_3, a_4) to
$$a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
Simplifying on the left we have:
$$\begin{bmatrix} a_1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & a_2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ a_3 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & a_4 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$$
The only solution for
$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
is $a_1 = a_2 = a_3 = a_4 = 0$. Therefore the list is linearly independent.

2. Using the method from the previous part we can conclude that the list
$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$
is linearly independent. Taking the span of this list we have:
$$\left\{ a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \mid a_1, a_2, a_3 \in \mathbf{R} \right\} = \left\{ \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix} \mid a_1, a_2, a_3 \in \mathbf{R} \right\} = \mathcal{T}_{2\times 2}(\mathbf{R})$$

---

3 Linear Independence, Basis and Dimension (LADR 2B/2C)

Learning Incomes.

- *Know what a vector space is, along with some common examples.*
- *Be able to manipulate equalities in a vector space*
- *Understand the definitions of span, linear independence in a vector space*
- *Understand the definition of basis in* $\mathbf{R}^n$

Learning Outcomes.

- *Understand the statement and proof of the Linear Dependence Lemma*
- *Be able to explain, by referring to theorems, how you know that the only subspaces of* $\mathbf{R}^3$ *are the zero vector, lines through the origin, planes through the origin and* $\mathbf{R}^3$ *itself.*
- *Understand the statement and proof of Theorem 3.3.*
- *Become familiar with the technique of Proof by Contradiction*
- *Understand the relationship between basis and unique representation through the statement and proof of Theorem 3.6*
- *Be able to explain how you know that every finite dimensional vector space has a basis and that every such basis has the same length*
- *Understand the definitions of* ***basis*** *and* ***dimension***

Newly Defined Terms and Notation.

- *basis, dimension,* $\dim V$

In Module 2 we extended our intuition for span and linear independence in $\mathbf{R}^n$ to be meaningful for every vector space. We continue this work this week as we explore ideas around basis and dimension.

Recall the following definition from MATH164.

Definition 3.1. *Let* $n \geq 1$ *be an integer and let* B *be a finite set of vectors in* $\mathbf{R}^n$*. We say* B ***is a basis*** *when*

1. B *is linearly independent; and*
2. $span(B) = \mathbf{R}^n$*.*

From MATH164, our intuition for a basis for $\mathbf{R}^n$ is as a smallest set of building blocks for all of the elements in $\mathbf{R}^n$. More precisely, for a basis B of $\mathbf{R}^n$, every vector of $\mathbf{R}^n$ can be expressed uniquely (i.e., built in just one way) as a linear combination of elements of B. We learned many facts about bases in MATH164:

- Every basis of $\mathbf{R}^n$ contains exactly n vectors.
- Any spanning collection of vectors of $\mathbf{R}^n$ can be turned in to basis by (possibly) tossing away some of the vectors.
- Any linearly independent set that is not a basis can be turned in to a basis by adding some more vectors.

Unsurprisingly, each of these facts remain true when we extend our idea of basis to any arbitrary vector space. Using these facts we prove that every finite dimensional vector space has a basis. For $\mathbf{R}^n$ we can always choose the standard basis vectors for our basis. In a future module, we will see that every finite dimensional vector space has an analogous *standard* basis.

Though span, linear independence and basis seem to naturally extend to arbitrary vector spaces, extending our idea of dimension for $\mathbf{R}^n$ takes a little more care. In $\mathbf{R}^n$, our intuition around the term dimension likely triggers in our minds some sort of notion of size. Even thought both $\mathbf{R}^2$ and $\mathbf{R}^3$ both contain infinitely many vectors, we perhaps intuit that $\mathbf{R}^3$ is, in some sense, bigger than $\mathbf{R}^2$.

In our discussion of bases of a finite dimensional vector space we will prove that every basis of a finite dimensional vector space contains the same number of elements. We will use this parameter, number of elements in a basis, to define the meaning of the term dimension for any finite dimensional vector space. This sense of size for a vector space will come in handy when we discuss functions that map one vector space to another. Just as with linear functions $T : \mathbf{R}^n \to \mathbf{R}^m$ we will be able to assert the existence of injections, surjections and bijections based only the dimensions of the respective vector spaces.

We begin this work by continuing our study of linear independence and dependence from Module 2.

Aside. *A quick note before we begin: If you are following along with the textbook, you may find that we have slightly changed the order and presentation of some of the material. In particular, some of the proofs you will find in this module are different than the ones in the text. This is intentional – in the module we introduce the technique of proof by contradiction. And so some proofs below are modified so that they use this technique.*

3.1 Span, Linear Independence and Basis Part II (LADR 2A/2B)

Let us begin by returning to our example from the previous module.

Consider the list $S = (x^2 + 2x + 3, 2x + 2, x^2 + 1)$ of polynomials in $\mathcal{P}_3(\mathbf{R})$. To determine if this list is linearly independent or not, we sought solutions (a_1, a_2, a_3) to:

$$a_1(x^2 + 2x + 3) + a_2(2x + 2) + a_3(x^2 + 1) = 0x^3 + 0x^2 + 0x + 0$$

We found the general solution:

$$\begin{aligned} a_1 &= -t \\ a_2 &= t \\ a_3 &= t \end{aligned}$$

for $t \in \mathbf{R}$.

And so there are infinitely many solutions, one for each value of $t \in \mathbf{R}$. Since $a_1 = a_2 = a_3 = 0$ is not the only solution, this list is linearly dependent. Choosing $t = -1$, gives:

$$(x^2 + 2x + 3) + (-1)(2x + 2) + (-1)(x^2 + 1) = 0$$

Rearranging yields

$$(x^2 + 2x + 3) + (-1)(2x + 2) = (x^2 + 1)$$

Just as we would expect for linearly dependence in $\mathbf{R}^n$, we can express one of the elements of the list as a linear combination as the others.

Aside. *We chose $t = -1$ because it led to something useful: expressing the third polynomial as a linear combination of the others. Choosing any other $t \neq 0$ would have permitted this, but the choice of $t = -1$ made it so we didn't have to work very hard to get the third polynomial as a linear combination of the others.*

And, in fact, removing this *redundant* polynomial does not change the span:

$$\begin{aligned} &span(x^2 + 2x + 3, 2x + 2, x^2 + 1) \\ &= \{a_1(x^2 + 2x + 3) + a_2(2x + 2) + a_3(x^2 + 1) \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{a_1(x^2 + 2x + 3) + a_2(2x + 2) + a_3((x^2 + 2x + 3) + (-1)(2x + 2)) \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{(a_1 + a_3)(x^2 + 2x + 3) + (a_2 - a_3)(2x + 2) \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{b_1(x^2 + 2x + 3) + b_2(2x + 2) \mid b_1, b_2 \in \mathbf{R}\} \\ &= span(x^2 + 2x + 3, 2x + 2) \end{aligned}$$

For the second-to-last equality, we notice when we let a_1, a_2 and a_3 range over all real numbers, then subsequently $a_1 + a_3$ and $a_2 - a_3$ will range over all real numbers. We let $b_1 = a_1 + a_3$ and $b_2 = a_2 - a_3$.

Keeping track of all of these polynomials is a pain. Let $f_1 = x^2+2x+3$, $f_2 = 2x+2$, $f_3 = x^2+1$. Repeating the work above, we have:

$$\begin{aligned} span(f_1, f_2, f_3) &= \{a_1 f_1 + a_2 f_2 + a_3 f_3 \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{a_1 f_1 + a_2 f_2 + a_3 (f_1 - f_2) \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{(a_1 + a_3) f_1 + (a_2 - a_3) f_2 \mid a_1, a_2, a_3 \in \mathbf{R}\} \\ &= \{b_1 f_1 + b_2 f_2 \mid b_1, b_2 \in \mathbf{R}\} \\ &= span(f_1, f_2) \end{aligned}$$

Looking at this work, we can almost forget about the fact that we have $f_1, f_2, f_3 \in \mathcal{P}_3(\mathbf{R})$. These same sorts of manipulations seem to be true no matter which vector space we are working in. We explore this further.

Let V be a vector space and let S be a linearly dependent list in V. Then, there exists some $v \in V$ so that removing v from S does not change the span of S. But, how do we know which $v \in V$ we can choose?

Let $S = (v_1, v_2, \ldots, v_k)$. If S is linearly dependent, then by definition there exists scalars $a_1, a_2, \ldots, a_k$ so that

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = 0$$

so that at least one of the scalars is not 0. If $a_k \neq 0$, then

$$v_k = \frac{a_1}{a_k} v_1 + \frac{a_2}{a_k} v_2 + \cdots \frac{a_{k-1}}{a_j} v_{k-1}$$

and

$$span(v_1, v_2, \ldots, v_k) = span(v_1, v_2, \ldots, v_{k-1})$$

But, what if $a_k = 0$? Well, if $a_{k-1} \neq 0$, then we have:

$$\begin{aligned} v_{k-1} &= \frac{a_1}{a_{k-1}} v_1 + \frac{a_2}{a_{k-1}} v_2 + \cdots \frac{a_{k-2}}{a_{k-1}} v_{k-2} + \frac{a_k}{a_{k-1}} v_k \\ &= \frac{a_1}{a_{k-1}} v_1 + \frac{a_2}{a_{k-1}} v_2 + \cdots \frac{a_{k-2}}{a_{k-1}} v_{k-2} \end{aligned}$$

and

$$span(v_1, v_2, \ldots, v_k) = span(v_1, v_2, \ldots, v_{k-2}, v_k)$$

But, what if $a_k = 0$ and $a_{k-1} = 0$? Well, we can do the same process with a_{k-2} (as long as it doesn't equal zero).

Since at least one of $a_1, a_2, \ldots, a_k$ is not zero, then we can scan this list from the right to find a non-zero scalar. Let j be the index of the rightmost scalar that is not zero. That is,

$a_j \neq 0$ but each of $a_{j+1}, a_{j+2}, \ldots a_k = 0$. Then:

$$\begin{aligned} v_j &= \frac{a_1}{a_j}v_1 + \frac{a_2}{a_j}v_2 + \cdots + \frac{a_{j-1}}{a_j}v_{j-1} + \frac{a_{j+1}}{a_j}v_{j+1} + \cdots + \frac{a_k}{a_j}v_k \\ &= \frac{a_1}{a_j}v_1 + \frac{a_2}{a_j}v_2 + \cdots + \frac{a_{j-1}}{a_j}v_{j-1} \end{aligned}$$

and

$$span(v_1, v_2, \ldots, v_k) = span(v_1, v_2, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k)$$

We state this result as a theorem so we may refer to it as necessary.

Theorem 3.2 (Linear Dependence Lemma)**.** *Let V be a vector space and let $S = (v_1, v_2, \ldots, v_m)$ be a linearly dependent list in V. Then there exists $j \in \{1, 2, \ldots, m\}$ so that:*

(a) $v_j \in span(v_1, v_2, \ldots, v_{j-1})$; and

(b) removing v_j from S does not change the span of S. That is,

$$span(S) = span(v_1, v_2, \ldots, v_{j-1}, v_{j+1}, \ldots, v_m)$$

Rather than write a formal proof of this fact, we take this moment to practice our skills in proof reading. The proof of this result in on page 34 (2.21 - Linear Dependence Lemma) of the course text. The approach of the proof in the text is identical to the argument we have outlined above. Take some time to read the proof and re-write the proof in you own words in your own notes.

.
.
.
.
.
.
.
.
.
.
.

(Developing our math literacy skills is one of the goals of this course. You can expect that on a future assessment you will be asked to read and understand a proof from the text. This is a good opportunity to practice this.)

The statement of the Linear Dependence Lemma not only tells us that we can remove an element from a linearly dependent list without changing the span, but also tells us how to find such an element. We find the rightmost non-zero scalar a_j in a non-trivial solution for:

$$a_1v_1 + a_2v_2 + \cdots + a_mv_m = 0$$

and then remove v_j.

Determining whether a list is linearly independent or not can be a cumbersome process: at some point we are faced with having to solve a system of linear equations and possibly manipulate a matrix. However, from MATH164, we had some other techniques to determine if a list was linearly independent. For example the list

$$\begin{pmatrix}1 & 2\end{pmatrix}, \begin{pmatrix}1 & 1\end{pmatrix}, \begin{pmatrix}-6 & 11\end{pmatrix}, \begin{pmatrix}2 & 9\end{pmatrix}$$

is not linearly independent in $\mathbf{R}^2$: there are too many vectors!

In $\mathbf{R}^2$, we recall that a list of three or more vectors must be linearly dependent. An analogous fact turns out to be true in any finite dimensional vector space. That is, for any finite dimensional vector space V, there exists an integer b so that any set of more than b points of the vector space is linearly dependent. For $\mathbf{R}^2$ we have $b = 2$ and in general we have $b = n$ for $\mathbf{R}^n$. It turns out that this number b, is the length of a linearly independent spanning list. Before we prove this fact, we first prove that the length of any linearly independent list is bounded above by the length of every spanning list.

Theorem 3.3. *Let V be a finite-dimensional vector space. If $L = (u_1, u_2, \dots u_m)$ is linearly independent list of elements of V and $S = (w_1, w_2, \dots, w_n)$ spans V, then $m \leq n$.*

Before we prove this theorem, let us take a moment to study the structure of this mathematical fact:

If $L = (u_1, u_2, \dots u_m)$ is linearly independent list of elements of V and $S = (w_1, w_2, \dots, w_n)$ spans V, then $m \leq n$

This is an *if, then* statement. If the hypothesis is true:

$L = (u_1, u_2, \dots u_m)$ is linearly independent list of elements of V and $S = (w_1, w_2, \dots, w_n)$ spans V

then the conclusion is true:

$$m \leq n.$$

We will verify this mathematical fact using a technique called *proof by contradiction*. In brief, we will assume that the hypothesis is true and the conclusion is false and then deduce something absurd (i.e., untrue). Since assuming the conclusion was false led to an absurdity (i.e., an untruth), then necessarily the conclusion must be true. When an author uses a *proof by contradiction* they will (usually) state it somewhere in the introduction to the proof.

If you have never seen the technique of *proof by contradiction*, or need refresher, go watch this:

- `https://www.youtube.com/watch?v=rV9esU9gHO8` (Proof by Contradiction Part I)
- `https://www.youtube.com/watch?v=J8N_yFXeA90` (Proof by Contradiction Part II)

The mathematical fact being proven in these videos isn't germane to MATH266, but these videos are a nice introduction to the idea of *Proof by Contradiction*. If you do not have access to YouTube and you are unfamiliar with *Proof by Contradiction*, please be in touch so I can point you towards other resources.

Aside. *In the videos above the instructor is teaching secondary-school students in Australia. In parts of the world that Canada generally compares itself to, topics like mathematical proof are a standard part of the secondary-school curriculum.*

If you haven't seen mathematical proof before this class, do not be ashamed, embarrassed or apologetic; your experience with mathematics is precisely the product of the mathematics you have been exposed to through your formal education – a factor you could have done little to control.

For those of you who attended secondary school in Saskatchewan, recognize that mathematical literacy for students in Saskatchewan is among the lowest in the country[2]*. If you are on the path to being an educator in this province, you will be uniquely positioned to push back against provincial governments whose priorities may not include public education. It is incumbent on all of us to not accept the status quo.*

In our proof we will proceed by contradiction. That is, we will assume L is linearly independent and S is spanning, but also that $m > n$.

We notice the following fact: since S spans V, adding any extra element of S will result in a linearly dependent list. Using the Linear Dependence Lemma, once we have added a new element to the front of S, we can remove one of the original elements of S without changing the span of S.

In our proof we proceed by adding elements of L to S one-by-one, and removing elements of S so that at every step S is still spanning. If $m > n$, then we will remove all of the original elements of S before we have added all of the elements of L. Once we have removed all of the original elements of S, we have a spanning list that contains only elements of L. Therefore L contains a spanning list. Since L contains a spanning list, then any element of L that is not part of the spanning piece can be expressed as linear combination of elements in the spanning list. Therefore L is not linearly independent. This is a contradiction. Therefore $n \leq m$.

[2] O'Grady, K., M. Deussing, T. Scerbina, Y. Tao, K. Fung, V. Elez, J. Monk. 2019b. "Measuring up: Canadian Results of the OECD PISA2018 Study." Council of Ministers of Education, Canada (CMEC).

Without our proof sketch in place, we proceed.

Proof. Let V be a vector space so that $L = (u_1, u_2, \dots u_m)$ is linearly independent and $S = (w_1, w_2, \dots, w_n)$ spans V. We want to show that $m \leq n$. We proceed by contradiction. To do so we will assume $m > n$ and then derive a contradiction.

Since S is spanning, we have $span(S) = V$. And since $u_1 \in V$ we have $u_1 \in span(S)$. That is, u_1 can be expressed as a linear combination of elements of S. Therefore the list $(u_1, w_1, \dots, w_n)$ is linearly dependent. By the Linear Dependence Lemma, we can find an element of this list to remove without changing the span of this list. That is, there exists $j \in \{1, 2, 3, \dots, n\}$ so that removing w_j from $(u_1, w_1, \dots, w_n)$ doesn't change the span. By the second part of the Linear Dependence Lemma, even once we have removed w_j, the list $(u_1, w_1, \dots, w_n)$ is still spanning.

Consider repeating this process with u_2. That is, add u_2 to $(u_1, w_1, \dots, w_n)$ and remove some element w_i so that the span does not change. As before when w_i and w_j are removed, the list $(u_2, u_1, w_1, \dots, w_n)$ is still spanning.

Consider now repeating this process of adding elements of L to S until all of the original elements of S have been removed. Since S originally had n elements, we have added all of $u_1, u_2, \dots, u_n$ to S. Similar to the reasoning above, adding all of $u_1, u_2, \dots, u_n$ to S and removing all of $w_1, w_2, \dots, w_n$ does not change the span of S. Since S is spanning, we have $span(u_1, u_2, \dots, u_n) = span(S) = V$.

Since $m > n$ and L has m elements, there is an element of L that is not one of $u_1, u_2, \dots, u_n$. In particular, u_{n+1} is an element of L that has not been added to S.

Since $span(S) = V$ we have $u_{n+1} \in span(S)$. However, since $span(S) = span(u_1, u_2, \dots, u_n)$, we have that $u_{n+1} \in span(u_1, u_2, \dots, u_n)$. Therefore there exist scalars $a_1, a_2, \dots a_n$ so that

$$u_{n+1} = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n$$

and not all of $a_1, a_2, \dots, a_n$ are 0.

Rearranging yields:

$$a_1 u_1 + a_2 u_2 + \cdots + a_n u_n + (-1) u_{n+1} = 0$$

Therefore L is linearly dependent. This is a contradiction: having L be linearly independent was part of our hypothesis. Thus if L is linearly independent we cannot have $m > n$. Therefore we must have $m \leq n$. □

Before we move on, the Linear Dependence Lemma has one more gem for us to mine. Let V be a finite-dimensional vector space and S be a list of elements of V that spans V. If S is linearly dependent, then by the Linear Dependence Lemma we can remove an element from S without changing the span of S. Since a list that contains a single vector is necessarily linearly independent, repeating the process of removing elements from S will eventually yield a spanning list that is linearly independent.

When $V = \mathbf{R}^n$ this is exactly the process we use to form a basis from a linearly dependent spanning set – we remove the *redundant* vectors until what remains is a basis. To consider this same idea for an arbitrary vector space, we first must define the meaning of the word basis for an arbitrary vector space. We proceed with this work in the following section.

Before we move on, let us return back to one of our threads from the previous module: what are all of the subspaces of $\mathbf{R}^3$?

From the previous module, we concluded that the zero vector, lines through the origin, planes through the origin and $\mathbf{R}^3$ itself were each subspaces of $\mathbf{R}^3$. By definition, each of the these subspaces is finite-dimensional: they are spanned by some list of vectors. Moreover, the span of every list of vectors in $\mathbf{R}^3$ is one of these objects. And so these objects make up all of the finite-dimensional subspaces $\mathbf{R}^3$. To prove that these are all possible subspaces of $\mathbf{R}^3$ we must prove that every subspace of $\mathbf{R}^3$ is necessarily finite dimensional.

Our previous result, Theorem 3.3 tells us that any list of vectors that is longer than any spanning list must necessarily be linearly dependent. Let V be a finite-dimensional vector space and let U be a subspace of V. Since V is finite-dimensional, there is a list of vectors that spans V. Let B be a shortest list of vectors that spans V. Let b be the number of vectors in B.

Consider now building a list of vectors of U so that at each step the list remains linearly independent. This process cannot continue indefinitely as the number of vectors in such a list cannot exceed b. This means that there is some list of vectors B_U so that adding any new vector in U to B_U results in a linearly dependent list. Therefore, every vector in U that is not in B_U can be expressed as a linear combination of vectors in B_U. Thus $span(B_U) = U$. This then implies that U is finite-dimensional.

Theorem 3.4. *Every subspace of a finite dimensional vector space is finite dimensional.*

Since every subspace of a finite-dimensional vector space is finite dimensional, and the span of any list of vectors in $\mathbf{R}^3$ is either the 0 vector, a line through the origin, a plane through the origin or $\mathbf{R}^3$, these are the only subspaces of $\mathbf{R}^3$.

3.1 Test your Understanding

1. Consider the equality:

$$0\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}+2\begin{bmatrix}1 & 0\\ 1 & 3\end{bmatrix}+(-1)\begin{bmatrix}2 & 0\\ 2 & 6\end{bmatrix}+0\begin{bmatrix}1 & 6\\ 4 & 9\end{bmatrix}+0\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$

   If we apply the Linear Dependence Lemma to remove an element from the list,

$$\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\begin{bmatrix}1 & 0\\ 1 & 3\end{bmatrix},\begin{bmatrix}2 & 0\\ 2 & 6\end{bmatrix},\begin{bmatrix}1 & 6\\ 4 & 9\end{bmatrix},\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\right)$$

   what is the value of j?

2. Let $L=((1,2,1),\ (2,1,1),\ (0,0,1),\ (2,3,0))$

   (a) Show that the set is linearly dependent using the definition of linear dependent.

   (b) Using the Linear Dependence Lemma, find an element of L that you can remove without changing the span of L.

3.1 Test your Understanding Solution

1. Let
$$S = (v_1, v_2, v_3, v_4, v_5) = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 2 & 6 \end{bmatrix}, \begin{bmatrix} 1 & 6 \\ 4 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right)$$

   In the equation
$$0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 2 \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} + (-1) \begin{bmatrix} 2 & 0 \\ 2 & 6 \end{bmatrix} + 0 \begin{bmatrix} 1 & 6 \\ 4 & 9 \end{bmatrix} + 0 \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

   We have $a_1 = 0, a_2 = 1, a_3 = -2, a_4 = a_5 = 0$. Since a_3 is the rightmost non-zero coefficient, we have $j = 3$ in the statement of the linear dependence lemma.

   And so, $span(S) = span(v_1, v_2, v_4, v_5)$.

2. To test if $L = ((1, 2, 1),\ (2, 1, 1),\ (0, 0, 1),\ (2, 3, 0))$ linearly dependent, we are looking for solutions to:
$$a_1(1, 2, 1) + a_2(2, 1, 1) + a_3(0, 0, 1) + a_4(2, 3, 0) = (0, 0, 0)$$

   Expressing vector equation as a system of linear equations we have
$$\begin{array}{rrrrl} a_1 & +2a_2 & +0a_3 & +2a_4 & = 0 \\ 2a_1 & +a_2 & +0a_3 & +3a_4 & = 0 \\ a_1 & +a_2 & +a_3 & +0a_4 & = 0 \end{array}$$

   When row reduced, the corresponding matrix is:
$$\begin{bmatrix} 1 & 0 & 0 & \frac{4}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \end{bmatrix}$$

   Therefore, the general solution to this system is:
$$\begin{aligned} a_1 &= -\frac{4}{3}t \\ a_2 &= -\frac{1}{3}t \\ a_3 &= \frac{5}{3}t \\ a_4 &= t \end{aligned}$$

   Setting $t = 1$ gives the following solution:
$$-\frac{4}{3}(1, 2, 1) - \frac{1}{3}(2, 1, 1) + -\frac{5}{3}(0, 0, 1) + 1(2, 3, 0) = (0, 0, 0)$$

   This is a non-trivial solution, and so L is linearly dependent. Applying the Linear Dependence Lemma, we have $j = 4$ and
$$span(L) = span((1, 2, 1),\ (2, 1, 1),\ (0, 0, 1))$$

3.2 Span, Linear Independence and Basis Part III

We continue with our examination of finite dimensional vector spaces by extending our definition for basis to work in any finite dimensional vector space:

Definition 3.5. *Let V be a finite-dimensional vector space and let B be a list of vectors in V. We say B* ***is a basis*** *when*

1. *B is linearly independent; and*
2. *$span(B) = V$.*

For example,

$$B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

is a basis for $\mathcal{M}_{2\times 2}(\mathbf{R})$, the set of 2×2 matrices with entries in $\mathbf{R}$. To see this, we look back at the our definition of basis:

1. *B is a linearly independent:* This was an exercise in Module 2.
2. *$span(B) = \mathcal{M}_{2\times 2}(\mathbf{R})$*: To see this, we notice:

$$\mathcal{M}_{2\times 2}(\mathbf{R}) = \left\{ \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \mid a_1, a_2, a_3, a_4 \in \mathbf{R} \right\}$$

and

$$\begin{aligned} span(B) &= \left\{ a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \mid a_1, a_2, a_3, a_4 \in \mathbf{R} \right\} \\ &= \left\{ \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \mid a_1, a_2, a_3, a_4 \in \mathbf{R} \right\} \end{aligned}$$

As with bases in $\mathbf{R}^n$, a basis in a finite dimensional vector space is a set of building blocks that lets us uniquely build any element of V. In fact, this condition is equivalent to a list being a basis. That is,

Theorem 3.6 (Criterion for Basis)**.** *Let V be a finite dimensional vector space. A list $B = (v_1, v_2, \ldots, v_k)$ is a basis for V if and only if every $u \in V$ there is exactly one collection of scalars $(a_1, a_2, \ldots, a_k)$ so that*

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = u$$

Recall from our discussion to meaning of the phrase *if and only if* in the statement of a theorem. To prove this theorem we must prove:

1. If B is a basis, then there is only one solution, $(a_1, a_2, \dots a_k)$, for $\sum_{i=1}^{k} a_i v_i = u$ for each $u \in V$; and
2. if there is only one solution, $(a_1, a_2, \dots a_k)$, for $\sum_{i=1}^{k} a_i v_i = u$ for each $u \in V$, then B is a basis.

Aside 3.7. *Reading proofs is difficult. Not only do you need to keep track of what all of the pieces of notation mean, but you also need to have an idea of the broad goal of each part of the proof.*

Before you read this proof, take a moment to remind yourself of the meaning of the definitions of ***linearly independent, span*** *and* ***basis****.*

The proof below is split in to two parts. In the first part we show that if B is a basis then for every $u \in V$ there is only one set of scalars $(a_1, a_2, \dots a_k)$ so that $\sum_{i=1}^{k} a_i v_i = u$. We do this by looking at a potentially different set of scalars $(c_1, c_2, \dots c_k)$ so that $\sum_{i=1}^{k} c_i v_i = u$ and then concluding that in fact we have $a_1 = c_1, a_2 = c_2, \dots, a_k = c_k$.

In the second half of the proof we show that if for every $u \in V$ there is only one set of scalars $(a_1, a_2, \dots a_k)$ so that $\sum_{i=1}^{k} a_i v_i = u$, then B is a basis. We do this by arguing that B satisfies the two parts of the definition of a basis: linear independence and spanning.

Ready slowly and carefully. Each sentence should follow from the previous sentences. Don't be discouraged if it doesn't make sense at first!

Proof. Let V be a finite-dimensional vector space.

1. Let B be a basis for V and let $u \in V$. Since B is a basis, then B spans V. Therefore $u \in span(B)$. Therefore there exists scalars $a_1, a_2, \dots, a_k$ such that

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = u$$

To show that this is the only solution, we consider another possible solution: $(c_1, c_2, \dots, c_k)$ so that

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = u$$

To show that the solution $(a_1, a_2, \dots, a_k)$ is the only possible solution, we show

$$(a_1, a_2, \dots, a_k) = (c_1, c_2, \dots, c_k)$$

To do so, we show $a_i - c_i = 0$ for every $i \in \{1, 2, \dots, k\}$.

Consider the following equation:

$$\begin{aligned} 0 &= u - u \\ &= (a_1 v_1 + a_2 v_2 + \cdots + a_k v_k) - (c_1 v_1 + c_2 v_2 + \cdots + c_k v_k) \\ &= (a_1 - c_1) v_1 + (a_2 - c_2) v_2 + \cdots + (a_k - c_k) v_k \end{aligned}$$

Since B is a basis and B is linearly independent, then the only linear combination of elements of B that gives 0 is the one where all of the coefficients are 0. Since we have:

$$0 = (a_1 - c_1)v_1 + (a_2 - c_2)v_2 + \cdots + (a_k - c_k)v_k$$

We must have $a_i - c_i = 0$ for every $i \in \{1, 2, \ldots, k\}$. Therefore $(a_1, a_2, \ldots, a_k) = (c_1, c_2, \ldots, c_k)$. And so $(a_1, a_2, \ldots, a_k)$ is the only solution to:

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = u$$

2. Assume now that for every $u \in U$ there is exactly one solution $(a_1, a_2, \ldots a_k)$ for:

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = u$$

Recalling the definition of basis, to show B is a basis we must show that it is linearly independent and it spans V. Since $0 \in U$, there is only one solution for

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$$

Since $(a_1, a_2, \ldots, a_k) = (0, , \ldots, 0)$ is a solution, such a solution must be the only solution. Therefore B is linearly independent.

To show B spans V we must show that every element of V is contained in the span of V. Since every element of V can be expressed as linear combination of elements if B in exactly one way, then every element of V can be expressed as a linear combination of elements of V. Therefore $span(B) = V$.

Since B is both spanning and linearly independent, necessarily B is a basis. □

Returning to our thought from the end of the previous section, we can use the Linear Dependence Lemma to iteratively remove vectors from a linearly dependent spanning list until we produce a linearly independent spanning list. Since a linearly independent spanning list is a basis, we have:

Theorem 3.8. *Let V be a finite dimensional vector space. Every spanning list can be reduced to a basis of V.*

This then implies directly that every finite-dimensional vector space has a basis: By definition, every finite-dimensional vector space is spanned by some list of vectors. If such a list is linearly independent then, by definition it is basis. Otherwise, it is linearly dependent and we can remove some vectors to produce a linearly independent spanning list. That is, we can produce a basis.

Theorem 3.9. *Every finite-dimensional vector space has a basis.*

3.2 Test your Understanding

1. Using Theorem 3.6, show that the following list is a basis for $\mathcal{P}_2(\mathbf{R})$, the real vector space of polynomials of degree at most 2.

$$L = \{1 + x, 1 + x^2, 1 + x + x^2\}$$

3.2 Test your Understanding Solution

1. Let $L = \{1 + x, 1 + x^2, 1 + x + x^2\}$. To use Theorem 3.6 we must show that for every $f \in \mathcal{P}_2(\mathbf{R})$ there is exactly one solution for:

$$a_1(1 + x) + a_2(1 + x^2) + a_3(1 + x + x^2) = f$$

Let $f = b_2x^2 + b_1x + b_0$

Expanding and simplifying on the left, we have:

$$(a_2 + a_3)x^2 + (a_1 + a_3)x + (a_1 + a_2 + a_3) = b_2x^2 + b_1x + b_0$$

Since two polynomials are equal exactly when their coefficients are equal, we want to show that the following system has exactly one solution:

$$\begin{array}{rrrl} & a_2 & +a_3 & = b_0 \\ a_1 & & +a_3 & = b_1 \\ a_1 & +a_2 & +a_3 & = b_2 \end{array}$$

The matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ row reduces to I_3. Therefore $Ax = b$ has a unique solution for every $b \in \mathbf{R}^3$. And so for every $f \in \mathcal{P}_2(\mathbf{R})$, there is exactly one solution for

$$a_1(1 + x) + a_2(1 + x^2) + a_3(1 + x + x^2) = f$$

And so by Theorem 3.6, L is a basis for $\mathcal{P}_2(\mathbf{R})$.

3.3 Dimension (LADR 2C)

Back in MATH164 our idea of dimension was probably rather fuzzy. For example, we likely think of $\mathbf{R}^4$ as having dimension 4 as every element of $\mathbf{R}^4$ is a list of 4 real numbers. However, this analogy doesn't nicely extend to arbitrary finite-dimensional vector spaces; not every finite-dimensional vector space is a list of numbers.

A fundamental fact of $\mathbf{R}^4$ is that every basis of $\mathbf{R}^4$ has exactly 4 vectors and, moreover, any collection of 4 linearly independent vectors in $\mathbf{R}^4$ is necessarily a basis.

To define dimension for arbitrary finite-dimensional vectors spaces we first prove that every basis of a finite-dimensional vector space V contains the same number of elements. We will call this number of elements the **dimension of V.**

Before we prove this fact, take a moment to review the statement of Theorem 3.3 from Section 3.1.

.
.
.
.
.
.
.
.

Theorem 3.3 tells us that a linearly independent list cannot be longer than a spanning list.

Let V be a finite dimensional vector space and let B_1 and B_2 be bases of V. Let b_1 be the length of B_1 and let b_2 be the length of B_2.

Since B_1 is a basis, B_1 is linearly independent. Since B_2 is a basis, B_2 spans V. Letting $B_1 = L$ and $B_2 = S$ in the statement of Theorem 3.3, tells us that $b_1 \leq b_2$.

Similarly, since B_2 is a basis, B_2 is linearly independent. Since B_1 is a basis, B_1 spans V. Letting $B_2 = L$ and $B_1 = S$ in the statement of Theorem 3.3, tells us that $b_2 \leq b_1$.

Since $b_1 \leq b_2$ and $b_2 \leq b_1$ we have $b_1 = b_2$

Theorem 3.10. *Let V be a finite-dimensional vector space. Every basis of V has the same number of elements.*

Definition 3.11. *Let V be a finite-dimensional vector space. The **dimension of** V is the number of elements in a basis for V. We denote this parameter as* $\dim V$.

This definition agrees with our intuition for $\mathbf{R}^n$. In $\mathbf{R}^n$ every basis has exactly n vectors and $\mathbf{R}^n$ has dimension n.

Returning to our example following the Definition 3.5, the list

$$B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

is a basis for $\mathcal{M}_{2\times 2}(\mathbf{R})$ and so $\dim \mathcal{M}_{2\times 2}(\mathbf{R}) = 4$.

Just as in $\mathbf{R}^n$, for any finite-dimensional vector space V, every linearly independent list with exactly $\dim V$ elements is necessarily a basis:

Theorem 3.12. *Let V be a finite-dimensional vector space. Every linearly independent list of length $\dim V$ is a basis for V*

Proof. Let V be a finite-dimensional vector space with dimension b. Let $L = v_1, v_2, \ldots, v_b$ be a linearly independent list. Let B be a basis for V. Since B is a basis for V and $\dim V = b$, then by Theorem 3.10 B has length b.

We proceed by contradiction. That is, we assume L is not a basis and deduce a statement we know to be false. This then implies that L must be a basis.

If L is not a basis, then there exists $u \in V$ so that $u \notin span(L)$. Since $u \notin span(L)$, the list $(v_1, v_2, \ldots, v_b, u)$ is linearly independent. Notice now that $(v_1, v_2, \ldots, v_b, u)$ has $b+1$ elements and is linearly independent. This contradicts the statement of Theorem 3.3 as B spans V and B has only b elements. Therefore L must be a basis for V. □

Theorem 3.13. *Let B be a finite-dimensional vector space. Every spanning list of length $\dim V$ is a basis for V.*

Proof. Let V be a finite dimensional vector space with dimension b. Let S be a spanning list of length b. We proceed by contradiction. That is, we assume S is not a basis.

If S is not a basis, then it must be that S is not linearly independent. Therefore S is linearly dependent. By Theorem 3.8 we can remove elements of S so that the resulting list has the same span, but is linearly independent. The resulting list is necessarily a basis as it spanning and is linearly independent. However, the resulting list has fewer than b entries. This contradicts the statement of Theorem 3.10 as $\dim V = b$.

Therefore S is a basis for V. □

Before we conclude our work for the week (and it has been a big week!), let us take a moment to think about how far we have come in a few short weeks of this course.

In Module 1, we defined an abstract mathematical object: a vector space. Vector spaces behave the same as vectors in $\mathbf{R}^n$, except the vectors inside a vector space may not be lists of elements of $\mathbf{R}^n$. Using only the definition of vector space and our mathematical wits, we have managed to conclude something quite surprising.

If V is a vector space and V is spanned by some list of vectors, then there exists a list B so that every vector in V can be expressed as a unique linear combination of vectors in B. Moreover, every such list B has the same length, and every list of this length that is either spanning or linearly independent has this property.

This single sentence more-or-less sums up what we have learned in this module:

- If V is a vector space and V is spanned by some list of vectors (*if V is finite dimensional*), then
- there exists a list B so that every vector in V can be expressed as a unique linear combination of vectors in B (*V has a basis, B. Every element can be expressed a unique linear combination of elements of B.*)
- Moreover, every such list B has the same length, (*Every basis has the same length. This length is the dimension of V*) and
- every list of this length that is either spanning or linearly independent has this property (*Any spanning or linearly independent list of length* $\dim V$ *is a basis*)

Finite-dimensional vector spaces can be built from some small collection of elements of the vector space. And no matter which collection of building blocks we use, we will always have the same quantity of building blocks. In $\mathbf{R}^2$ we can take $(1, 0)$ and $(0, 1)$ as our building blocks. We could also take $(3, 1)$ and $(1, 0)$. In $\mathcal{P}_5(\mathbf{R})$ we can take $1, x, x^2, x^3, x^4, x^5$ as our building blocks. Much like in $\mathbf{R}^2$, these aren't our only choice of building blocks.

In this way, we can think of a vector space as being the entire collection of objects that can be built from some fixed set of building blocks. The beauty of abstraction here is that this fact is true for every finite dimensional vector space. If some collection of mathematical objects satisfies the definition of finite-dimensional vector space, then necessarily these building blocks exist.

3.3 Test your Understanding

1. Determine the dimension of $\mathcal{M}_{2\times 3}(\mathbf{R})$, the real vector space of 2×3 matrices.
2. Briefly explain how you know $\mathcal{M}_{m\times n}(\mathbf{R})$ has dimension mn.

3.3 Test your Understanding Solution

1. To find the dimension of $\mathcal{M}_{2\times 3}(\mathbf{R})$ it suffices to count the number of elements in a basis for $\mathcal{M}_{2\times 3}$. Consider the list:

$$B = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \right)$$

   Since

$$a_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + a_5 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + a_6 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{bmatrix}$$

   we have $span(B) = \mathcal{M}_{2\times 3}$. To have

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

   we must have $a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = 0$. Therefore B is linearly independent.

   Since B is spanning and linearly independent, it follows that B is a basis. Since $\mathcal{M}_{2\times 3}(\mathbf{R})$ has a basis with 6 elements, it follows that $\mathcal{M}_{2\times 3}(\mathbf{R})$ has dimension 6.

2. Similar to the work in the previous part, we have form a basis for $\mathcal{M}_{m\times n}(\mathbf{R})$ by taking the list of $m \times n$ matrices that have a single entry as 1 and all other entries as 0. There are mn such matrices, therefore $\mathcal{M}_{m\times n}(\mathbf{R})$ has dimension mn

---

4 Linear Maps Part I (LADR 3A/3B)

Learning Incomes.

- *recall the definitions of injective, surjective and bijective functions*
- *recall the definitions of domain and codomain of a function*
- *recall the definition of subspace and the theorem that saves us time in checking if a subset is a subspace.*

Learning Outcomes.

- *Understand the meaning of the term* ***linear map*** *and be able to see how this definition works for a variety of familiar contexts.*
- *Be able to find the image of any element of the domain of a linear map given the image of each element of a basis for the domain.*
- *Understand the relationship between null* $T = \{0\}$ *and injectivity*
- *Understand the connection between null space of a matrix and null space of a linear map.*
- *Understand the connection between column space of a matrix and range of a linear map.*

Newly Defined Terms and Notation.

- *linear map,* Tu*, null space of a linear map, range of a linear map*

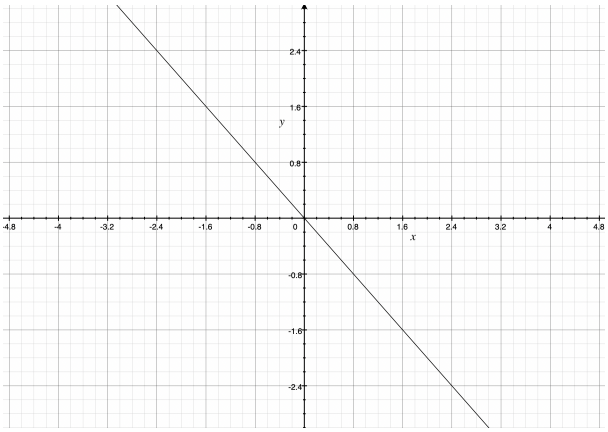
We have spent the first three modules of this course introducing vector spaces. For most concepts, our strategy so far has been to think about what the concept means in $\mathbf{R}^n$ and then realize that we can still make sense of it when we switch from $\mathbf{R}^n$ to some other vector space. For example, knowing what a basis in $\mathbf{R}^n$ helps us immensely when we come to consider this concept in an arbitrary vector space V.

In Module 4 we will begin a study of a concept that is not part of the main focus of MATH164 – linear maps. In this context, the word *map* is a synonym for *function*. Our time spent in calculus courses have made us very familiar with *linear maps*, but not in the way we might expect.

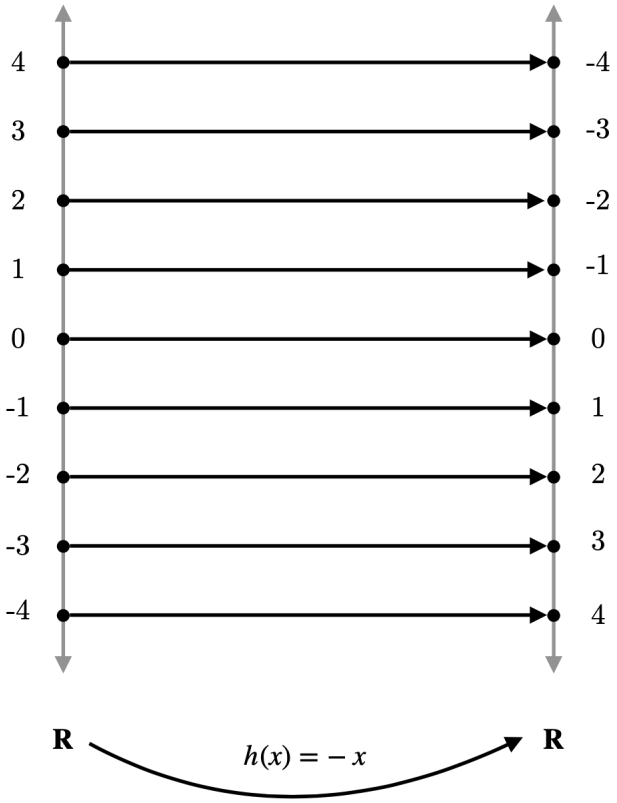
In this module we study linear maps whose domain and codomain are vector spaces. Realizing that $\mathbf{R}$ is a vector space helps us use some of our intuition from the study of linear functions from $\mathbf{R}$ to $\mathbf{R}$ as we study linear functions of vector spaces.

We begin with an example.

Consider the function $h : \mathbf{R} \to \mathbf{R}$ so that $h(x) = -x$ for every $x \in \mathbf{R}$. Asked to draw a picture of this function, I expect that most of us would draw this:



Let us take a moment to think about a different way of visualizing this function:



(There are lots of arrows missing here! We have just drawn in some of them so we can make

sense of it.)

On the left we have a copy of $\mathbf{R}$, the real line. This copy of $\mathbf{R}$ is the domain of our function, h. On the right we have a second copy of $\mathbf{R}$. This copy of $\mathbf{R}$ is the codomain of our function, h.

Our function matches up elements of the domain with elements of the codomain so that a value r in the domain gets matched up with the value $-r$ in the codomain. For example, the arrow from -1 to 1 corresponds to the statement $h(-1) = 1$.

Thinking about these two copies of $\mathbf{R}$ as being the same thing, our function has transformed our copy of $\mathbf{R}$ by *flipping* all of the values. That is to say, our function moves the point r to the point $-r$. When we transform our real line using the function h the object that results is another copy of the real line.

For every $s \in \mathbf{R}$ there exists $r \in \mathbf{R}$ so that $h(r) = s$. Every element of the codomain is the image of some element of the domain. In other words, this function is surjective. This function is also injective. If $r_1 \neq r_2$, then $h(r_1) \neq h(r_2)$. Together these two facts tell us that h is a bijection.

This function also has some more features for us to notice – it preserves sums and scalar multiples. For example $h(1) + h(3) = -1 + -3 = -4 = h(4) = h(1 + 3)$

For all $y, z \in \mathbf{R}$, we have $h(y + z) = -(y + z) = -y + -z = h(y) + h(z)$. And for all $\lambda \in \mathbf{R}$ we have $h(\lambda y) = -(\lambda y) = \lambda(-y) = \lambda h(y)$.

Functions that preserve sums and scalar multiples are already familiar to us, even if we haven't noticed it before. Consider the function

$$D : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_1(\mathbf{R})$$

so that $D(f) = \frac{d}{dx} f$. For example

$$D(x^2 + 1) = \frac{d}{dx}(x^2 + 1) = 2x$$

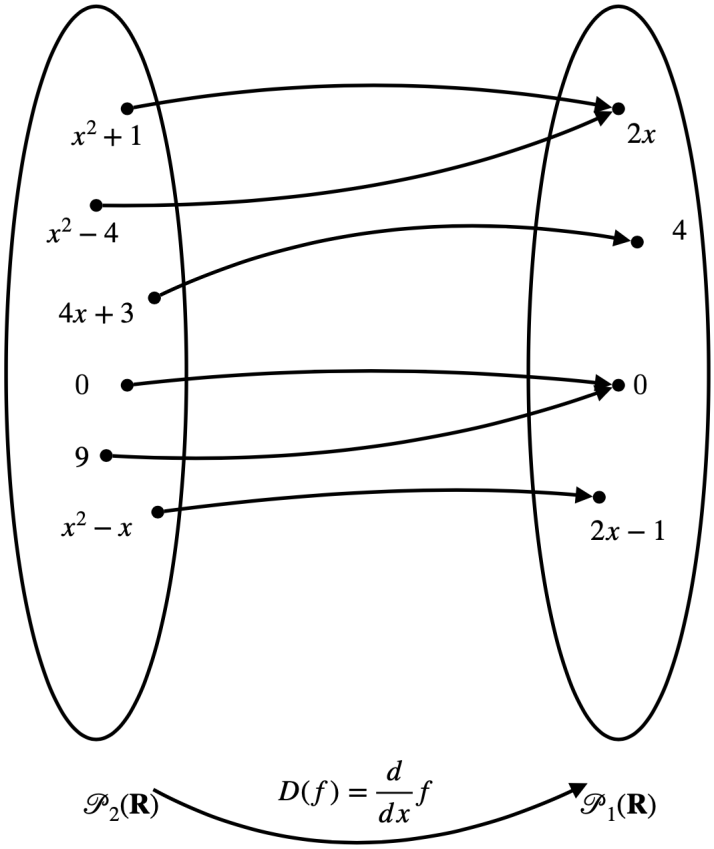
Just like our function above, D preserves sums and scalar multiples:

$$\begin{aligned} D(a_2x^2 + a_1x + a_0) &= \frac{d}{dx}(a_2x^2 + a_1x + a_0) \\ &= 2a_2x + a_1 + 0 \\ &= \frac{d}{dx}(a_2x^2) + \frac{d}{dx}(a_1x) + \frac{d}{dx}(a_0) \\ &= D(a_2x^2) + D(a_1x) + D(a_0) \end{aligned}$$

and

$$\begin{aligned} D(\lambda(a_2x^2 + a_1x + a_0)) &= \frac{d}{dx}(\lambda(a_2x^2 + a_1x + a_0)) \\ &= \frac{d}{dx}(\lambda a_2x^2 + \lambda a_1x + \lambda a_0) \\ &= 2\lambda a_2x + \lambda a_1 + \lambda 0 \\ &= \lambda(2a_2x + a_1 + 0) \\ &= \lambda D(a_2x^2 + a_1x + a_0) \end{aligned}$$

Though drawing a picture of this function, like we did with h above, is a little bit less intuitive, it is still worth thinking about.



(Again, there a lots of arrows missing here.)

This time, however, we see that our function is not injective – there are lots of functions that have the same derivative. For example every element of $\mathcal{P}_2(\mathbf{R})$ of the form $0x^2 + 0x + c$ maps to the zero function under D.

In Module 4 we begin our study of functions from one vector space to another that preserve sums and scalar multiples. We call such functions *linear maps*. Though the definition of linear map is seemingly uninteresting, it belies deep structure that relates matrices and functions.

4.1 The Vector Space of Linear Maps (LADR 3A)

Definition 4.1. *Let V and W be vector spaces and let $T : V \to W$ be a function, We say T **is linear** when for all $u, v \in V$ and all $\lambda \in \mathbf{F}$ we have*

1. *$T(u + v) = T(u) + T(v)$; and*
2. *$T(\lambda u) = \lambda T(u)$*

*When T is a linear function we may say T **is a linear map.***

When T is a linear function, we may write Tv to refer to $T(v)$.

Aside 4.2. *At this point, sometimes denoting $T(v)$ as Tv will be a cause for confusion for us. With the notation $T(v)$ it is more clear to us that T is a function and v is in the domain of T. This information is harder for us to intuit when we denote the image of v as Tv. When written this way, Tv looks more like multiplication than like a function. Taking the product of an object denoted by a majuscule letter and one denoted by a minuscule letter perhaps reminds us the product of a matrix and a vector, Av. Though it may take us a few weeks to get to the payoff, this notational interpretation (evaluating a function looks like matrix multiplication) will be worth the momentary confusion we face in this module.*

Our two examples from the introduction, $h : \mathbf{R} \to \mathbf{R}$ and $D : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_1(\mathbf{R})$ are examples of linear maps. To further our understanding, we consider a non-example.

Let $A : \mathbf{R} \to \mathbf{R}$ so that $A(x) = |x|$. To confirm that A is not a linear map, we must show that it violates one of the two requirements in the definition of linear map.

For $u = 1$ and $v = -1$, we have $A(1) + A(-1) = |1| + |-1| = 1 + 1 = 2$. But $A(1 + (-1)) = T(0) = |0| = 0$. Since there is at least one pair u, v so that $A(u + v) \neq A(u) + A(v)$, the function A does not satisfy the definition of linear. Therefore A is not a linear map.

Though the definition of linear mapping tells us that we have $T(u + v) = T(u) + T(v)$ for all $u, v \in V$, in fact it extends to arbitrary quantities of terms. That is

$$T(v_1 + v_2 + \cdots + v_m) = Tv_1 + Tv_2 + \cdots + Tv_m$$

We omit the proof of this fact.

Aside 4.3. *If you are coming from a statistics background, you are likely quite familiar with a very practical example of linear maps. For any experiment, the set of all possible random variables defines a vector space. The expectation of a random variable $E[X]$ satisfies the following two properties:*

1. *$E[X + Y] = E[X] + E[Y]$; and*
2. *$E[cX] = cE[X]$.*

Expectation is a linear map whose domain is the collection of all random variables and whose codomain is $\mathbf{R}$.

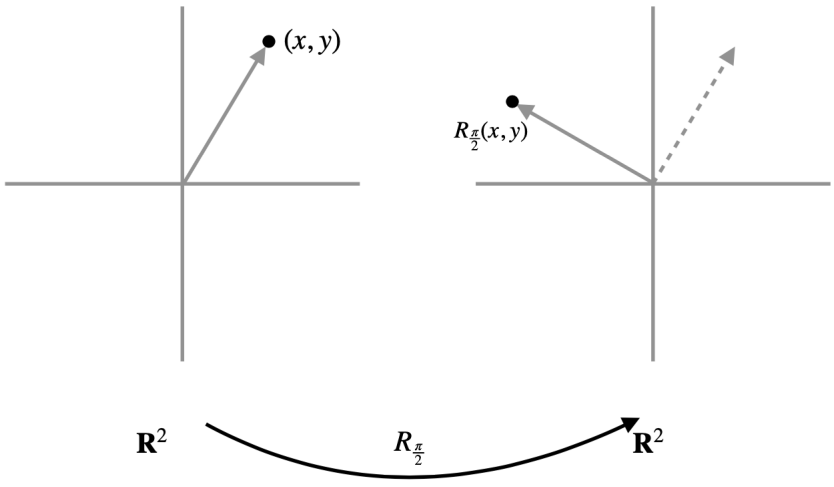
Our goal in the study of linear maps is to understand their structure. At this point, this goal isn't particularly meaningful and so let us jump ahead to the result we will end up with:

Linear maps and matrices are the same[3] thing.

Don't worry too much if this doesn't make much sense right now. Though we may have seen something about matrices and linear maps in MATH164, from our study so far we have no reason at all to suspect that we should be able to connect linear maps of arbitrary vector spaces with matrices.

From our work in Module 3 and our experiences in MATH164, we understand a basis for a vector space to be a set of building blocks that we can use to uniquely create any element of our vector space. We head out on our journey to understand the structure of linear maps by first thinking about linear maps and how they treat a basis.

To connect basis and linear map, let us consider another example, counter clockwise rotation in $\mathbf{R}^2$ in $\frac{\pi}{2}$ radians. Let $R_{\frac{\pi}{2}} : \mathbf{R}^2 \to \mathbf{R}^2$ so that $R(x, y)$ gives the resulting vector when (x, y) is rotated $\frac{\pi}{2}$ radians counter clockwise.



We won't take the time to verify that $R_{\frac{\pi}{2}}$ is linear, but be assured that it is.

[3]once we define very carefully what we mean when we say *same*

Let us consider the action of $R_{\frac{\pi}{2}}$ on a basis for $\mathbf{R}^2$. Consider the standard basis: $B = ((1,0),(0,1))$. Rotating each of these two vectors counterclockwise by $\frac{\pi}{2}$ radians results in the list $((0,1),(-1,0))$. Notably, this list is a basis for $\mathbf{R}^2$.

Consider now the basis $B' = ((1,1),(1,0))$. Rotating each of these vectors yields the list $((-1,1),(0,1))$, which again is a basis for $\mathbf{R}^2$.

Thinking back to our example, $h : \mathbf{R} \to \mathbf{R}$ given by $h(x) = -x$, when we drew our picture of h in the introductory section, we observed that h transformed $\mathbf{R}$ by *flipping* all of the values to produce a copy of $\mathbf{R}$. In this same vein, we can consider $R_{\frac{\pi}{2}}$ to have transformed $\mathbf{R}^2$ by rotating it counter clockwise by $\frac{\pi}{2}$ radians.

From these two examples, perhaps we suspect that the image of a basis of vector space of a linear map is another basis. Immediately, however, we can find a counter example to this statement.

Let U and V be vector spaces and consider the function $T : U \to V$ so that $Tu = 0$ for all $u \in U$. On Assignment 3 we will verify that T is linear. In this case, however, the image of a basis for U is a list where all entries are the zero vector. Any list that contains the zero vector cannot be linearly independent, and so we have not produced a basis.

Rather than the statement:

the image of a basis of vector space of a linear map is another basis

being true, what turns out the be true is the following:

When U and V are vector spaces of the same dimension, if B_U is a basis for U and B_V is a basis for V, then there is exactly one linear map T so that the image of B_U under T is B_V.

More precisely, we have:

Theorem 4.4. *Let U and V be vector spaces of dimension n. If $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_n)$ respectively are bases of U and V, then there is a unique linear map $T : U \to V$ so that*

$$Tu_i = v_i$$

for all $i \in \{1, 2, 3, \ldots, n\}$.

Before we dive in to this proof, let us contextualize in a pair of familiar real vector spaces, $\mathbf{R}^2$ and $\mathcal{P}_1(\mathbf{R})$. Both of these vector spaces have dimension 2. Our theorem requires us to have a basis for each of these vector spaces. Let us choose the almost least interesting basis we can think of for $\mathbf{R}^2$, $B_{\mathbf{R}^2} = (u_1, u_2) = ((1,1),(1,0))$. But to make our workings slightly

helpful for learning, let us choose a basis for $\mathcal{P}_1(\mathbf{R})$ that is slightly more interesting. Let $B_{\mathcal{P}_1(\mathbf{R})} = (v_1, v_2) = (x+1, x-1)$.

Our theorem tells us that there is precisely one linear map $T : \mathbf{R}^2 \to \mathcal{P}_1(\mathbf{R})$ so that

$$\begin{aligned} T(1,1) &= x+1 \\ T(1,0) &= x-1 \end{aligned}$$

Our theorem doesn't tell us anything else, only that this map T exists. So let's take a moment to see what we can learn about this linear map, T.

Consider the vector $u = (3,2)$. Since

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

we must have

$$\begin{aligned} T(3,2) &= T(2(1,1) + 1(1,0)) \\ &= T(2(1,1)) + T(1(1,0)) \\ &= 2(T(1,1)) + 1T(1,0) \\ &= 2(x+1) + 1(x-1) \\ &= 2x + 1 + x - 1 \\ &= 3x \end{aligned}$$

(Don't just quickly skip by the last set of equalities. Take some time to make sure you are convinced that each line follows from the previous one. These equalities follow from the fact that T is a linear map. Take a moment to review the definition of **linear** if these equalities confuse you.)

As B_U is a basis for $\mathbf{R}^2$, for any $u \in \mathbf{R}^2$, there exists scalars c_1 and c_2 so that

$$u = c_1(1,1) + c_2(0,1)$$

We can use this information to compute Tu as follows:

$$\begin{aligned} Tu &= T(c_1(1,1) + c_2(0,1)) \\ &= T(c_1(1,1)) + T(c_2(0,1)) \\ &= c_1T(1,1) + c_2T(0,1) \\ &= c_1(x+1) + c_2(x-1) \end{aligned}$$

The image of u with respect to T is the given by the linear combination of the elements of $B_{\mathcal{P}_1(\mathbf{R})}$ where the coefficients are the ones we would use if we wrote u as a linear combination

of the elements of $B_{\mathbf{R}^2}$. In turns out that this strategy works for any vector spaces U and V with the same dimension. That is, if $B_U = (u_1, u_2, \ldots, v_n)$ is a basis for U and $B_V = (v_1, v_2, \ldots, v_n)$ is a basis for V, then for

$$u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$$

we have

$$Tu = a_1v_1 + a_2u_2 + \cdots + a_nv_n$$

With this thought in mind, we turn to the proof of our theorem.

Proof. Let U and V be vector spaces of dimension n and let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_n)$ respectively be bases of U and V. To prove our theorem we want to show two things:

1. There exists a linear map $T : U \to V$ that satisfies $Tu_i = v_i$ for all for all $i \in \{1, 2, 3, \ldots, n\}$, and
2. the linear map T is the only linear map that has this property.

To define a function from U to V we must define the image of each element of U. Since B_U is a basis for U, for every $u \in U$ there exists scalars $c_1, c_2, \ldots, c_n$ so that

$$u = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

And so to define our function, we define the image of elements of the form:

$$c_1u_1 + c_2u_2 + \cdots + c_nu_n$$

Let us define $T : U \to V$ so that

$$T(c_1u_1 + c_2u_2 + \cdots + c_nu_n) = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

We first verify that $Tu_i = v_i$ for all $i \in \{1, 2, 3, \ldots, n\}$. When $c_i = 1$ and all other scalars are 0 we have:

$$T(c_1u_1 + c_2u_2 + \cdots + c_nu_n) = Tu_i$$

and

$$T(c_1u_1 + c_2u_2 + \cdots + c_nu_n) = c_1v_1 + c_2v_2 + \cdots + c_nv_n = v_i$$

Therefore $Tu_i = v_i$ for all for all $i \in \{1, 2, 3, \ldots, n\}$.

We now verify that T is linear.

For $x = b_1u_1 + b_2u_2 + \cdots + b_nu_n$ and $y = d_1u_1 + d_2u_2 + \cdots + d_nu_n$ we have

$$\begin{aligned} T(x + y) &= T\left((b_1 + d_1)u_1 + (b_2 + d_2)u_2 + \cdots + (b_n + d_n)u_n\right) \\ &= (b_1 + d_1)v_1 + (b_2 + d_2)v_2 + \cdots + (b_n + d_n)v_n \\ &= (b_1v_1 + b_2v_2 + \cdots + b_nv_n) + (d_1v_1 + d_2v_2 + \cdots + d_nv_n) \\ &= Tx + Ty \end{aligned}$$

For any $\lambda \in \mathbf{F}$ we have

$$\begin{aligned} T(\lambda x) &= T\left(\lambda(b_1u_1 + b_2u_2 + \cdots + b_nu_n)\right) \\ &= T\left(\lambda b_1u_1 + \lambda b_2u_2 + \cdots + \lambda b_nu_n\right) \\ &= \lambda b_1v_1 + \lambda b_2v_2 + \cdots + \lambda b_nv_n \\ &= \lambda\left(b_1v_1 + b_2v_2 + \cdots + b_nv_n\right) \\ &= \lambda T(x) \end{aligned}$$

Since T satisfies both parts of the definition of linear, necessarily T is linear.

We now show T is the only linear map that satisfies $Tu_i = v_i$ for all $i \in \{1, 2, 3, \ldots, n\}$. To show that T is the only such linear map, we consider another linear map $T' : U \to V$ and show that in fact $Tu = T'u$ for all $u \in U$. For this end, suppose $T' : U \to V$ so that $T'u_i = v_i$ for all $i \in \{1, 2, 3, \ldots, n\}$. Let $u = c_1u_1 + c_2u_2 + \cdots + c_nu_n$. Since T' is linear we have:

$$\begin{aligned} T'u &= T'(c_1u_1 + c_2u_2 + \cdots + c_nu_n) \\ &= T'(c_1u_1) + T'(c_2u_2) + \cdots + T'(c_nu_n) \\ &= c_1T'(u_1) + c_2T'(u_2) + \cdots + c_nT'(u_n) \\ &= c_1v_1 + c_2v_2 + \cdots + c_nv_n \\ &= Tu \end{aligned}$$

Therefore T and T' are the same function. And so T is the only linear map that satisfies $Tu_i = v_i$ for all $i \in \{1, 2, 3, \ldots, n\}$. □

Since every finite dimensional linear map has a basis, necessarily we can find a linear map between any pair of finite dimensional vector spaces with the same dimension.

Corollary 4.5. *For every pair U and V of vector spaces with the same dimension, there exists a linear map $T : U \to V$.*

As linear maps are functions, we can define addition and scalar multiplication in the same way we define it for any function. For linear maps $T, S : U \to V$ and $\lambda \in \mathbf{F}$, define

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

With this definition of addition and scalar multiplication, we have the following:

Theorem 4.6. *For vector spaces U and V, the set of all linear maps from U to V is a vector space.*

It takes a little bit of work to actually prove this theorem is true, but the proof isn't particularly interesting, and so we omit it. *(Yes, I realize that by asserting this proof isn't interesting I am loosely implying that some of these proofs are interesting... – Chris)*

In our journey to understand how we can say that linear maps are matrices in disguise, this result perhaps should not surprise us too much. Indeed, the set of matrices of a fixed dimension form a vector space. And so if linear maps and matrices are the same[4], then so should linear maps form a vector space. Similarly, if linear maps and matrices are the same[5], then perhaps there are things we know about matrices that have analogues for linear maps. Indeed, in the following section we consider defining null space and column space for linear maps.

[4]see footnote 3
[5]see footnote 4

4.1 Test Your Understanding

1. Let $T : \mathbf{R} \to \mathbf{R}$ so that $Tx = x^2 + 2x + 1$ for all $x \in \mathbf{R}$. Using the definition of linear map, show that T is not linear.

2. Let $U = \mathcal{M}_{2\times 2}$ and let $V = \mathbf{R}^4$. Let

$$B_U = \left(\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

be a basis for U and let $B_V = (e_1, e_2, e_3, e_4)$ be the standard basis for V. Let T be the unique linear map so that $Tu_i = v_i$ for all $1 \in \{1, 2, 3, 4\}$. Compute

$$T \begin{bmatrix} 4 & 1 \\ 3 & 0 \end{bmatrix}$$

4.1 Test Your Understanding Solution

1. We have $T(1) = 4$ and $T(-1) = 0$, but $T(1 + -1) = T(0) = 1$. Since $T(1) + T(-1) \neq T(1 + -1)$, the function T is not linear.

2. The proof of Theorem 4.4 tells us that the image of $u \in U$ with respect to T is the given by the linear combination of the elements of B_V where the coefficients are the ones we would use if we wrote u as a linear combination of the elements of B_U.

   We have
   $$\begin{bmatrix} 4 & 1 \\ 3 & 0 \end{bmatrix} = 2 \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + 1 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 0 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

   Therefore

   $$T \begin{bmatrix} 4 & 1 \\ 3 & 0 \end{bmatrix} = 2e_1 + e_2 + 3e_3 + 0e_4 = (2, 1, 3, 0)$$

4.2 Null Space and Range of a Linear Map (LADR 3B)

Returning to some familiar ground from MATH164 (and from Module 2) recall the definition of null space for a matrix A:

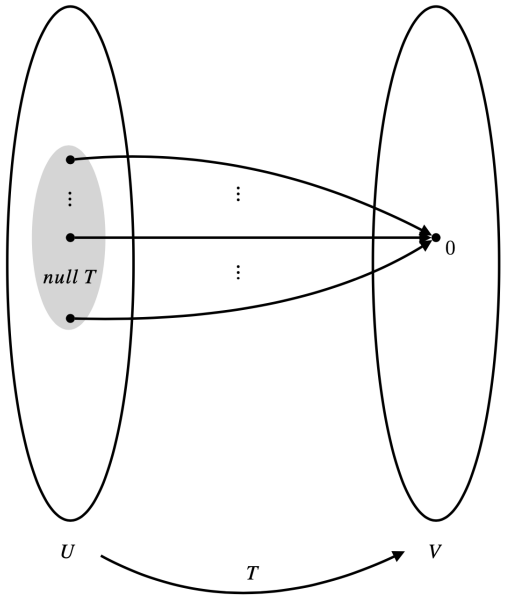
$$null\ A = \{u \mid Au = 0\}$$

In Module 2 we showed that when A is an $n \times m$ matrix with elements in $\mathbf{R}$, then $null\ A$ is a subspace of $\mathbf{R}^m$. We showed this by checking that $null\ A$ was non-empty and closed with respect to addition and scalar multiplication.

Informally, we can think the null space of A to be the set of vectors of $\mathbf{R}^m$ that go to zero when they are multiplied by A. We define null space for a linear transformation analogously.

Definition 4.7. *Let U and V be vector spaces and let $T : U \to V$ be a linear map. The* ***null space of*** *T is the set of elements of U whose image is 0. That is*

$$null\ T = \{u \in U \mid Tu = 0\}$$



Recall our differentiation example from our introductory section:. Consider the function

$$D : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_1(\mathbf{R})$$

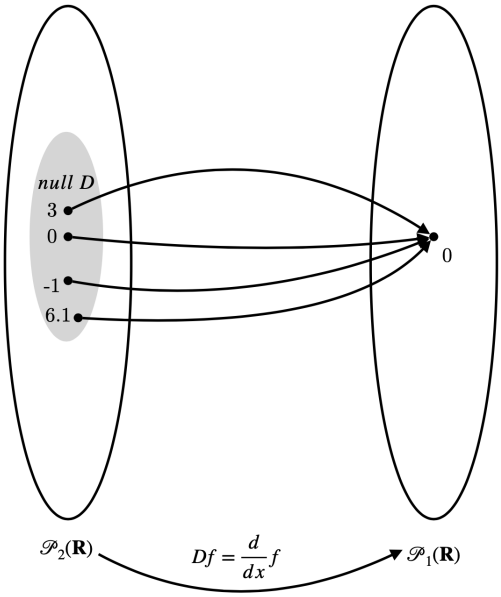
so that $Df = \frac{d}{dx}f$.

The null space of D is the set of polynomials of degree at most 2 whose derivative is 0.

$$null\ D = \{f \in \mathcal{P}_2(\mathbf{R}) \mid Df = 0\} = \{f \in \mathcal{P}_2(\mathbf{R}) \mid \frac{d}{dx}f = 0\}$$

A polynomial has derivative equal to zero if and only if it is a constant function. Therefore

$$null\ D = \{f \mid f(x) = c,\ \ c \in \mathbf{R}\}$$



Just as the null space of an $n \times m$ matrix is a subspace of $\mathbf{R}^m$, so too is null space of a linear map a subspace of its domain. For a linear map $T : U \to V$, we have $null\ T \subseteq U$. We apply Theorem 2.5 to show $null\ T$ is a subspace of U. Take a moment to recall the statement of Theorem 2.5.

.
.
.
.
.
.
.
.
.

Notice

$$T(0) = T(0 + 0) = T(0) + T(0)$$

Adding the additive inverse of $T(0)$ in V to both sides, gives $T(0) = 0$. Therefore $0 \in null\ T$ and so $null\ T$ is non-empty.

Let $x, y \in \mathit{null}\ T$. Since T is linear we have

$$T(x+y) = T(x) + T(y) = 0 + 0 = 0$$

Therefore $x + y \in \mathit{null}\ T$. And so, $\mathit{null}\ T$ is closed with respect to addition.

Let $\lambda \in \mathbf{F}$. Since T is linear we have

$$T(\lambda x) = \lambda T(x) = \lambda 0 = 0$$

Therefore $\lambda x \in \mathit{null}\ T$. And so, $\mathit{null}\ T$ is closed with respect to scalar multiplication.

By Theorem 2.5, it follows that $\mathit{null}\ T$ is a subspace of U.

Aside 4.8. *There is something slightly sneaky happening when we write $x + y \in \mathit{null}\ T$. In our work above we could have just as easily started by assigning the label* `z` *to the result of the sum $x + y$ (i.e., let $z = x + y$) and then subsequently showing $z \in \mathit{null}\ T$. By writing $x + y \in \mathit{null}\ T$ we are assigning label* `x+y` *to the element that results from adding x and y in U.*

Returning to our example from the introductory section, recall the linear map $f : \mathbf{R} \to \mathbf{R}$ given by $h(x) = -x$. In this case we have $\mathit{null}\ h = \{0\}$. The value $0 \in \mathbf{R}$ is the only one for which $h(x) = 0$. In this case, the null space of the linear map is not terribly interesting, it consists of just the zero element. Linear maps whose null space is exactly the zero vector can be fully classified:

Theorem 4.9. *Let $T : U \to V$, we have $\mathit{null}\ T = \{0\}$ if and only if T is injective.*

Despite the fact that familiarity with word *injective* is part of the learning incomes for this module, let us take a moment to remind ourselves of the definition of injective:

Definition. *Let A and B be sets and let f be a function from A to B. We say f* ***is an injection*** *when each element of the domain has a unique image. When f is an injection we say that h* ***is injective***

To check if a function f is injective, we consider all possible pairs x_1, x_2 in the domain with $x_1 \neq x_2$ and show we have $f(x_1) \neq f(x_2)$. For example, our function $h : \mathbf{R} \to \mathbf{R}$ given by $h(x) = -x$ is injective: if $x_1 \neq x_2$, then we can conclude $h(x_1) \neq h(x_2)$ as $-x_1 \neq -x_2$.

On the other hand, our linear map $D : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_1(\mathbf{R})$ so that $Df = \frac{d}{dx}f$ is not injective. We can find pairs of functions in $\mathcal{P}_2(\mathbf{R})$ that have the same derivative. For example, $D(x^2+1) = D(x^2+6)$. Since two elements of the domain, $\mathcal{P}_2(\mathbf{R})$ have the same image, the linear map is not injective.

Returning now to the statement of Theorem 4.9, let us first consider an injective linear map $T : U \to V$. From our work above, we have $T(0) = 0$, therefore $0 \in \mathit{null}\ T$. Since T is injective, for any $u \in U$ with $u \neq 0$ we have $T(u) \neq T(0)$. Therefore, for every $u \in U$ with $u \neq 0$ we have $Tu \neq 0$. Therefore, for every $u \in U$ with $u \neq 0$ we have $u \notin \mathit{null}\ T$. And so $\mathit{null}\ T = \{0\}$.

Consider now a linear map $T' : U \to V$ so that T' is not injective. Therefore there exist $u_1, u_2 \in U$ with $u_1 \neq u_2$, but $T'u_1 = T'u_2$. And so

$$0 = T'u_1 - T'u_2 = T'(u_1 + (-u_2))$$

Since $u_1 \neq u_2$ we have $u_1 + (-u_2) \neq 0$. Since $T'(u_1 + (-u_2)) = 0$ we have $u_1 + (-u_2) \in \mathit{null}\ T'$. Therefore $\mathit{null}\ T' \neq \{0\}$.

Together these two arguments (the previous two paragraphs) tell us that the null space of a linear map contains only 0 exactly when the linear map is injective. In other words, we have $\mathit{null}\ T = \{0\}$ if and only if T is injective.

On Assignment 2 you are asked to show that the column space of an $n \times m$ matrix is a subspace of $\mathbf{R}^n$. The question prompts you to use a similar method as we did in Module 2 to show that the null space of an $n \times m$ matrix is a subspace of $\mathbf{R}^m$. That our work for showing the null space of a matrix is a subspace translated almost exactly to showing that the null space of a linear map is a subspace, perhaps suggests that the work that shows the column space of a matrix is a subspace will also show that the column space (?) of a linear map is a subspace.

But wait – there are no columns anywhere in a linear map? What should column space mean in the context of a linear map?

One way to think about the meaning of the column space of a matrix (and, I expect how it was introduced to us in MATH164) is

> The column space, *col* A, of a matrix is the span of the columns of the matrix.

Since there are no columns anywhere in sight in a linear map, let us massage this statement a little bit to get something more useful. Let A be an $n \times m$ matrix and let $v_1, v_2, \ldots v_m$ denote the columns of A.

$$A = [v_1\ v_2\ \cdots\ v_m]$$

Since A is an $n \times m$ matrix, each of the vectors v_i have n entries. That is, they are elements of $\mathbf{R}^n$

From our statement above we have

$$\mathit{col}\ A = \mathit{span}(v_1, v_2, \ldots, v_m) = \{c_1v_1 + c_2v_2 + \cdots + c_mv_m \mid c_1, c_2, \ldots, c_m \in \mathbf{R}\}$$

Consider the vector

$$c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

and the product Ac.

By definition of matrix multiplication we have

$$Ac = c_1 v_1 + c_2 v_2 + \cdots + c_m v_m$$

And so

$$col\ A = span(v_1, v_2, \ldots, v_m) = \{c_1 v_1 + c_2 v_2 + \cdots + c_m v_m \mid c_1, c_2, \ldots, c_m \in \mathbf{R}\} = \{Ac |\ c \in \mathbf{R}^m\}$$

The column space of A is the set of all the resulting elements of $\mathbf{R}^n$ we can obtain by multiplying A by all possible vectors in $\mathbf{R}^m$. In other words,

$$col\ A = \{v \mid \text{there exists } c \in \mathbf{R}^m \text{ so that } Ac = v\}$$

Naively swapping A for a linear map $T : U \to V$ gives:

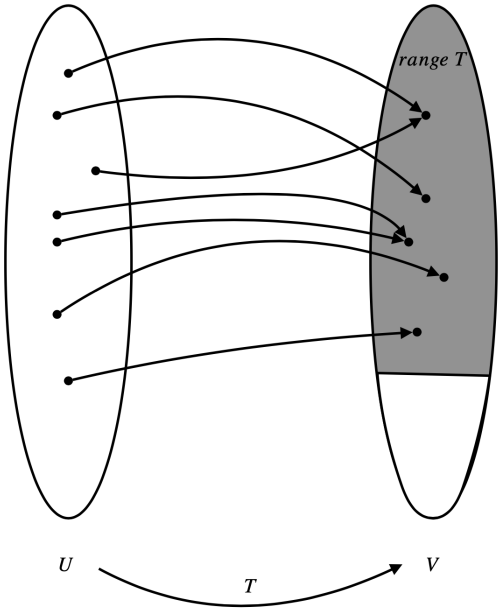
$$col\ T = \{v \in V \mid \text{there exists } c \in U \text{ so that } Tc = v\}$$

This definition seems plausible, though calling it the column space doesn't really much sense anymore.

Definition 4.10. *Let U and V be vector spaces and let $T : U \to V$ be a linear map. The* ***range of*** *T is the set of elements of V that have a pre-image in U with respect to T. That is*

$$range\ T = \{v \in V \mid there\ exists\ u \in U\ so\ that\ Tu = v\}$$

Another way to understand the meaning of range is as the elements of the codomain that are actually mapped to by the linear map.



For $h : \mathbf{R} \to \mathbf{R}$ given by $h(x) = -x$ we have

$$range\ h = \{v \in \mathbf{R} \mid \text{there exists } u \in \mathbf{R} \text{ so that } hu = v\}$$

For every $v \in \mathbf{R}$ we have $h(-v) = v$. Therefore each element of the codomain, $\mathbf{R}$, has a pre-image in the domain, $\mathbf{R}$.

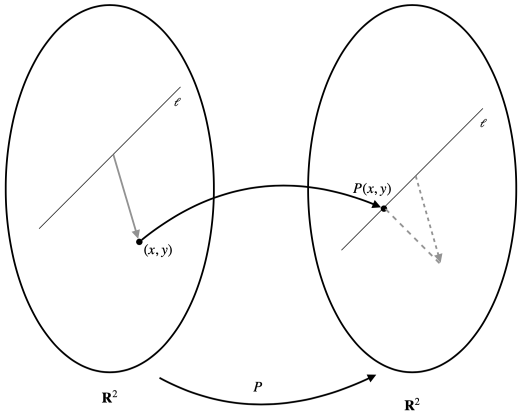
$$range\ h = \{v \in \mathbf{R} \mid \text{there exists } u \in \mathbf{R} \text{ so that } hu = v\} = \mathbf{R}$$

Similarly for $D : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_1(\mathbf{R})$ given by $Df = \frac{d}{dx}f$, we have the same behaviour. For every $g \in \mathcal{P}_1(\mathbf{R})$ there exists $f \in \mathcal{P}_2(\mathbf{R})$ so that $Df = g$. (I.e, every polynomial has an anti-derivative.) And so

$$range\ D = \{g \in \mathcal{P}_1(\mathbf{R}) \mid \text{there exists } f \in \mathcal{P}_2(\mathbf{R}) \text{ so that } Df = g\} = \mathcal{P}_1(\mathbf{R})$$

To see some different behaviour, let us consider a new example.

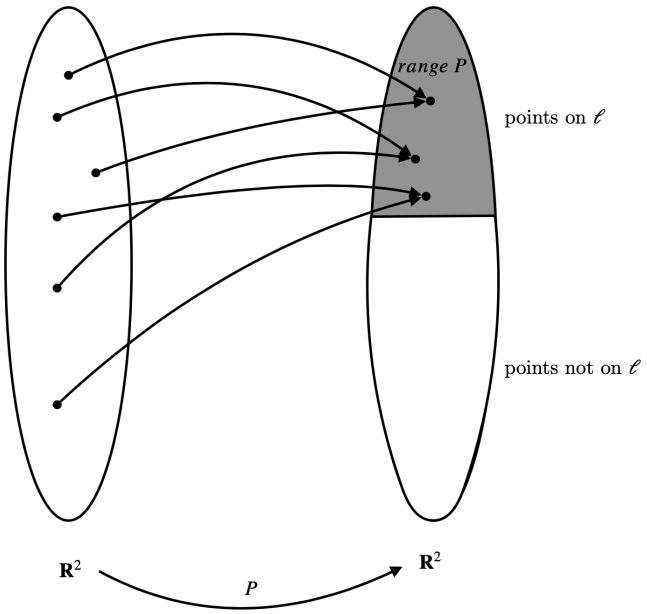
Let ℓ be the line in $\mathbf{R}^2$ that passes through the origin and the point $(1, 1)$. Let $P : \mathbf{R}^2 \to \mathbf{R}^2$ be the function that projects vectors onto ℓ. That is, for $(x, y) \in \mathbf{R}^2$, $P(x, y)$ is the point on ℓ that is closest to (x, y).



We won't take the time to check that P is linear; delving in to the required notation isn't worth the trouble.

Let us take a few moments to think about the set *range* P. The set *range* P is the set of elements of the codomain of P that are the image of some element of the domain of P.

For a point p on ℓ, any point on the line through p that is perpendicular to ℓ will map to p under P. And so every point on ℓ is the image of some element of the domain. And further since we are projecting onto ℓ, only the points on ℓ can be the image of elements of the domain. Therefore *range* $P = \ell$. In this example the codomain of P, $\mathbf{R}^2$, and the range of P, ℓ, are not the same.



Thinking back to our example, $h : \mathbf{R} \to \mathbf{R}$ given by $h(x) = -x$, when we drew our picture of h in the introductory section, we observed that h transformed $\mathbf{R}$ by *flipping* all of the values to produce a copy of $\mathbf{R}$. In this same vein, we can consider P to have transformed $\mathbf{R}^2$ into a copy of ℓ.

On Assignment 2, you were asked to show the column space of a matrix is a subspace. Unsurprisingly, the same fact turns out to be true for the range of a linear map.

Aside. *Depending on when you read these notes, you are either going to be pleased or annoyed by what is about the happen. The work we do subsequently is almost exactly the solution to question 4 on Assignment 2. If you are reading this ahead of the Assignment 2 due date (Friday February 12), feel free to adapt the work below to be a solution to question 2.*

For a linear map $T : U \to V$, we have *range* $T \subseteq V$. We apply Theorem 2.5 to show *range* T is a subspace of V. Take a moment to recall the statement of Theorem 2.5.

.
.
.
.
.
.
.
.
.

Notice

$$T(0) = T(0 + 0) = T(0) + T(0)$$

Adding the additive inverse of $T(0)$ in V to both sides, gives $T(0) = 0$. Therefore $0 \in$ *range* T and so *range* T is non-empty.

Let $v_1, v_2 \in$ *range* T. Since $v_1, v_2 \in$ *range* T, there exists u_1 and u_2 so that $T(u_1) = v_1$ and $T(u_2) = v_2$. Since T is linear we have

$$T(u_1 + u_2) = T(u_1) + T(u_1) = v_1 + v_2$$

Since $T(u_1 + u_2) = v_1 + v_2$ we have $v_1 + v_2 \in$ *range* T. And so, *range* T is closed with respect to addition.

Let $\lambda \in \mathbf{F}$. Since T is linear we have

$$T(\lambda u_1) = \lambda T(u_1) = \lambda v_1$$

Therefore $\lambda v_1 \in$ *range* T. And so, *range* T is closed with respect to scalar multiplication. By Theorem 2.5, it follows that *range* T is a subspace of V.

We end our work in Module 4 with a quick preview of what is to come in Module 5. In MATH164 you were likely introduced to some form of the *Rank-Nullity Theorem*:

Theorem 4.11 (The Rank Nullity Theorem)**.** *For an $n \times m$ matrix A we have*

$$n = rank\ A + nullity\ A$$

In this context, *rank* A denotes to the dimension of the column space of A and *nullity* A denotes the dimension of the null space of A. Translating this theorem into one for linear maps, we arrive at the following result.

Theorem (The Fundamental Theorem of Linear Maps)**.** *Let U be a finite-dimensional vector space and let V be a vector space. For every linear map $T : U \to V$ we have*

$$\dim\ U = \dim range\ T + \dim null\ T$$

In Module 5 we provide a proof of this theorem and explore some of its consequences on the existence of injective and surjective linear maps.

4.2 Test Your Understanding 4.2

1. Let $f(x) = x^2 - 4$. Show that f is not injective.
2. In our discussion of range, we considered a linear map P that projects vectors on to a line ℓ. What is the null space of this linear map?
3. In our proof that *null* T is a subspace of U, we have the sentence

   $$\textit{For any } u \in U \textit{ we have } T(0) = T(0u) = 0T(u) = 0.$$

   The notation 0 appears four times in this sentence. The first appearance refers to the additive identity in U. What do the other appearances refer to?

4.2 Test Your Understanding Solutions

1. We have $f(2) = f(-2) = 0$. Therefore there exists a pair $x_1, x_2 \in \mathbf{R}$ so that $f(x_1) \neq f(x_2)$. Therefore f is not injective.

2. By definition we have

$$null\ P = \{u \in \mathbf{R}^2 \mid Tu = 0\}$$

In $\mathbf{R}^2$, 0 refers to the origin, $(0, 0)$. For a point p on ℓ, any point on the line through p that is perpendicular to ℓ will map to p under P. Therefore points on the line through $(0, 0)$ that is perpendicular to ℓ will map to $(0, 0)$ under P. Therefore

$$null\ P = \{u \mid u = t(1, -1), t \in \mathbf{R}\}$$

3. Recall the sentence:

$$For\ any\ u \in U\ \ we\ have\ T(0) = T(0u) = 0T(u) = 0.$$

The first appearance of 0 refers to the additive identity in U. . The second and third appearance of 0 refer to the scalar 0 in $\mathbf{F}$. The final appearance of 0 refers to the additive identity in V.

---

5 Linear Maps Part II (LADR 3B/3C)

Learning Incomes.

- *Understand the definition of linear map, null space of a linear map and range of a linear map.*
- *Recall the meaning of the words injective, surjective and bijective.*

Learning Outcomes.

- *Understand the statement and proof of the Fundamental Theorem Linear Maps.*
- *Be able to construct the matrix of a linear map.*
- *Be able to construct the matrix of an element of a vector space with respect to a give basis.*
- *Have at least a fuzzy understanding of what same means when we talk about matrices and linear maps being the same.*

Newly Defined Terms and Notation.

- *matrix of a linear map,* $\mathcal{M}(T)$*, matrix of a vector,* $\mathcal{M}(v)$

In Module 4, amongst other things we extended our intuition for null space and column space of a matrix for linear maps. The resulting subspaces, the null space and the range, generalize the null space and the column space of matrices. That is, for a linear map $T : U \to V$ we have

$$\textit{null } T = \{u \mid Tu = 0\} \text{ and } \textit{range } T = \{v \in V \mid \text{there exists } u \in U \text{ so that } Tu = v\}$$

In MATH164, we learned about a relation between the null space of a matrix and the column space of a matrix:

Theorem (The Rank Nullity Theorem)**.** *For an* $n \times m$ *matrix* A *we have*

$$n = \textit{rank } A + \textit{nullity } A$$

Given that our goal in studying linear maps is to relate them to matrices, it stands to reason that the relationship between the respective dimensions of $\mathbf{R}^n$, the null space of a matrix and the column space of a matrix should extend to the dimensions of a vector space, a null space and a range of a linear map. That is, we can expect:

Theorem (The Fundamental Theorem of Linear Maps)**.** *Let U be a finite-dimensional vector space and let V be a vector space. For every linear map $T : U \to V$ we have*

$$\dim\ U = \dim range\ T + \dim null\ T$$

In reading the statement of this theorem, there are a few things to consider. Recall first our definition for dimension:

Definition. *Let W be a finite-dimensional vector space. The* ***dimension of*** *W is the number of elements in a basis for W. We denote this parameter as* $\dim W$*.*

We notice that **dimension** only has meaning when considering a finite dimensional vector space. The hypothesis of the statement of The Fundamental Theorem of Linear Maps requires only that U is finite dimensional. And so, we wonder, how do we know that each of *null* T and *range* T is finite dimensional (and thus have dimension)? Answering this question leads us to a proof of The Fundamental Theorem of Linear Maps. We undertake this work in Section 5.1.

Following this work, we return to our central question in the study of linear maps: what is the meaning of the word *same* when we write

Linear maps and matrices are the *same*.

In Section 5.2 we begin to answer this question by first giving a way to associate a matrix with a linear map. We then observe that operations on linear maps (addition and scalar multiplication) are precisely mimicked by corresponding operations for matrices (addition and scalar multiplication).

A key ingredient to all of our work in this module is the following result from Module 3.

Theorem 5.1 (Criterion for Basis)**.** *Let V be a finite dimensional vector space. A list $B = (v_1, v_2, \ldots, v_k)$ is a basis for V if and only if every $v \in V$ there is exactly one collection of scalars $(a_1, a_2, \ldots, a_k)$ so that*

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = v$$

For example, let $V = \mathcal{P}_2(\mathbf{R})$, $B_V = (3, x^2 + 1, x + 3)$ and $v = 2x^2 + x + 1$. The Criterion for Basis theorem tells us that there is exactly one collection of scalars a_1, a_2, a_3 so that

$$2x^2 + x + 1 = a_1(3) + a_2(x^2 + 1) + a_3(x + 3)$$

As we have done in our previous work, we can rearrange on the right side to get

$$2x^2 + x + 1 = a_2 x^2 + a_3 x + (3a_1 + a_2 + 3a_3)$$

A pair of polynomials are equal exactly when they have the same coefficients. And so to have equality here we must have

$$\begin{aligned} a_2 &= 2 \\ a_3 &= 1 \\ 3a_1 + a_2 + 3a_3 &= 1 \end{aligned}$$

Solving this system of three equations and three unknowns gives us the unique list of scalars (a_1, a_2, a_3) so that

$$a_1(3) + a_2(x^2 + 1) + a_3(x + 3)$$

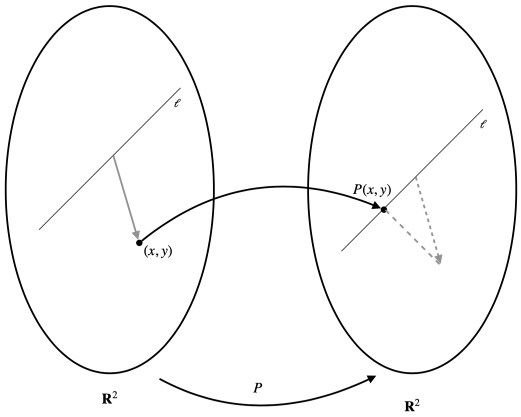
5.1 The Fundamental Theorem of Linear Maps (LADR 3B)

Theorem 5.2 (The Fundamental Theorem of Linear Maps)**.** *Let U be a finite-dimensional vector space and let V be a vector space. For every linear map $T : U \to V$ we have*

$$\dim\ U = \dim range\ T + \dim null\ T$$

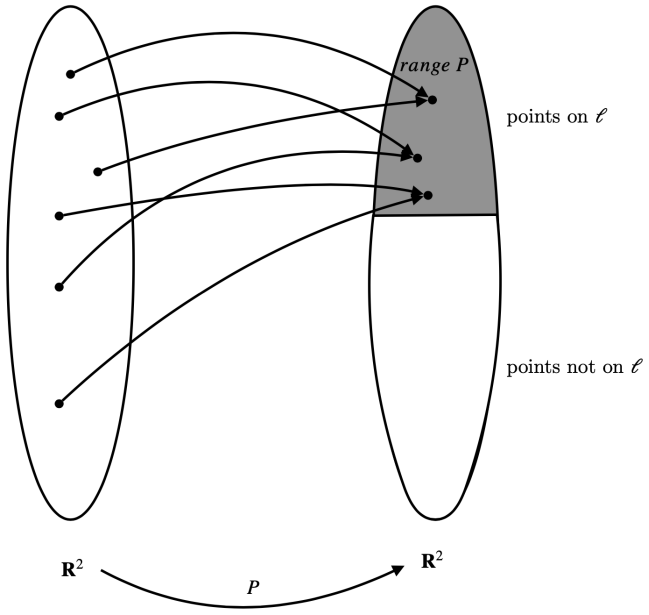
Before we dive into justifying the Fundamental Theorem of Linear Maps, let us first return to some of our examples from the previous module so that we can understand the statement of this theorem.

Let ℓ be the line in $\mathbf{R}^2$ that passes through the origin and the point $(1, 1)$. Let $P : \mathbf{R}^2 \to \mathbf{R}^2$ be the function that projects vectors onto ℓ. That is, for $(x, y) \in \mathbf{R}^2$, the image of (x, y) under P is the point on ℓ that is closest to (x, y).



Let us take a few moments to think about the set *range* P. The set *range* P is the set of elements of the codomain of P that are the image of some element of the domain of P.

For a point p on ℓ, any point on the line through p that is perpendicular to ℓ will map to p under P. And so every point on ℓ is the image of some element of the domain. And further, only the points on ℓ can be the image of elements of the codomain. Therefore *range* $P = \ell$.



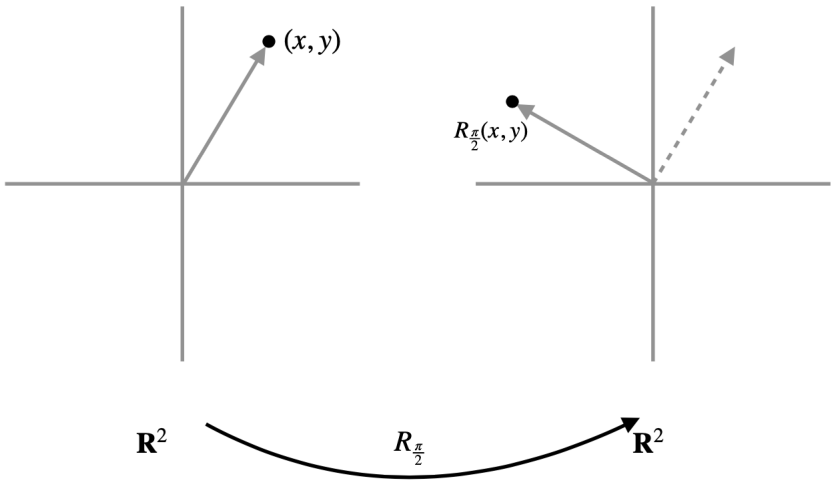
Since $\ell = span((1,1))$ we have $dim\ range\ P = 1$. Since $\mathbf{R}^2$ has dimension 2, the Fundamental Theorem of Linear Maps tells us that the dimension of the null space of P should be 1.

Following our work in Test Your Understanding 4.2, we found

$$null\ P = \{u \mid u = t(1,-1), t \in \mathbf{R}\}.$$

And so, $null\ P = span((1,-1))$. Therefore $\dim range\ P = 1$, as expected.

Returning to another example from the previous module, let $R_{\frac{\pi}{2}} : \mathbf{R}^2 \to \mathbf{R}^2$ so that $R_{\frac{\pi}{2}}(x,y)$ gives the resulting vector when (x,y) is rotated $\frac{\pi}{2}$ radians counter clockwise.



We have

$$range\ R_{\frac{\pi}{2}} = \{v \in \mathbf{R}^2 \mid \text{there exists } u \in \mathbf{R}^2 \text{ so that } R_{\frac{\pi}{2}}(u) = v\}$$

For any vector $v \in \mathbf{R}^2$, rotating v clockwise by $\frac{\pi}{2}$ radians results in a vector u so that $R_{\frac{\pi}{2}}(u) = v$. Therefore

$$range\ R_{\frac{\pi}{2}} = \mathbf{R}^2.$$

And so, dim $range\ R_{\frac{\pi}{2}} = 2$.

The statement of the Fundamental Theorem of Linear Maps then implies dim $null\ \mathbf{R}_{\pi/2} = 0$.

Wait? Huh? What should it mean for a vector space to have dimension 0?

Thinking about rotation by $\frac{\pi}{2}$ radians counter clockwise, the only vector that results in $(0, 0)$ is the zero-vector itself. Therefore

$$null\ R_{\frac{\pi}{2}} = \{0\}$$

So that the statement of the Fundamental Theorem of Linear Maps makes sense, we would need dim $\{0\} = 0$. This doesn't seem to violate anything obvious (except perhaps our sensibilities) and so let us define the dimension of the trivial vector space to be 0.

Aside. *We could, quite reasonably, define the dimension of the trivial vector space,* $\{0\}$*, to not exist, rather than being* 0*. Doing so would not be incorrect, but would require us to modify the statement of the Fundamental Theorem of Linear Maps to account for the case where one of the relevant subspaces was trivial. This would complicate the statement of this theorem; we would need to account for the case where null* $T = \{0\}$ *And so as a community we decide to define the dimension of the trivial vector space to be* 0*. Doing so makes the*

statement of the Fundamental Theorem of Linear Maps sleek and efficient. (Analogous to this choice, consider the fact that $0! = 1$*. The only good reason to define* $0!$ *to be equal to* 1 *is convenience. Doing so makes it easier to consider binomial coefficients without having to deal with special cases.)*

It is easy forget that, in some sense, all of this math notational stuff is made up. The notation we use, the definitions we have, the way that we express theorems are all choices made by the mathematical community. Using different notation or making different choices for how we express our ideas is not objectively wrong. We define terms so that we have a common vocabulary to express complex ideas. We are free to personally violate mathematical convention without producing objectively incorrect work, but in doing so we make it difficult to communicate our ideas.

*For a lot of students a source of frustration (i.e., losing marks on assessment) is when the grader tells you that what you have written is wrong, even though your solution follows a correct line of reasoning. This often results for a student using mathematical notation or vocabulary in a non-standard way. When this occurs the grader cannot tell the difference between a misunderstanding in mathematical content and work that does not adhere to standard mathematical notation convention. In the absence of evidence of the latter, the grader will often assume the former. The burden of clear mathematical communication is on the writer, not the reader. (*I recognize the irony of my pointing this out, as, at times, I am sure these notes are less than clear – Chris *)*

In our two examples, both the null space and the range were finite dimensional, and so it made sense to talk about their dimension. How can we be sure that this is always the case?

For a linear map $T : U \to V$, the null space of T is a subspace of U. The hypothesis of the statement of the Fundamental Theorem of Linear Maps requires that U is finite dimensional. Are we certain that a subspace of a finite-dimensional vector space is also finite dimensional?

We faced this question when we considered the subspaces of $\mathbf{R}^3$. In our study of this question, we concluded the following (Theorem 3.4):

Theorem. *Every subspace of a finite dimensional vector space is finite dimensional.*

And so, since *null* T is a subspace of U and U is finite dimensional, then *null* T is finite dimensional.

We turn now to convincing ourselves that *range* T is necessarily finite dimensional. From our work in the previous module, we concluded that the range of T is a subspace of V. However, the hypothesis of the Fundamental Theorem of Linear Maps does not require V to be finite-dimensional. And so the statement of Theorem 3.4 is of no help.

Recalling the definition of finite-dimensional from Module 2 (Definition 2.10), to conclude that *range* T is finite dimensional (and thus the notation dim *range* T has meaning) it

suffices to show that *range* T is spanned by some set of vectors of V. In fact, we will do something more useful; we find a method to construct a basis for *range* T. Since a basis for a vector space necessarily spans the vector space, the existence of a basis for *range* T implies directly that *range* T is finite-dimensional (and thus, the notation $\dim$ *range* T has meaning!)

Let U be a finite dimensional vector space, let V be a vector space and let $T : U \to V$ be a linear map. Since U is finite dimensional, the null space of T (a subspace of U) is also finite dimensional. Since *null* T is finite dimensional, then it necessarily has a basis (Theorem 3.10). Let $B_{null} = (u_1, u_2, \ldots, u_k)$ be a basis for the null space of T.

Since B_{null} is a basis, this list is linearly independent. If B_{null} is not a basis for U then there exists vectors of U that are not in the span of B_{null}. Let $w_1 \in U$ be an element of U that is not in the span of B_{null}. Since w_1 is not in the span of B_{null}, then adding w_1 to B_{null} yields a linearly independent list.

If $(u_1, u_2, \ldots, u_k, w_1)$ is not a basis for U then there still exist elements of U that are not in span of $(u_1, u_2, \ldots, u_k, w_1)$. Consider continuing to add elements to this list so that the resulting list remains linearly independent. Since every linearly independent list of length $\dim U$ is a basis for U, eventually we will construct a basis for U:

$$B_U = (u_1, u_2, \ldots, u_k, w_1, w_2, \ldots, w_t)$$

Since this list is a basis for U and has $k + t$ elements, we have $\dim U = k + t$. We claim $B_{range} = (Tw_1, Tw_2, \ldots, Tw_t)$ is a basis for *range* T

To show B_{range} is a basis for *range* T, we show that B_{range} satisfies the two parts of the definition of basis. That is, we show the span of B_{range} is *range* T and B_{range} is linearly independent. We first show $span(B_{range}) =$ *range* T.

Consider $v \in$ *range* T. Since $v \in$ *range* T, there exists $u \in U$ so that $Tu = v$. Since $u \in U$ and B_U is a basis for U, u can be uniquely expressed as a linear combination of elements of B_U:

$$u = a_1u_1 + a_2u_2 + \ldots a_ku_k + a_{k+1}w_1 + \cdots + a_{k+t}w_t$$

Since $Tu = v$ and T is linear we have

$$\begin{aligned} v &= Tu \\ &= T\left(a_1u_1 + a_2u_2 + \ldots a_ku_k + a_{k+1}w_1 + \cdots + a_{k+t}w_t\right) \\ &= T(a_1u_1) + T(a_2u_2) + \ldots T(a_ku_k) + T(a_{k+1}w_1) + \cdots + T(a_{k+t}w_t) \\ &= a_1Tu_1 + a_2Tu_2 + \cdots + a_kTu_k + a_{k+1}Tw_1 + \cdots + a_{k+t}Tw_t \end{aligned}$$

Recall $B_{null} = (u_1, u_2, \ldots, u_k)$ is a basis for *null* T. Therefore $Tu_i = 0$ for all $i \in \{1, 2, \ldots, k\}$.

And so

$$\begin{aligned} v &= Tu \\ &= a_1Tu_1 + a_2Tu_2 + \dots a_kTu_k + a_{k+1}Tw_1 + \cdots + a_{k+t}Tw_t \\ &= a_1(0) + a_2(0) + \cdots + a_k(0) + a_{k+1}Tw_1 + \cdots + a_{k+t}Tw_t \\ &= a_{k+1}Tw_1 + \cdots + a_{k+t}Tw_t \end{aligned}$$

Therefore v can be expressed as a linear combination of elements of B_{range}. Therefore every element of $range\ T$ can be expressed as a linear combination of elements of B_{range}. And so $span(B_{range}) = range\ T$.

We now show that B_{range} is linearly independent. To so do, we show that the only solution to

$$0 = c_1Tw_1 + \cdots + c_tTw_t$$

is $c_1 = c_2 = \cdots = c_t = 0$.

Since T is linear, we have

$$\begin{aligned} 0 &= c_1Tw_1 + \cdots + c_tTw_t \\ &= T(c_1w_1 + c_2w_2 + \cdots + c_tw_t). \end{aligned}$$

Therefore $c_1w_1 + c_2w_2 + \cdots + c_tw_t \in null(T)$. Since B_{null} is a basis for $null(T)$, we can express $c_1w_1 + c_2w_2 + \cdots + c_tw_t$ as a linear combination of the elements of B_{null}. That is, there exists scalars $d_1, d_2, \dots, d_k$ so that

$$c_1w_1 + c_2w_2 + \dots c_tw_t = d_1u_1 + d_2u_2 + \dots d_ku_k$$

Rearranging we have

$$0 = d_1u_1 + d_2u_2 + \dots d_ku_k + (-c_1)w_1 + (-c_2)w_2 + \dots (-c_t)w_t$$

This is a linear combination of the elements of B_U that sum to 0. Since B_U is a basis for U, B_U is a linearly independent list. By definition, the only way for a linearly dependent list to sum to 0 is to have all of the scalars be 0. And so, $c_1 = c_2 = \cdots = c_t = 0$.

Therefore B_{range} is linearly independent. Since B_{range} is linearly independent and B_{range} spans $range\ T$, it follows that B_{range} is a basis for $range\ T$.

Since B_{range} is spanned by some list of vectors, it then follows that $range\ T$ is finite dimensional. And thus referring to the dimension of the range of T is the statement of the Fundamental Theorem of Linear Maps has meaning.

It turns out that hiding in our work above is a proof for the Fundamental Theorem of Linear Maps. Since B_{null} is a basis for the null space of T, the number of elements of B_{null} is exactly the dimension of the null space. In our work above we had

$$\dim\ null\ T = k$$

We added elements to this basis to produce a basis for U. We added exactly t additional elements and so B_U has $k + t$ elements. And so we had

$$\dim U = k + t$$

We used the t elements that we added, $w_1, w_2, \ldots, w_t$, to form a basis for *range* T. The list B_{range} is a basis for *range* T. This list has exactly t elements and so *range* T has dimension t. That is

$$\dim \ \textit{range}\, T = t$$

Putting all of this together, we have:

$$\dim \ U = \dim \textit{range}\, T + \dim \textit{null}\, T$$

One of the themes of MATH164 was solving systems of linear equations. One of the fundamental results from MATH164 is that a system with more variables than equations necessarily has at least one solution. This fact turns out to be a consequence of the Fundamental Theorem of Linear Maps. We return to this on Assignment 3.

Now that we are convinced about the Fundamental Theorem of Linear Maps, let us take a time to think about what it is telling us.

Our intuition for dimension tells us that $\mathbf{R}^2$ is *smaller* than $\mathbf{R}^3$. Sure $2 < 3$ but why does the relative dimension of these two vector spaces mean anything about their relative size? In fact, how can we ever make sense of what it means for one vector space to be smaller than another without knowing what size means for a vector space?

Aside 5.3. *If you have taken MATH163, this previous paragraph may have reminded you of the discussion of cardinality in that class. When size refers to cardinality, all finite-dimensional vector spaces have the size – they all have the same cardinality as* $\mathbf{R}$*. When we talk about the relative sizes of vector spaces we are not talking about the cardinality of the underlying set of elements.*

Let U be a finite dimensional vector space, let V be a vector space and let $T : U \to V$ be a linear map. We have

$$\dim \ U = \dim \textit{range}\, T + \dim \textit{null}\, T$$

That dimension is a non-negative integer tells us

$$\dim \textit{range}\, T \leq \dim \ U \qquad \text{and} \qquad \dim \textit{null}\, T \leq \dim \ U$$

This first inequality seems strange. The vector space *range* T is a subspace of V. There is no reason to think that the dimension of some subspace of V should be bounded by the dimension of some other vector space, U.

When this bound, $\dim range\ T \leq \dim\ U$, holds with equality, i.e., when $\dim range\ T = \dim\ U$, we must have $\dim null\ T = 0$. Similarly, when $\dim null\ T = 0$ this bound must hold with equality. And so, we have

$$\dim range\ T = \dim\ U \qquad \text{if and only if} \quad null\ T = \{0\}$$

In Module 4 we fully classified those linear maps for which $null\ T = \{0\}$; these are exactly the linear maps that are injective. And so

$$T : U \to V \text{ is injective if and only if } \dim range\ T = \dim\ U.$$

On the other hand, since $range\ T$ is a subspace of V, necessarily $\dim range\ T \leq \dim V$. And so if $\dim V < \dim U$, we must have $\dim range\ T \neq \dim\ V$. This then implies that T is not injective.

Theorem 5.4. *Let U and V be linear maps. If $\dim V < \dim U$, then no linear map from U to V is injective.*

Recalling the definition of surjective, saying $range\ T = V$ is identical to saying T is surjective. If $\dim U < \dim V$, then

$$\dim range\ T \leq \dim\ U < \dim V$$

In other words:

Theorem 5.5. *Let U and V be linear maps. If $\dim V > \dim U$, then no linear map from U to V is surjective.*

These theorems fall directly out of the statement of the Fundamental Theorem of Linear Maps. For now, we leave them as curiousities. We return to them in Module 6.

5.1 Test Your Understanding

1. Let U be a finite-dimensional vector space so that $\dim U$ is odd. Let $T : U \to U$ be a linear map. Explain how you know $\textit{null } T \neq \textit{range } T$.

2. Let $T : \mathbf{R}^3 \to \mathbf{R}^3$ be a linear map so that $T(x, y, z) = (x, y + z, y + z)$. One can compute $\textit{null } T = \{t(0, 1, -1) \mid t \in \mathbf{R}\}$. And so the vector $(0, 1, -1)$ is a basis for the null space. Find a basis for $\textit{range } T$ by first extending the basis for $\textit{null } T$ to be a basis for $\mathbf{R}^3$

5.1 Test Your Understanding Solution

1. By the Fundamental Theorem of Linear Maps we have

$$\dim U = \dim null; T + \dim range\ T$$

Since $\dim U$ is odd, we must have $\dim null; T \neq \dim range\ T$. Since $null\ T$ and $range\ T$ do not have the same dimension, they cannot be the same vector space. Therefore $null\ T \neq range\ T$.

2. Using the method that justified that the notation $\dim range\ T$ has meaning, we can construct a basis for $range\ T$ by first extending a basis for $null\ T$ to be a basis of $U = \mathbf{R}^3$ and then taking the image of the added vectors with respect to T. Given that $(0, 1, -1)$ is a basis for the null space, we want to find a linearly independent list of length 3 of vectors in $\mathbf{R}^3$ that includes $(0, 1, -1)$.

   By inspection, the list $(0, 1, -1), (1, 0, 0), (2, 2, -1)$ is linearly independent. Therefore it is a basis for $\mathbf{R}^3$. Therefore $T(1, 0, 0), T(2, 2, -1)$ is a basis for $range\ T$. Computing, we find

$$\begin{aligned} T(1, 0, 0) &= (1, 0, 0) \\ T(2, 2, -1) &= (2, 1, 1) \end{aligned}$$

   Therefore $(1, 0, 0), (2, 1, 1)$ is a basis for $range\ T$.

   (Other answers are possible here, depending on the choice of basis for $\mathbf{R}^3$ that includes $(0, 1, -1)$)

5.2 Matrices and Linear Maps (LADR 3C)

Our driving motivation in our study of linear maps so far has been drawing parallels between concepts for matrices and concepts for linear maps. We finally reach the point where we make this relationship explicit.

Let U and V be finite dimensional vector spaces. Since each of U and V are finite dimensional, they necessarily each have a basis. Let B_U and B_V be bases respectively for U and V. Let $T : U \to V$ be a linear map.

Since B_V is a basis for V, every element of V can be expressed uniquely as a linear combination of elements of B_V. (This is what the Criterion for Basis Theorem tells us)

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

By expressing an element of V as a linear combination of elements of B_V we generate a list of coefficients: $(a_1, a_2, a_3, \ldots, a_n)$

For example, let $V = \mathcal{P}_2(\mathbf{R})$, $B_V = (3, x^2 + 1, x + 3)$ and $v = x^2 + x + 1$. We have

$$x^2 + x + 1 = -1(3) + 1(x^2 + 1) + 1(x + 3)$$

Here the coefficients are $a_1 = -1$, $a_2 = 1$ and $a_3 = 1$.

We use this observation to generate a matrix for T that depends on the choice of B_U and B_V. Since $Tu_i \in V$ for every $u_i \in B_u$ we can express the image of any element of u (with respect to T) as a linear combination of elements of B_V. Subsequently, we can generate a list of coefficients that correspond to the image of each element of B_U.

Continuing with our example, let $U = \mathcal{P}_3(\mathbf{R})$ and let $B_U = (1, x, x^2, x^3)$. For $T = D$, where $D : \mathcal{P}_3(\mathbf{R}) \to \mathcal{P}_2(\mathbf{R})$ is given by $Df = \frac{d}{dx} f$ we can compute Dg for each $g \in B_U$.

$$\begin{aligned} D(1) &= 0 \\ D(x) &= 1 \\ D(x^2) &= 2x \\ D(x^3) &= 3x^2 \end{aligned}$$

Since the codomain of D is $\mathcal{P}_2(\mathbf{R})$ we can express each Dg as a linear combination of elements of B_V:

$$\begin{aligned} D(1) &= 0(3) + 0(x^2 + 1) + 0(x + 3) \\ D(x) &= \frac{1}{3}(1) + 0(x^2 + 1) + 0(x + 3) \\ D(x^2) &= -2(3) + 0(x^2 + 1) + 2(x + 3) \\ D(x^3) &= -1(3) + 3(x^2 + 1) + 0(x + 3) \end{aligned}$$

For element of B_U we get a corresponding list of coefficients. For $g = 1$ the coefficient list is $(0, 0, 0)$. For $g = x$, the coefficient list is $(1/3, 0, 0)$. For $g = x^2$ the coefficient list is $(-2, 0, 2)$. For $g = x^3$ the coefficient list is $(-1, 3, 0)$.

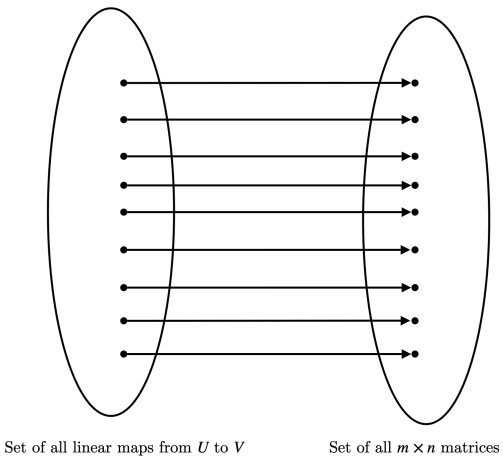
We form a matrix by taking these coefficient lists and making them the columns of a matrix.

$$\begin{bmatrix} 0 & 1/3 & -2 & -1 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

For vector spaces U and V we need to have in mind a particular basis for each of U and V and linear map $T : U \to V$. In our example above, if we change an element of either basis we will end up with a different matrix. And so there will be many possible matrices that correspond to a particular linear map.

Definition 5.6. *Let U and V be vector spaces and let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. Let $T : U \to V$ be a linear map. The* ***matrix of*** *T* ***with respect to*** *B_U* ***and*** *B_V is the $m \times n$ matrix, $\mathcal{M}(T)$, so that for all $i \in \{1, 2, \ldots, n\}$ the entries in the ith column of $\mathcal{M}(T)$ are the coefficients of the linear combination of elements of B_V that is equal to Tu_i.*

For vector spaces U and V and bases B_U and B_V this construction is a bijection. In other words we match up exactly the set of $m \times n$ matrices with the set of all linear maps from U to V.



At this point, our meaning of the word *same* becomes a little more clear. Given a linear map and a basis for each of its domain and codomain, we can construct a unique matrix corresponding to that linear map. Further, every matrix can be realized this way.

If we have any hope to treat linear maps and their corresponding matrices as the *same*, then we would expect our operations for linear maps (addition and scalar multiplication) to behave the same as the operations for the corresponding matrices. That is to say, for linear maps U and V and a pair of linear maps $S, T : U \to V$ can we expect the matrix for the linear map of $S + T$ (i.e., $\mathcal{M}(S + T)$) to be the same as the matrix we would get when we add $\mathcal{M}(S)$ and $\mathcal{M}(T)$?

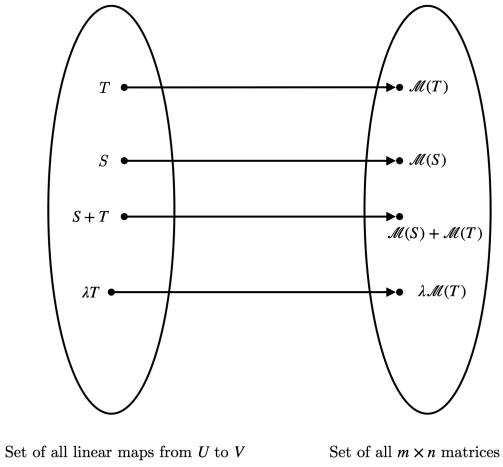
Surprisingly, the answer turns out to be yes!

Theorem 5.7. *Let U and V be vector spaces and let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. For every pair $S, T : U \to V$ of linear maps we have*

$$\mathcal{M}(S) + \mathcal{M}(T) = \mathcal{M}(S + T)$$

Theorem 5.8. *Let U and V be vector spaces and let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. For a linear map $T : U \to V$ and a scalar $\lambda \in \mathbf{F}$ we have*

$$\lambda\mathcal{M}(T) = \mathcal{M}(\lambda T)$$



We omit the proofs of these facts because introducing the required notation to prove them provides us with a barrier to our understanding that is not worth spending the time to overcome.

One way to conceptualize a vector space is by thinking about it as a collection of objects and some structure imparted by the two operations of the vector space. For example, the vector space $\mathbf{R}^2$ is more than just a set of vectors – it is the structure that we get by equipping the vectors with a meaning for addition and scalar multiplication.

These two results are telling us that the structure of the vector space of all linear maps between a pair of vector spaces is the same as the structure of the set of all matrices of an appropriate dimension. Whether we consider these objects with the notation of linear maps or with the notation of the corresponding matrices, we get the same structure. This is a tough idea to get our heads around, so we'll leave it for now and return to it in Module 6. If the word *isomorphism* is familiar to you, this is the concept that this paragraph is discussing.

Just as we can associate each linear map with a matrix, we can associate an element of a n-dimensional vector space with an element of $\mathbf{R}^n$. For example, let $U = \mathcal{P}_2(\mathbf{R})$ and let $B_U = (3, x^2 + 1, x + 3)$ be a basis for U. Since B_U is a basis for U, every element of U can be uniquely expressed as a linear combination of elements of B_U. For instance:

$$x^2 + x + 1 = -1(3) + 1(x^2 + 1) + 1(x + 3)$$

For every $f \in \mathcal{P}_2(\mathbf{R})$ we will get a unique triple of coefficients a_1, a_2 and a_3. In the example above we have $a_1 = -1$, $a_2 = 1$ and $a_3 = 1$. Thus, with respect to this basis, we can match up every element of $\mathcal{P}_2(\mathbf{R})$ with an element of $\mathbf{R}^3$. For $f(x) = x^2 + x + 1$ the corresponding element of $\mathbf{R}^3$ is $(-1, 1, 1)$.

Notice, however, that this correspondence depends on the choice of basis for U. If we chose $B'_U = (2, x^2 + 1, x + 3)$ we would have

$$x^2 + x + 1 = -\frac{3}{2}(2) + 1(x^2 + 1) + 1(x + 3)$$

In this case the corresponding element of $\mathbf{R}^3$ would be $\left(-\frac{3}{2}, 1, 1\right)$.

In general we define the matrix of an element of a finite dimensional vector space as follows:

Definition 5.9. *Let U be a vector space, let $B_U = (u_1, u_2, \ldots, u_n)$ be a basis for U and let $u \in U$. The **matrix of** u **with respect to** B_U, denoted $\mathcal{M}(u)$, is the element of $\mathbf{R}^n$ so that for all $i \in \{1, 2, \ldots, n\}$ the ith entry of $\mathcal{M}(u)$ is the coefficient of u_i when u is expressed as a linear combination of the elements of B_U.*

Continuing with the example above, the matrix of $x^2 + x + 1$ with respect to the basis $(2, x^2 + 1, x + 3)$ is

$$\begin{bmatrix} -\frac{3}{2} \\ 1 \\ 1 \end{bmatrix}$$

As noticed above, the choice of basis directly effects the resulting matrix of a vector. However, this is not true for the zero vector. No matter which basis we choose for U, the coefficients will always be 0. Since a basis is linearly independent, the only way for a linear combination of basis vectors to sum to 0 is to have all the coefficients be 0.

Lemma 5.10. *Let U be a vector space and let B_U be a basis for U. We have $\mathcal{M}(u) = 0$ if and only if $u = 0$.*

Aside. *Why have we called this a* lemma *rather than a* theorem*? We use the word* lemma *for mathematical truths that are interesting to us in the context of helping us prove a larger mathematical truth (i.e., a theorem). When something is called a* lemma *you can expect it will show up somewhere in the proof of a theorem. This lemma is used implicitly as part of a result in Section 5.2. On an upcoming assessment you will be asked to spot where it has been used.*

When we introduced linear maps, we dealt with some notional awkwardness. Rather than writing $T(u)$ we write Tu. Expressed as Tu, evaluating T at u *looks* the same as multiplying a matrix by a vector. Now that we can express both a linear map and a element of a vector space as a vector in $\mathbf{R}^n$, this notational choice becomes a little more clear. We can find the matrix of the vector Tu by multiplying the matrix of T with the matrix of u.

Let us return to our familiar example differentiation of polynomials of degree 3. Let $U = \mathcal{P}_3(\mathbf{R})$ and $V = \mathcal{P}_2(\mathbf{R})$ and consider $D : \mathcal{P}_3(\mathbf{R}) \to \mathcal{P}_2(\mathbf{R})$ given by $Df = \frac{d}{dx}$. Let $B_U = (1, x, x^2, x^3)$ and $B_V = (1, x, x^2)$. To find the matrix of D with respect to these two bases, we first compute Du for each $u \in B_U$.

$$\begin{aligned} D1 &= 0 \\ Dx &= 1 \\ Dx^2 &= 2x \\ Dx^3 &= 3x^2 \end{aligned}$$

Given that we have chosen the basis $B_V = (1, x, x^2)$, there is little work to do find the corresponding lists of coefficients.

$$\begin{aligned} D1 &= 0(1) + 0(x) + 0(x^2) \\ Dx &= 1(1) + 0(x) + 0(x^2) \\ Dx^2 &= 0(1) + 2(x) + 0(x^2) \\ Dx^3 &= 0(1) + 0(x) + 3(x^2) \end{aligned}$$

By definition, these coefficients give the us the columns of the matrix $\mathcal{M}(D)$

$$\mathcal{M}(D) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Consider $f = 4x^3 + x^2$. To find the matrix for f we must first express f as a linear combination of elements of B_U.

$$4x^3 + x^2 = 0(1) + 0(x) + 1(x^2) + 4(x^3)$$

Therefore

$$\mathcal{M}(f) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 4 \end{bmatrix}$$

Consider separately the product $\mathcal{M}(D)\mathcal{M}(f)$ and evaluating D at f, Df.

We have $Df = 12x^2 + 2x$ and

$$\mathcal{M}(D)\mathcal{M}(f) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 12 \end{bmatrix}$$

Expressing Df as matrix with respect to the basis B_V we have

$$\mathcal{M}(Df) = \begin{bmatrix} 0 \\ 2 \\ 12 \end{bmatrix}$$

Therefore $\mathcal{M}(D)\mathcal{M}(f) = \mathcal{M}(Df)$.

Such an equality turns out to be true in general.

Theorem 5.11. *Let U and V be vector spaces let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. For all $u \in U$ we have*

$$\mathcal{M}(Tu) = \mathcal{M}(T)\mathcal{M}(u)$$

When we introduced null space and range of a linear map in the previous module, we did so by analogy to the corresponding concepts for matrices. We realized that by massaging the definition of null space and column space of a matrix, we could make sense of these ideas of linear maps. In turns out that the correspondence is much more direct that we might have expected. If we take the vectors corresponding to the elements of the null space of a linear map, we get exactly the vectors of the null space of the corresponding matrix.

Continuing our example above, we have

$$\textit{null } D = \{f(x) = t \mid t \in \mathbf{R}\}$$

Using techniques from MATH164, we find

$$\textit{null } \mathcal{M}(D) = \left\{ \begin{bmatrix} t \\ 0 \\ 0 \\ 0 \end{bmatrix} \mid t \in \mathbf{R} \right\}$$

Let us consider now the matrices of elements of *null* D. The null space of D consists of functions of the form $f(x) = t$. With respect to the basis $(1, x, x^2, x^3)$, the matrix of an element $f \in$ *null* D is

$$\mathcal{M}(f) = \begin{bmatrix} t \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

These are exactly the elements that make up *null* $\mathcal{M}(D)$.

We prove this fact in general.

Theorem 5.12. *Let U and V be vector spaces let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. We have $u \in$ null T if and only if $\mathcal{M}(u) \in$ null $\mathcal{M}(T)$.*

Proof. Let U and V be vector spaces and let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V.

Let $u \in$ *null* T. Therefore $Tu = 0$. Therefore $\mathcal{M}(Tu) = \mathcal{M}(0) = 0$. Since $\mathcal{M}(Tu) = \mathcal{M}(T)\mathcal{M}(u)$ we have $\mathcal{M}(T)\mathcal{M}(u) = 0$. Therefore $\mathcal{M}(u) \in$ *null* $\mathcal{M}(T)$.

Consider now $\mathcal{M}(u) \in$ *null* $\mathcal{M}(T)$. Since $\mathcal{M}(u) \in$ *null*$\mathcal{M}(T)$ we have $\mathcal{M}(T)\mathcal{M}(u) = 0$. And so $\mathcal{M}(Tu) = 0$. Therefore $Tu = 0$. And so $u \in$ *null* T. □

The statement of this theorem is very technical and quite abstract, but it is telling us something useful. To know the null space of a linear map it is enough to find the null space of the corresponding matrix.

Unsurprisingly, the corresponding statement for the range of a linear map and the column space of its corresponding matrix also turns out to be true.

Theorem 5.13. *Let U and V be vector spaces let $B_U = (u_1, u_2, \ldots, u_n)$ and $B_V = (v_1, v_2, \ldots, v_m)$ respectively be bases for U and V. We have $v \in$ range T if and only if $\mathcal{M}(v) \in$ col $\mathcal{M}(T)$.*

The main theme of this module has been noting that every linear map can be represented as a matrix and every matrix corresponds to a linear map. As we looked the implications of this fact, we looked at a method to construct a vector in $\mathbf{R}^n$ from any element of a vector space of dimension n. This allowed us to make explicit the connection between the null space and column space of a linear map and null space and column space of the corresponding matrix. Just as matrices are linear maps in disguise, vectors (as we understood them from MATH164) are in fact representations of elements of a vector space of dimension n. We explore this further in the coming module.

Aside. *Have you ever wondered why this subject is called Linear Algebra?*

*For mathematicians, an **algebra** is any system in which there are some collection of mathematical objects that can be added and multiplied. And so the term Linear Algebra refers to the study of the vector space of linear maps. In this module we have learned that matrices are linear maps in disguise. And so in Linear Algebra when we study the algebra of linear maps we use their representation as matrices and instead study matrices.*

5.2 Test Your Understanding

1. Let $U = \mathbf{R}^2$ and $V = \mathbf{R}^2$. Let $R_{\pi/2} : U \to V$ be the linear map that rotates vectors by $\pi/2$ counter-clockwise. Find $\mathcal{M}(R_{\pi/2})$ with respect to the standard bases $B_U = B_V = (e_1, e_2)$.
2. Let $U = \mathbf{R}^2$ and $V = \mathbf{R}^2$. Let $P : U \to V$ be the projection of vectors on to ℓ, the line spanned by $(1, 1)$. Find $\mathcal{M}(P)$ with respect to the bases $B_U = (1, 0), (-3, -1)$ and $B_V = (1, 1), (2, 1)$.
3. Let $U = \mathcal{M}_{3\times 3}(\mathbf{R})$ and $V = \mathcal{P}_4(\mathbf{R})$ and let B_U and B_V respectively be bases for U and V. Let $T : U \to V$ be a linear map. What are the dimensions of the matrix $\mathcal{M}(T)$? Explain.

5.2 Test Your Understanding Solution

1. Columns of $\mathcal{M}(R_{\pi/2})$ correspond to the images with respect to $R_{\pi/2}$ of elements of B_U. By observation, we have

$$R_{\pi/2}(e_1) = (0,1)$$
$$R_{\pi/2}(e_2) = (-1,0)$$

We now express these images as a linear combination of elements of B_V:

$$(0,1) = 0e_1 + 1e_2$$
$$(-1,0) = -1e_1 + 0e_2$$

Therefore

$$\mathcal{M}(R_{\pi/2}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

2. Columns of $\mathcal{M}(P)$ correspond to the images with respect to P of elements of B_U. We compute:

$$P(1,0) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

By linearity of P we have

$$P(-3,-1) = P(-3e_1) + P(-1e_2) = \left(-\frac{3}{2}, -\frac{3}{2}\right) + \left(-\frac{1}{2}, -\frac{1}{2}\right) = (-2,-2)$$

We now express these images as a linear combination of elements of B_V:

$$\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}(1,1) + 0(2,1)$$
$$(-2,-2) = -2(1,1) + 0(2,1)$$

Therefore

$$\mathcal{M}(P) = \begin{bmatrix} \frac{1}{2} & -2 \\ 0 & 0 \end{bmatrix}$$

3. From our work in Test Your Understanding on Module 4. we have $\dim U = 9$ and $\dim V = 5$. Therefore B_U and B_V respectively are of length 9 and length 5. Columns of $\mathcal{M}(T)$ correspond to the images with respect to T of elements of B_U. Therefore $\mathcal{M}(T)$ has 9 columns. The entries in a column are the coefficients of the image of the corresponding element of B_U when expressed as a linear combination of elements in B_V. Since B_V has 5 entries, such linear combinations will have 5 coefficients. Therefore $\mathcal{M}(T)$ has 5 rows.

---

6 Linear Maps Part III (LADR 3D)

Learning Incomes.

- *be comfortable with the meanings of the terms composition,injective, surjective, bijective and inverse and how they apply to functions.*
- *understand the relationship between an injective linear map and the null space of that linear map*
- *understand the relationship between a surjective linear map and the range of that linear map.*
- *remember how to use the null space or range of a matrix to determine if it is invertible.*
- *recall how to use the Criterion for Basis theorem to show that a list is a basis*
- *be able to construct the matrix of a linear map with respect to a pair of bases*
- *be able to construct the matrix of an element of a vector space with respect to a basis*

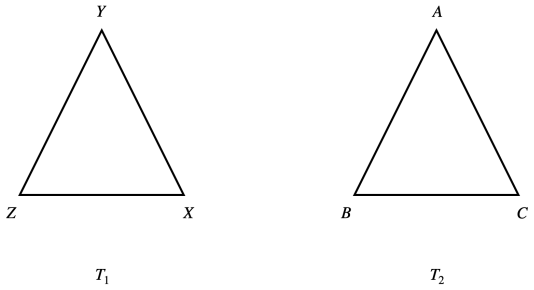
Learning Outcomes.

- *Understand what same means when we refer to matrices and linear maps being the same.*
- *Given bases for a pair of vector spaces, be able to determine if the vector spaces are isomorphic*
- *Understand the relationship between an invertible linear operator and an invertible matrix*
- *Given a linear map, be able to determine if it is invertible by computing its matrix with respect to some pair of bases*
- *Given a linear map, be able to determine if there is a pair of bases so that the matrix with respect to those bases is the identity matrix*

Newly Defined Terms and Notation.

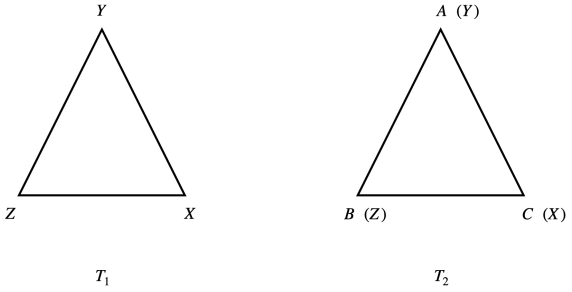
- *isomorphism, isomorphic,* $U \cong V$*, identity linear map, linear operator,* $\mathcal{L}(U, V)$

We start with a seemingly benign question – are these two triangles the same?



One supposes the answer to this question depends on what we mean when we use the word *same*. If the names of the points of the triangle help to define what a triangle is, then these are not the same triangle; the triangles do not have the same set of points.

However, if we only care about the lengths of the sides of the triangle, then we can relabel the points of T_2 so that we arrive at the triangle T_1 by assigning each of the labels in T_2 a corresponding label in T_1.



In thinking about how we can relabel the points of T_2 so that we arrive at the triangle T_1, we've assigned each of the labels in T_2 a corresponding label in T_1 and vice versa. This correspondence is a function $f : \{A, B, C\} \to \{X, Y, Z\}$ so that

$$f(A) = Y$$
$$f(B) = Z$$
$$f(C) = X$$

But, not every function from $\{A, B, C\}$ to $\{X, Y, Z\}$ turns T_2 in to T_1. For example the function: $f' : \{A, B, C\} \to \{X, Y, Z\}$ so that

$$f'(A) = Y$$
$$f'(B) = Y$$
$$f'(C) = X$$

doesn't give a relabelling of the points of T_2 with the labels of the points of T_1 as one of the labels of T_1 is repeated. In other words, f' is not a bijection.

For another example, the function:

$$f''(A) = X$$
$$f''(B) = Y$$
$$f''(C) = Z$$

also doesn't give us a valid correspondence. The length of the line between B and C (i.e. $\overline{BC}$ and the length of the line between Y and Z (i.e., $\overline{YZ}$) are not the same. In other words $\overline{BC} \neq \overline{f''(B)f''(C)}$.

And so, in order to know that two triangles are the same, we must be able to find a way to relabel the points of the first triangle with the labels of the points of the second triangle so that corresponding side lengths are preserved. The world of geometry we would express this idea with the word *congruence*. And we would write:

$$\Delta ABC \cong \Delta XYZ$$

In general, for a pair of triangles ΔDEF and ΔUVW we have $\Delta DEF \cong \Delta UVW$ when there is a bijection $f : \{D, E, F\} \to \{U, V, W\}$ so that for all $S, T \in \{D, E, F\}$ we have $\overline{ST} = \overline{f(S)f(T)}$. In the example above, we have $\Delta ABC \cong \Delta XYZ$ as $f : \{A, B, C\} \to \{X, Y, Z\}$ so that

$$f(A) = Y$$
$$f(B) = Z$$
$$f(C) = X$$

is a bijection and

$$\overline{XY} = \overline{\phi(X)\phi(Y)}$$
$$\overline{YZ} = \overline{\phi(Y)\phi(Z)}$$
$$\overline{XZ} = \overline{\phi(X)\phi(Z)}$$

Since this function is a bijection, the inverse of this function, namely f^{-1}, necessarily exists. We have $f^{-1} : \{X, Y, Z\} \to \{A, B, C\}$ so that

$$f^{-1}(Y) = A$$
$$f^{-1}(Z) = B$$
$$f^{-1}(X) = C$$

The use of the word *inverse* here and its use in the term *additive inverse* is no mere coincidence. Recall that an additive inverse in a vector space is the thing we have to add to an element to get our additive identity, 0. Perhaps, we can define the inverse of a function using some notion of identity for a function.

Definition 6.1. *Let A be a set. The* ***identity function*** *on A is the function $I : A \to A$ so that $I(a) = a$ for every $a \in A$*

Let A be a set. With a little bit of work, one can verify that for all functions $g : A \to A$ we have $g \circ I = g$. That is, composing a function with the identity function doesn't change the function. With respect to the operation of function composition, the function I plays the same role as 0 does with respect to addition. And so, continuing with the comparison to additive inverse and identity, the inverse of a function g^{-1} with respect to composition behaves the same as an additive inverse does with respect to addition. That is

$$g \circ g^{-1} = I$$

In other words, if $g : A \to A$ is a bijection, then for all $a \in A$ we have

$$g \circ g^{-1}(a) = a$$

Let us turn now to think about how these ideas might apply to vector spaces. Thinking back to congruent triangles, our observation is that change in how things are labelled does not change the structure of the object. When we talked about triangles being the same the structure we were interested in was the lengths of the sides. Thinking now about vector spaces the structure we are interested in is addition and scalar multiplication.

Let $U = \mathcal{P}_2(\mathbf{R})$ and $V = \mathbf{R}^3$. The set U is a set of polynomials and the set V is a set of vectors. Polynomials and vectors are not the same thing, and so $U \neq V$. However, thinking about U and V as vector spaces it turns out that these are two different ways to label the same vector space.

Just as relabelling the points of a triangle doesn't change the underlying structure of the triangle, let us consider relabelling elements of $\mathcal{P}_2(\mathbf{R})$ so that $a_2x^2+a_1x+a_0$ gets relabelled as the vector (a_2, a_1, a_0). With respect to this relabelling, we wonder if the underlying structure of the vector space is preserved.

The structure of our vector space is provided by our notions of addition and scalar multiplication. If we add a pair of elements of $\mathcal{P}_2(\mathbf{R})$ we have

$$(a_2x^2 + a_1x + a_0) + (b_2x^2 + b_1x + b_0) = (a_2 + b_2)x^2 + (a_1 + b_1)x + (a_0 + b_0)$$

If we relabel these elements of $\mathcal{P}_2(\mathbf{R})$ as elements of $\mathbf{R}^3$ we have the sum:

$$(a_2, a_1, a_0) + (b_2, b_1, b_0) = (a_2 + b_2, a_1 + b_1, a_0 + b_0)$$

Adding a pair of elements in $\mathcal{P}_2(\mathbf{R})$ and then relabelling to be an element of $\mathbf{R}^3$ gives the same result as if we first relabel our elements of $\mathcal{P}_2(\mathbf{R})$ as elements of $\mathbf{R}^3$ and then perform our addition.

The same is true when we consider scalar multiplication. Multiplying an element of $\mathcal{P}_2(\mathbf{R})$ by a scalar and then relabelling to be an element of $\mathbf{R}^3$ gives the same results as if we first relabel our element of $\mathcal{P}_2(\mathbf{R})$ and then multiply by a scalar.

$$\lambda(a_2x^2 + a_1x + a_0) = \lambda a_2x^2 + \lambda a_1x + \lambda a_0$$

$$\lambda(a_2, a_1, a_0) = (\lambda a_2, \lambda a_1, \lambda a_0)$$

Let $\phi : U \to V$ be the function that relabels elements of $\mathcal{P}_2(\mathbf{R})$ as elements of $\mathbf{R}^3$. For example, we have $\phi(2x^2 + x + 4) = (2, 1, 4)$. This function is a bijection; it exactly matches up elements of $\mathcal{P}_2(\mathbf{R})$ with elements of $\mathbf{R}^3$.

Since this function is a bijection, we have that ϕ^{-1} exists and satisfies the property:

$$\phi \circ \phi^{-1}(v) = v$$

for each $v \in \mathbf{R}^3$.

Aside. *The symbol ϕ is the lower-case Greek letter "phi". It is sometimes pronounced as* fee*, but it is often pronounced as* f-eye*.*

Try not to let the inclusion letters from alphabets other than our familiar Latin alphabet intimidate you. The use of various pieces of notation to denote various objects is a choice we make to be consistent with the choices that others make. For example, we seem to always use the letters f, g and h to denote polynomials. And we seem to always use x as the name of a variable. So far in this course, we have used capital letters to denote linear transformations. Computer scientists seem to always use i, j and k as their local incrementing variable in the loop data structures. (I always struggle to choose what variable to use next when I need to nest more than three loops...) And, in some mathematical communities, we consistently use ϕ to denote an isomorphism.

These choices are cultural. They are not inherently right or wrong; but by choosing notation consistent with what our readers expect, it eases their path.

Recall the sentence above:

Adding a pair of elements in $\mathcal{P}_2(\mathbf{R})$ and then relabelling to be an element of $\mathbf{R}^3$ gives the same result as if we first relabel our elements of $\mathcal{P}_2(\mathbf{R})$ as elements of $\mathbf{R}^3$ and then perform our addition.

Expressing this into a slick and slightly confusing piece of mathematical jargon yields

$$\phi(f+g) = \phi(f) + \phi(g)$$

Reading from the left, the first addition sign denotes addition in $\mathcal{P}_2(\mathbf{R})$. The second addition sign denotes addition in $\mathbf{R}^3$

Similarly, recall the statement

Multiplying an element of $\mathcal{P}_2(\mathbf{R})$ by a scalar and then relabelling to be an element of $\mathbf{R}^3$ gives the same result as if we first relabel our element of $\mathcal{P}_2(\mathbf{R})$ and then multiply by a scalar.

Expressing this into a slick and slightly confusing piece of mathematical jargon yields

$$\phi(\lambda f) = \lambda\phi(f)$$

Before we dive into the full blown mathematics of this in the coming section, let us take a moment to first generalize a little bit. Let U and V be vector spaces. If there is a bijective function $\phi : U \to V$ so that for all $u_1, u_2 \in U$ and all $\lambda \in \mathbf{F}$ we have

1. $\phi(u+v) = \phi(u) + \phi(v)$ and
2. $\phi(\lambda f) = \lambda\phi(f)$,

then U and V are the same vector space, but perhaps, with different labels on the elements. The function ϕ converts the label for every element of U into a corresponding label for an element of V while preserving the structure of the vector space. This means that for every true statement we can make about U, we can make an analogous true statement for V. For example, if such a function exists, then we necessarily have $\dim U = \dim V$; we can convert any basis for U into a basis for V.

6.1 Vector Space Isomorphism (LADR 3D)

If you have taken other upper-level mathematics courses you might have encountered the word *isomorphism*. This word (which is likely new for most of us) captures the sense of sameness we discussed in the introduction to this module; two mathematical things being isomorphic means that the objects have the same internal structure but perhaps different labels.

The word *isomorphic* finds its roots in ancient greek. *Iso* comes from *isos*, which means equal and *morphic* comes from *morph* meaning shape or structure.

Aside. *An isosceles is a triangle where exactly two sides are equal. Latin won the naming game for triangles – the equi in equilateral triangle comes from Latin.*

Okay, enough etymology, back to linear algebra.

Definition 6.2. *Let U and V be vector spaces. We say U* ***is isomorphic to*** *V when there exists a bijection $\phi : U \to V$ so that for all $u_1, u_2 \in U$ and all $\lambda \in \mathbf{F}$ we have*

1. *$\phi(u_1 + u_2) = \phi(u_1) + \phi(u_2)$; and*
2. *$\phi(\lambda u_1) = \lambda\phi(u_1)$*

In this context we say ϕ ***is an isomorphism from*** *U* ***to*** *V. When U is isomorphic to V we write $U \cong V$.*

From our example, in the previous section, we have that $\mathcal{P}_2(\mathbf{R})$ and $\mathbf{R}^3$ are isomorphic. We can express this as: $\mathcal{P}_2(\mathbf{R}) \cong \mathbf{R}^3$. The function $\phi : \mathcal{P}_2(\mathbf{R}) \to \mathbf{R}^3$ given by

$$\phi(a_2x^2 + a_1x + a_0) = (a_2, a_1, a_0)$$

is an isomorphism.

Aside. *If terms in mathematics were chosen to make things as easy as possible for students, then we would probably use the word isomorphic rather than congruent was talking about triangles that are the same.*

Notice that the definition of isomorphic speaks of *existence* of an isomorphism. In almost all cases isomorphisms are not unique. Therefore lots of ways we can match up elements of $\mathcal{P}_2(\mathbf{R})$ and $\mathbf{R}^3$ that preserves structure. For example, we could relabel $a_2x^2 + a_1x + a_0$ as $(-a_0, -a_1, -a_2)$ and use this relabelling to conclude that $\mathcal{P}_2(\mathbf{R})$ and $\mathbf{R}^3$ have the same structure.

Aside. *For those of you who have taken MATH163, isomorphism is an excellent example of an equivalence relation. If we declare two finite-dimensional vector spaces to be related when they are isomorphic, the resulting equivalence classes partition the collection of all finite dimensional vector spaces into classes that have the same dimension. In what follows, we will see that the equivalence classes of real vector spaces of dimension n always contains a copy of $\mathbf{R}^n$. When we talk about being able to represent a finite dimensional real vector space as $\mathbf{R}^n$, we are choosing the vector space of matrices to be the representative element for an equivalence class.*

For those of you who have taken MATH327, two graphs are isomorphic when there exists a bijection from one vertex set to the other that preserves the existence of edges.

For those of you who knows the meaning of the word group, two groups are isomorphic when there is a bijection from one group to the other than preserves the outcome of the group operation.

Recall for a moment the definition of linear map from Module 4:

Definition. *Let V and W be vector spaces and let $T : V \to W$ be a function, We say T* ***is linear*** *when for all $u, v \in V$ and all $\lambda \in \mathbf{F}$ we have*

1. *$T(u) + T(v) = T(u + v)$; and*
2. *$T(\lambda u) = \lambda T(u)$*

When T is a linear function we say T ***is a linear map.*** *When T is a linear function, we may write Tv to refer to $T(v)$.*

Our definition for isomorphism looks a lot like our definition for linear map! The only difference between an isomorphism and a linear map is that an isomorphism required to be a bijection. And so, isomorphisms are exactly the same as bijective linear maps.

Theorem 6.3. *Let U and V be vector spaces. We have $U \cong V$ if and only if there exists a bijective linear map $T : U \to V$*

In Modules 4 and 5 we spent most of our time thinking about linear maps. Here we apply some of that knowledge to see what we can learn about isomorphism for vector spaces. In particular, we wonder which pairs of vector spaces are isomorphic.

Recall the Fundamental Theorem of Linear Maps:

Theorem (The Fundamental Theorem of Linear Maps)**.** *Let U be a finite-dimensional vector*

space and let V be a vector space. For every linear map $T : U \to V$ we have

$$\dim\ U = \dim range\ T + \dim null\ T$$

At the end of the previous module we looked at two consequences of this theorem:

Theorem. *Let U and V be linear maps. If $\dim V < \dim U$, then no linear map from U to V is injective.*

Theorem. *Let U and V be linear maps. If $\dim V > \dim U$, then no linear map from U to V is surjective.*

Since an isomorphism is a bijective linear maps, and a bijective function is, by definition, both injective and surjective, these theorems give us a pair of conditions for the non-existence of an isomorphism.

If $\dim V < \dim U$, then no linear map form U to V is injective. Therefore no such linear map is bijective, Therefore no such linear map is an isomorphism.

$\dim V > \dim U$, then no linear map from U to V is surjective. Therefore no such linear map is bijective, Therefore no such linear map is an isomorphism.

And so, if U and V are finite-dimensional vector spaces and $U \cong V$, then necessarily $\dim U = \dim V$. Perhaps unexpectedly, the converse of the previous statement is also true: if $\dim U = \dim V$, then $U \cong V$.

Theorem 6.4. *Let U and V be finite dimensional real vector spaces. We have $U \cong V$ if and only if $\dim U = \dim V$.*

Proof. Let U and V be finite dimensional vector spaces. Let B_U and B_V respectively be bases for U and V.

Assume $\dim U = \dim V$ and let $n = \dim U = \dim V$. Let $B_U = (u_1, u_2, \ldots, u_n)$ and let $B_V = (v_1, v_2, \ldots, v_n)$ be bases for U and V respectively. By Theorem 4.4 there is a unique linear map $T : U \to V$ so that $Tu_i = v_i$ for all $i \in \{1, 2, 3, \ldots n\}$. To show that $U \cong V$ is suffices to show T is an isomorphism. Since T is linear, to show T is an isomorphism it suffices to show that it is a bijection.

Recall from the proof of Theorem 4.4 we have

$$T(a_1u_1 + a_2u_2 + \cdots + a_nu_n) = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

To show T is a bijection we show that it is an injection and a surjection. To show T an injection we show T satisfies the definition of injection. That is, we show for all $u, u' \in U$ we have $Tu \neq Tu'$ whenever $u \neq u'$.

By the Criterion for Basis Theorem, there exists distinct lists of scalars $c_1, c_2, \ldots, c_n$ and $c_1', c_2', \ldots c_n'$ so that

$$u = c_1u_1 + c_2u_2 + \cdots + c_nu_n$$
$$u' = c_1'u_1 + c_2'u_2 + \cdots + c_n'u_n$$

If $u \neq u'$, then there exists at least one $i \in \{1, 2, 3, \ldots, n\}$ so that $c_i \neq c_i'$.

By construction we have

$$Tu = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$
$$Tu' = c_1'v_1 + c_2'v_2 + \cdots + c_n'v_n$$

Since $v_1, v_2, \ldots, v_n$ is a basis for V and $c_i \neq c_i'$, it follows by the Criterion for Basis Theorem that $Tu \neq Tu'$. Therefore T is injective.

To show T is a surjection we show T satisfies the definition of surjection. That is, we show for all $v \in V$ that there exists $u \in U$ so that $Tu = v$. Since B_V is a basis for V, the Criterion for Basis Theorem tells us that there is exactly one list of scalars $(d_1, d_2, \ldots, d_n)$ so that

$$v = d_1v_1 + d_2v_2 + \cdots + d_nv_n$$

Let $u = d_1u_1 + d_2u_2 + \cdots + d_nu_n$. By construction we have $Tu = v$. Therefore T is surjective.

Since T is both injective and surjective, it follows by definition that T is bijective. Since T is a bijective linear map, then by definition T is an isomorphism. Therefore $U \cong V$.

Assume now $U \cong V$. Let $\phi : U \to V$ be an isomorphism. Let ϕ is an isomorphism it is injective. By Theorem 4.9, it follows $null\ \phi = \{0\}$. Since $null\ \phi = \{0\}$, by the Fundamental Theorem of Linear Maps we have

$$\dim U = \dim range\ T$$

Since ϕ is isomorphism, is surjective. Therefore $range\ \phi = V$. Therefore $\dim range\ T = \dim V$. Combining this with the equality above yields

$$\dim U = \dim V$$

□

Since this theorem is an *if and only if* statement, it tells us that for finite-dimensional vector spaces the statements

$$U \text{ and } V \text{ are isomorphic}$$

and

are synonyms. This means, that know if a pair of finite-dimensional vector spaces are isomorphic, it suffices to compare their respective dimensions.

For example, we have $\mathcal{P}_3(\mathbf{R}) \cong \mathcal{M}_{2\times 2}(\mathbf{R})$; both of these vector spaces have dimension 2. On the other hand, $\mathcal{P}_3(\mathbf{R}) \ncong \mathbf{R}^3$; these vector spaces do not have the same dimension.

Let us take a moment to think about the case $V = \mathbf{R}^n$ in the statement of Theorem 6.4. As $\dim \mathbf{R}^n = n$, Theorem 6.4 is telling us that every real vector space of dimension n is isomorphic to $\mathbf{R}^n$. Informally, this means that every vector space of dimension n is $\mathbf{R}^n$ in disguise.

Aside. *Wow!*

The definition of isomorphism tells that for every n-dimension vector space U we should be able to find a bijective linear map $T : U \to \mathbf{R}^n$. Recall from Module 5 the following definition:

Definition. *Let U be a real vector space, let $B_U = (u_1, u_2, \ldots, u_n)$ be a basis for U and let $u \in U$. The **matrix of u with respect to** B_U, denoted $\mathcal{M}(u)$, is the element of $\mathbf{R}^n$ so that for all $i \in \{1, 2, \ldots, n\}$ the ith entry of $\mathcal{M}(u)$ is the coefficient of u_i when u is expressed as a linear combination of the elements of B_U.*

Let U be a n-dimensional real vector space and let B_U be a basis for U. Once U and B_U are in place, the construction of $\mathcal{M}(u)$ is a function in disguise! The name of the function is $\mathcal{M}$. The domain of the function is U and the codomain in $\mathbf{R}^n$. (That is, $\mathcal{M} : U \to \mathbf{R}^n$) The notation $\mathcal{M}(u)$ denotes the image of u with respect to this function; the matrix of u with respect to B_U.

Theorem 6.5. *Let U be a real n-dimensional vector space and let B_U be a basis for U. The function $\mathcal{M} : U \to \mathbf{R}^n$ is an isomorphism from U to $\mathbf{R}^n$.*

We omit this proof for now. Understanding the proof of this theorem requires of us a firm understanding of many concepts from the first half of this course. And so it will be very useful come midterm time.

Just as $\mathcal{M} : U \to \mathbf{R}^n$ defines an isomorphism, so too does the construction of a matrix for linear map, $\mathcal{M}(T)$. For vector spaces U and V, let $\mathcal{L}(U, V)$ denote the set of all linear maps from U to V.

Theorem 6.6. *Let U and V be real vector spaces respectively of dimensions n and m. Let B_U and B_V respectively be bases for U and V. With respect to the bases B_U and B_V, the function $\mathcal{M} : \mathcal{L}(U, V) \to \mathcal{M}_{m\times n}(\mathbf{R})$ is an isomorphism from $\mathcal{L}(U, V)$ to $\mathcal{M}_{m\times n}(\mathbf{R})$.*

Corollary 6.7. *Let U and V be finite dimensional real vector spaces respectively of dimensions n and m. We have $\mathcal{L}(U, V) \cong \mathcal{M}_{m\times n}(\mathbf{R})$.*

Aside. *As we have seen, sometimes mathematical facts get titles other than* theorem. *A* corollary *is a mathematical fact whose truth follows almost direct from a theorem. The proof of Corollary 6.7 follows by from the statement of Theorem 6.6 by recalling the definition of isomorphism: the existence of an isomorphism between vector spaces is synonymous with saying the two vector spaces are isomorphic.*

We began our study of linear maps with an eye towards the following statement:

Linear maps and matrices are the same[6] thing.

Isomorphic is the answer to the question – what does *same* mean in this context? For finite-dimensional vector spaces U and V respectively of dimension n and m, the structure of the vector space $\mathcal{L}(U, V)$ is the same as the structure of the vector space $\mathcal{M}_{m\times n}(\mathbf{F})$. With respect to a basis for each of $\mathcal{L}(U, V)$ and $\mathcal{M}_{m\times n}(\mathbf{F})$, the function $\mathcal{M} : \mathcal{L}(U, V) \to \mathcal{M}_{m\times n}(\mathbf{F})$ tells us how to relabel linear maps as matrices. Since $\mathcal{M}_{m\times n}(\mathbf{F})$ has dimension $m \times n$, Theorem 6.4 tells us the dimension of $\mathcal{L}(U, V)$.

Corollary 6.8. *Let U and V be finite dimensional real vector spaces. We have*

$$\dim \mathcal{L}(U, V) = (\dim U) \cdot (\dim V).$$

The results above are written specifically for real vector spaces. For complex vector spaces, we have analogous results. Finite-dimensional complex vector spaces are isomorphic exactly when they have the same dimension. A complex vector space of dimension n is isomorphic to the complex vector space $\mathbf{C}^n$. If U and V are finite-dimensional complex vector spaces, then $\mathcal{L}(U, V) \cong \mathcal{M}_{m\times n}(\mathbf{C})$. We will return to complex numbers and complex vector spaces in Module 7.

Aside. *Depending on your mathematical background, reasons for being in this course and your mathematical goals, your reaction to the statements of Theorem 6.4 and Corollary 6.7 may vary considerably from your peers in this course. However, if you are ever wondering what sort of things interest and surprises mathematicians, it is statements like these. Prima*

[6] once we define very carefully what we mean when we say *same*

facie, there is no reason at all to expect that all finite-dimensional vector spaces are $\mathbf{R}^n$ *in disguise, or that every linear map is a matrix in disguise. However, our careful study of these objects revealed to us a hidden truth about objects that on the surface seem to have nothing in common but in fact are the same.*

6.1 Test Your Understanding

1. Are $\mathcal{M}_{4\times 3}(\mathbf{R})$ and $\mathbf{R}^7$ isomorphic? Justify your response.
2. Let V be a finite-dimensional real vector space. What is the dimension of V if $\mathcal{L}(V, V) \cong \mathbf{R}^9$?

6.1 Test Your Understanding Solution

1. No. We have $\dim \mathcal{M}_{4\times 3} = 12$ and $\dim \mathbf{R}^7 = 7$. Therefore $\mathcal{M}_{3\times 3} \ncong \mathbf{R}^7$.
2. If $\mathcal{L}(V, V) \cong \mathbf{R}^9$, then $\dim \mathcal{L}(V, V) = 9$. Therefore $\dim V = 3$.

6.2 Invertibility and Linear Operators (LADR 3D)

Throughout our study of systems of linear equations and matrices in MATH164, we have paid special attention to square matrices; the matrices that correspond to systems of linear equations where the number of variables and the number of equations was the same. Such systems gave us the full gamut of possibilities for the number of solutions to such a system. The corresponding square matrices gave us a path to study matrix inverses and powers of matrices.

In thinking about how we can transplant such ideas to our study of linear maps, we lean heavily on the representation of linear maps as matrices. The dimensions of a matrix corresponding to a linear map depend on the dimension of the domain and codomain of the linear map. And so with an eye towards transplanting ideas in the study of square matrices to linear maps,we consider the study of linear maps for which the domain and codomain are the same vector space. The matrix of such a linear map is necessarily a square matrix. In this section, we focus specifically on invertibility of such linear maps.

When we learned about invertible matrices in MATH164 we saw a number of different conditions that told us that a square matrix was invertible. There are lots of ways to express these ideas. Let us write one down so that we have one to refer to.

Theorem. *Let A be an $n \times n$ matrix. The following three statements are equivalent:*

i. *A is invertible.*

ii. *$\text{null } A = \{0\}$.*

iii. *$\text{col } A = \mathbf{R}^n$.*

In Module 4, we have already examined the meaning for statements *ii.* and *iii.* in the context of linear maps. From our discussion above in the introductory to this module, we recognize invertiblity of a function to mean the same as the being bijective. And so perhaps such a theorem is true when transplanted to the context of linear maps. To prove a theorem for linear maps analogous to the one above, we recall a pair of facts from Modules 4 and 5.

Recall from Module 4, the following theorem:

Theorem. *Let $T : U \to V$, we have $\text{null } T = \{0\}$ if and only if T is injective.*

In Module 5 during our discussion of the consequences of the Fundamental Theorem of Linear Maps we stated the following about a linear map $T : U \to V$.

Recalling the definition of surjective, saying $\text{range } T = V$ is identical to saying T is surjective.

And so for $T : V \to V$ we replace *null* $A = 0$ with *injective* and *col* $A = \mathbf{R}^n$ with *surjective* and arrive at the following result.

Theorem 6.9. *Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear map. The following three statements are equivalent.*

i. T is invertible.

ii. T is injective.

iii. T is surjective.

To prove that these three statements are equivalent, we will prove that the truth of one statement implies the truth of the successive statement. That is, we show:

1. If T is invertible, then T injective;
2. if T is injective, then T is surjective; and
3. if T is surjective, then T is invertible.

By showing this, we in fact show the following:

- If any of statements $i., ii.$ or $iii.$ is true, then all three are true; and
- if any of statements $i., ii.$ or $iii.$ is false, then all three are false.

These latter two facts are what we mean when we say the three statements *are equivalent.*

Proof. Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear map.

If T is invertible, then T is a bijection. Therefore T is injective. And so $i.$ implies $ii.$

If T is injective, then by Theorem 4.9 we have *null* $T = \{0\}$. By the Fundamental Theorem of Linear Maps is then follows that $\dim V = \dim \mathit{range}\ T$. Therefore *range* $T = V$ and so T is surjective. And so $ii.$ implies $iii.$

Assume now T is surjective. Therefore *range* $T = V$. Therefore $\dim V = \dim \mathit{range}; T$. And so $\dim \mathit{null}\ T = 0$. Therefore *null* $T = \{0\}$. By Theorem 4.9 it follows that T is injective, Since T is both injective and surjective, by definition T is a bijection. In other words, T is invertible. And so $iii.$ implies $i.$

Since $i.$ implies $ii.$, $ii.$ implies $iii.$ and $iii.$ implies $i.$, all three statements are equivalent. □

One may wonder, why we consider $T : V \to V$ in the statement of this previous theorem, rather than $T : U \to V$ where U and V have the same dimension? Theorem 6.4 tells us that if U and V have the same dimension, then we can relabel the elements of U to so as to turn U in to a copy of V. And when U and V are linear maps of the same dimension, we can simplify our work by assuming $U = V$.

This standing assumption leads us to a new piece of terminology.

Definition 6.10. *Let V be a vector space and let $T : V \to V$ be a linear map. We call T a* ***linear operator****.*

The word operator is meant to remind us of the word *operation*. The output of an operation on a set is another element of the set. And so when the codomain of a linear map is the same as the domain, we refer to the linear map as a linear operator. For a vector space V, let $\mathcal{L}(V, V)$ denote the set of linear operators.

Returning now to the statement of Theorem 6.9 – In MATH164 we related invertibility of a matrix to the null space and column space of a matrix. In Module 5, we related the null space and column space of a matrix to the null space and range of some corresponding linear map. And so perhaps we can relate invertibility of a linear map and invertibility of a corresponding matrix. To do this, we take a moment to look more carefully at the definition of invertible matrix.

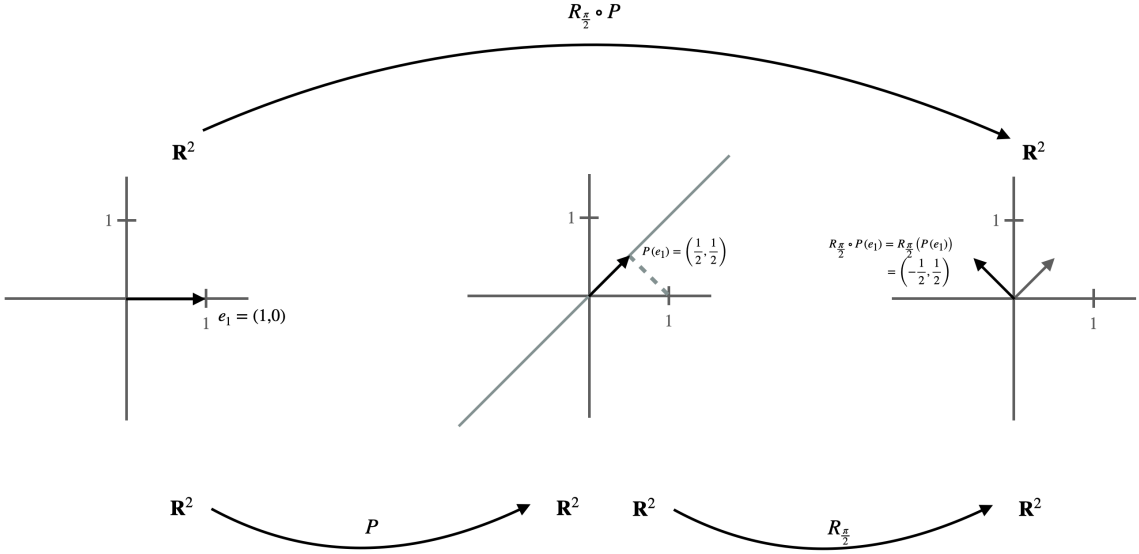
Definition 6.11. *Let A be an $n \times n$ matrix. We say A* ***is invertible*** *when there exists a matrix B so that $AB = I_n$. When $AB = I_n$ we say B* ***is the inverse of*** *A and we denote B by A^{-1}*

To be able to talk about matrix invertibility, we first need to be able to talk about matrix multiplication. To be able to talk about invertibility of functions, we first need to be able to talk about function composition. Let us take a moment to better understand the meaning of composition of functions in the context of linear maps.

We return to two previous examples of linear maps from $\mathbf{R}^2 \to \mathbf{R}^2$; projection and rotation. Let $P : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map that projects vectors on to the line ℓ spanned by $(1, 1)$. And let $R_{\frac{\pi}{2}} : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map that rotates vectors counterclockwise by $\frac{\pi}{2}$ radians.

The composition of these functions $R_{\frac{\pi}{2}} \circ P$ is the linear map that first projects a vector on to the line ℓ and then rotates the result by $\frac{\pi}{2}$ radians. For example

$$R_{\frac{\pi}{2}} \circ P(e_1) = R_{\frac{\pi}{2}}\left(P(e_1)\right) = R_{\frac{\pi}{2}}\left(\frac{1}{2}, \frac{1}{2}\right) = \left(-\frac{1}{2}, \frac{1}{2}\right)$$



Just as our two operations in the vector space of linear maps (addition and scalar multiplication) are emulated by matrix operations for the corresponding matrices, so too can we relate composition of linear maps to a familiar operation for matrices – matrix products.

As $R_{\frac{\pi}{2}} \circ P : \mathbf{R}^2 \to \mathbf{R}^2$ is a linear map, it can be represented by a matrix with respect to some bases. With respect to the standard basis (e_1, e_2) for both the domain and the codomain, the matrix of this linear map is:

$$\mathcal{M}(R_{\frac{\pi}{2}} \circ P) = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

The linear map $R_{\frac{\pi}{2}} \circ P$ is formed from the composition of 2 linear maps: $R_{\frac{\pi}{2}}$ and P. With respect to the standard basis vectors, we have

$$\mathcal{M}(R_{\frac{\pi}{2}}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathcal{M}(P) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Taking the product of these two matrices, we find

$$M(R_{\frac{\pi}{2}})\mathcal{M}(P) = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \mathcal{M}(P \circ R_{\frac{\pi}{2}})$$

The matrix of a composition of linear maps is equal to the product of the matrices of each of the linear maps in the composition. More carefully, we have

Theorem 6.12. *Let V be a vector space. Let $S, T : V \to V$ be linear maps. Let B_{V_1}, B_{V_2} and B_{V_3} respectively be bases for V We have*

$$\mathcal{M}(S)\mathcal{M}(T) = \mathcal{M}(S \circ T)$$

where

- *$\mathcal{M}(S)$ is the matrix of S with respect to B_{V_1} and B_{V_2};*
- *$\mathcal{M}(T)$ is the matrix of S with respect to B_{V_2} and B_{V_3}; and*
- *$\mathcal{M}(S \circ T)$ is the matrix of $S \circ T$ with respect to B_{V_1} and B_{V_3}.*

In our example above we have $V = \mathbf{R}^2$, $T = P$, $S = R_{\frac{\pi}{2}}$ and $B_{V_1} = B_{V_2} = B_{V_3} = (e_1, e_2)$.

We omit the proof of this theorem because introducing the required notation to prove it provides us with a barrier to our understanding that is not worth spending the time to overcome.

Aside. *In MATH164 you learned about matrix multiplication for matrices that were not square. To multiply a pair of matrices, their respective dimensions need to be consistent, in some sense. Multiplication of non-square matrices corresponds to composition of linear maps whose domain and codomain need not be the same dimension. For such linear maps a statement analogous to Theorem 6.12 is true, but it is a pain to write down.*

With Theorem 6.12 we can now state a result that perhaps shouldn't be too surprising given the general theme of our study of linear maps and matrices.

Theorem 6.13. *Let V be a finite dimensional vector space and let $T : V \to V$ be a linear map. Let B_V and B'_V be bases for V. The matrix of T with respect to B_V and B'_V is invertible if and only if T is invertible.*

This proof requires a few ingredients which will take us some time to fully digest. We'll explore the proof of Theorem 6.13 on Assignment 4. We will spend the remainder of thinking about some of the pieces that we will need for the proof.

Let V be a vector space and let B_V be a basis for V. Let $I : V \to V$ be the unique linear map so that $Iv_i = v_i$ for all $i \in \{1, 2, \ldots, n\}$. The Criterion for Basis Theorem tells us that for each $v \in V$, there exists unique scalars $a_1, a_2, \ldots, a_n$ so that

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

Therefore

$$\begin{aligned} Iv &= I(a_1v_1 + a_2v_2 + \cdots + a_nv_n) \\ &= a_1v_1 + a_2v_2 + \cdots + a_nv_n \\ &= v \end{aligned}$$

Unsurprisingly, the image of each element of V with respect to I is the element itself.

Constructing the matrix for this linear map with respect to these bases B_V and B_V, we first compute the image of each element of B_V with respect to I.

$$\begin{aligned} Iv_1 &= v_1 \\ Iv_2 &= v_2 \\ &\vdots \\ Iv_n &= v_n \end{aligned}$$

We then express each image as a linear combination of the basis of the codomain:

$$\begin{aligned} Iv_1 &= 1(v_1) + 0(v_2) + 0(v_3) + \cdots + 0(v_{n-1}) + 0(v_n) \\ Iv_1 &= 0(v_1) + 1(v_2) + 0(v_3) + \cdots + 0(v_{n-1}) + 0(v_n) \\ &\vdots \\ Iv_n &= 0(v_1) + 0(v_2) + 0(v_3) + \cdots + 0(v_{n-1}) + 1(v_n) \end{aligned}$$

Recalling the definition of the matrix of a linear map, the entries in the ith column are the coefficients when Iv_i is expressed as linear combination of the elements of the basis of the codomain. Therefore

$$\mathcal{M}(I) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = I_n$$

Just as we call the matrix I_n, the **identity matrix** let us give the same name to this linear map, I.

Definition 6.14. *Let V be a vector space. The **identity linear map** $I : V \to V$ is the unique linear map so that $Iv = v$ for all $v \in V$.*

However, depending on our choice of bases when we construct $\mathcal{M}(I)$ we may not have $\mathcal{M}(I) = I_n$. For example, for $I : \mathbf{R}^2 \to \mathbf{R}^2$, the matrix of I with respect to the bases (e_1, e_2)

and $(-e_1, e_2)$ is

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Conversely it is possible to have the matrix of a linear map be the identity matrix without the linear map being the identity linear map. Let V be a vector space and let $B_V = (v_1, v_2, \ldots, v_n)$ and $B'_V = (v'_1, v'_2, \ldots, v'_n)$ be bases for V. By Theorem 4.4 there is a unique linear map $T : V \to V$ so that $Tv_i = v'_i$ for all $i \in \{1, 2, \ldots, n\}$.

Constructing the matrix for this linear map with respect to B_V and B'_V, we first compute the image of each element of B_V with respect to T.

$$\begin{aligned} Tv_1 &= v'_1 \\ Tv_2 &= v'_2 \\ &\vdots \\ Tv_n &= v'_n \end{aligned}$$

We then express each image as a linear combination of the basis of the codomain:

$$\begin{aligned} Iv_1 &= 1(v'_1) + 0(v'_2) + 0(v'_3) + \cdots + 0(v'_{n-1}) + 0(v'_n) \\ Iv_1 &= 0(v'_1) + 1(v'_2) + 0(v'_3) + \cdots + 0(v'_{n-1}) + 0(v'_n) \\ &\vdots \\ Iv_n &= 0(v'_1) + 0(v'_2) + 0(v'_3) + \cdots + 0(v'_{n-1}) + 1(v'_n) \end{aligned}$$

Recalling the definition of the matrix of a linear map, the entries in the ith column are the coefficents when v_i is expressed as linear combination of the elements of the basis of the codomain. Therefore

$$\mathcal{M}(T) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} = I_n$$

And so we wonder, for which linear operators $T : V \to V$ can we find a pair of bases B_V and B'_V so that $\mathcal{M}(T) = I_n$, where $\mathcal{M}(T)$ is the matrix of T with respect to B_V and B'_V. As we have seen above, when T is unique linear map that maps elements of B_V to elements of B'_V, the matrix of T with respect to B_V and B'_V is I_n.

For an example, let us return back to $\mathbf{R}_{\frac{\pi}{2}} : \mathbf{R}^2 \to \mathbf{R}^2$. This linear operator maps the basis (e_1, e_2) to the basis $(e_2, -e_1)$. Let us construct the matrix of $\mathbf{R}_{\frac{\pi}{2}}$ with respect to these two bases:

Constructing the matrix for this linear map with respect to (e_1, e_2) and $(e_2, -e_1)$, we first compute the image of each element of (e_1, e_2) with respect to $\mathbf{R}_{\frac{\pi}{2}}$.

$$\mathbf{R}_{\frac{\pi}{2}} e_1 = (0, 1)$$
$$\mathbf{R}_{\frac{\pi}{2}} e_2 = (-1, 0)$$

We then express each image as a linear combination of the basis $(e_2, -e_1)$

$$(0, 1) = 1(e_2) + 0(-e_1)$$
$$(-1, 0) = 0(e_2) + 1(-e_1)$$

Therefore, with respect to the bases (e_1, e_2) and $(e_2, -e_1)$ we have

$$\mathcal{M}(\mathbf{R}_{\frac{\pi}{2}}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

Returning now to the general case, let us consider a finite-dimensional vector space V, a linear operator T, and a pair of bases B_V and B'_V so that the matrix of T with respect to B_V and B'_V is the identity matrix. Let $B_V = \{v_1, v_2, \dots v_n\}$ and $B'_V = \{v'_1, v'_2, \dots v'_n\}$. That the first column of $\mathcal{M}(T)$ is the column vector whose first entry is 1 and all other entries are 0 means that when we express the image of v_1 with respect to T as a linear combination of elements of B'_V we have

$$Tv_1 = 1(v'_1) + 0(v'_2) + \cdots + 0(v'_n)$$

Therefore $Tv_1 = v'_1$.

A similar argument tells us that $Tv_i = v'_i$ for all $i \in \{1, 2 \dots, n\}$. Therefore $T : V \to V$ is the unique linear map so that $Tv_i = v'_i$ for all $i \in \{1, 2 \dots, n\}$.

Putting all of this together, we have:

Lemma 6.15. *Let V be a vector space of dimension n and let $B_V = (v_1, v_2, \dots, v_n)$ and $B'_V = (v'_1, v'_2, \dots, v'_n)$ be bases for V. Let T be a linear operator. The matrix of T with respect to B_V and B'_V is the identity matrix if and only if T is the unique linear map so that $Tv_i = v'_i$ for all $i \in \{1, 2, \dots, n\}$.*

With this lemma and the statement of Theorem 6.9, we can use invertibility to fully classify those linear operators for which there is a pair of bases so that the matrix of the linear operator with respect to those bases is the identity matrix.

Let V be a vector space, let B_V and B'_V be bases of V. Let $B_V = (v_1, v_2, \dots v_n)$ and $B'_V = (v'_1, v'_2, \dots v'_n)$. Let T be the linear operator so that $Tv_i = v'_i$ for all $i \in \{1, 2, \dots, n\}$.

With an eye towards applying the statement of Theorem 6.9, let us think about the null space of this linear operator. That is, for which $v \in V$ do we have $Tv = 0$?

By the linearity of T for $v = a_1v_1 + a_2v_2 + \ldots a_nv_n$ we have

$$\begin{aligned} Tv &= T(a_1v_1 + a_2v_2 + \cdots + a_nv_n) \\ &= a_1Tv_1 + a_2Tv_2 + \cdots + a_nTv_n \\ &= a_1v_1' + a_2v_2' + \cdots + a_nv_n' \end{aligned}$$

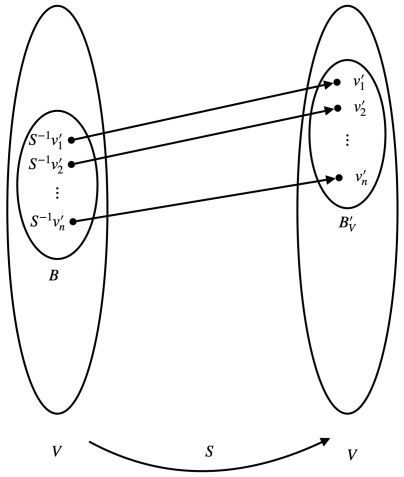
And so to have $Tv = 0$ we must have

$$0 = a_1v_1' + a_2v_2' + \cdots + a_nv_n'$$

Since B_V' is a basis for V, the list B_V' is linearly independent. Therefore we must have $a_1 = a_2 = \cdots = a_n = 0$. And so $Tv = 0$ only for $v = 0v_1 + 0v_2 + \cdots + 0v_n$. In other words, $null\ T = \{0\}$.

By Theorem 4.9, T is injective. And by Theorem 6.9, T is invertible. And so, by Lemma 6.15, if there exists bases so that the matrix with respect T is the identity matrix, then T is invertible.

Conversely, let S be an invertible linear operator for V. Since S is invertible, for each $v' \in B_V'$, the notation $S^{-1}v'$ has meaning. For $v' \in B_V'$, the notation $S^{-1}v'$ refers to the element of the domain, V whose image is v'. Let $B = (S^{-1}v_1', S^{-1}v_2', \ldots, S^{-1}v_n')$



If B is a basis for V, then S is the unique linear map that maps elements of the basis B to elements of the basis B_V'. And so from Lemma 6.15, we could conclude that the matrix of S with respect to B and B_V' is the identity matrix. Consequently we then could conclude that if S is invertible, then there exists a pair of bases for V so that the matrix with respect to those bases is the identity matrix.

To show B is a basis we apply the Criterion for Basis Theorem. Let $v \in V$ and let

$$v = a_1 S^{-1} v_1' + a_2 S^{-1} v_1' + \cdots + a_n S^{-1} v_1'$$

and

$$v = b_1 S^{-1} v_1' + b_2 S^{-1} v_1' + \cdots + b_n S^{-1} v_1'$$

To show B is a basis it suffices to show $a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n$.

We compute Sv. Since S is linear we have:

$$\begin{aligned} Sv &= S\left(a_1 S^{-1} v_1' + a_2 S^{-1} v_1' + \cdots + a_n S^{-1} v_1'\right) \\ &= a_1 S\left(S^{-1} v_1'\right) + a_2 S\left(S^{-1} v_2'\right) + \ldots a_n S\left(S^{-1} v_n'\right) \end{aligned}$$

and

$$Sv = b_1 S\left(S^{-1} v_1'\right) + b_2 S\left(S^{-1} v_2'\right) + \ldots b_n S\left(S^{-1} v_n'\right)$$

Since S^{-1} is the inverse of S we have $S(S^{-1} v_i') = v_i'$ for all $i \in \{1, 2, \ldots, n\}$. Therefore

$$Sv = a_1 v_1' + a_1 v_2' + \cdots + a_1 v_n'$$

and

$$Sv = b_1 v_1' + b_1 v_2' + \cdots + b_1 v_n'$$

Since $(v_1', v_2', \ldots v_n')$ is a basis for V, the Criterion for Basis Theorem implies $a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n$. Therefore B is a basis for V.

From our work above, we have the following statement:

Theorem 6.16. *Let V be a vector space and let T be a linear operator. There exists bases B_V and B_V' so that the matrix of T with respect to B_V and B_V' is the identity matrix if and only if T is invertible.*

Contextualizing, this theorem tells us that by choosing our bases carefully, we can represent the linear operator $\mathbf{R}_{\frac{\pi}{2}}$ as the matrix I_2. On the other hand, no matter how we choose our bases, we cannot represent the projection linear operator P as the identity matrix.

6.2 Test Your Understanding

1. Let V be a finite dimensional vector space and let T be a linear operator so that $null\ T = \{0\}$. Explain how you know T is invertible.
2. In our work above, we showed $\mathcal{M}(P)\mathcal{M}(R_{\frac{\pi}{2}}) = \mathcal{M}(P \circ R_{\frac{\pi}{2}})$.
   (a) Compute $P \circ R_{\frac{\pi}{2}}(e_1)$. Notice $P \circ R_{\frac{\pi}{2}}(e_1) \neq R_{\frac{\pi}{2}} \circ P(e_1)$
   (b) Without actually doing the computation, explain how you know
   $$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \neq \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$
3. Explain how you know $\mathcal{M}(P) \neq I_2$ no matter which bases are used to construct $\mathcal{M}(P)$.

6.2 Test Your Understanding Solution

1. By Theorem 4.9, if *null* $T = \{0\}$, then T is injective. By Theorem 6.9, since T is injective, T is invertible. Therefore if T is a linear operator so that *null* $T = \{0\}$, then T is invertible.

2. (a) We have $P \circ R_{\frac{\pi}{2}}(e_1) = P(R_{\frac{\pi}{2}}(e_1))$. Rotating e_1 counterclockwise by $\pi/2$ radians yields e_2. Therefore

$$P \circ R_{\frac{\pi}{2}}(e_1) = P(R_{\frac{\pi}{2}}(e_1)) = P(e_2)$$

   Projecting e_2 on to ℓ yields $\left(\frac{1}{2}, \frac{1}{2}\right)$. Therefore $P \circ R_{\frac{\pi}{2}}(e_1) = \left(\frac{1}{2}, \frac{1}{2}\right)$.

   From our work above, we have $R_{\frac{\pi}{2}} \circ P(e_1) = \left(-\frac{1}{2}, \frac{1}{2}\right)$.

   (b) Since $R_{\frac{\pi}{2}} \circ P(e_1) \neq P \circ R_{\frac{\pi}{2}}(e_1)$ with respect to the standard basis vectors we have

$$\mathcal{M}(P \circ R_{\frac{\pi}{2}}) \neq \mathcal{M}(R_{\frac{\pi}{2}} \circ P)$$

   And so by Theorem 6.12 we have $\mathcal{M}(P)\mathcal{M}(R_{\frac{\pi}{2}}) \neq \mathcal{M}(R_{\frac{\pi}{2}})\mathcal{M}(P)$. And so

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \neq \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

3. We have *null* $P \neq \{0\}$. By Theorem 4.9, P is not injective. By Theorem 6.9, P is not invertible. And so by Theorem 6.16, we have $\mathcal{M}(P) \neq I_2$ no matter which bases are used to construct $\mathcal{M}(P)$.

---

7 Polynomials with Complex Coefficients (LADR 4)

Learning Incomes.

- *Remember the meaning of the terms eigenvalue and eigenvector of a matrix.*
- *Remember the relationship between determinant and invertibility of a matrix.*

Learning Outcomes.

- *Understand the relationship between linear factors and zeros of a polynomial.*
- *Understand the statement of the Fundamental Theorem of Algebra and why is it not true for polynomials in $\mathcal{P}(\mathbf{R})$.*
- *Understand how we know that every complex polynomial of degree n factors into exactly n linear factors.*

Newly Defined Terms and Notation.

- *polynomial factor*

In thinking about how our work in MATH266 mirrors our work in MATH164, one would expect that the topic of eigenvalues and eigenvectors would show up eventually. Recall that an **eigenvector** of a matrix A is a vector $v \neq 0$ so that $Av = \lambda v$ for some scalar λ. Informally, we interpret this to mean that multiplying A by v has the effect of *stretching* v by a factor of λ. In this context, we say that λ **is an eigenvalue of** A **and** v **is a corresponding eigenvector**

For a quick refresher of the meaning of eigenvalues and eigenvectors for a matrix, this video does a much better job than I could ever hope to do `https://www.youtube.com/watch?v=PFDu9oVAE-g`

Continuing our work in extending concepts for matrices to be concepts for linear maps, we arrive at the following definitions.

Definition. *Let V be a vector space, let T be a linear operator on V and let $\lambda \in \mathbf{F}$. We say λ* ***is an eigenvalue of*** *T when there exists $v \in V$ with $v \neq 0$ so that $Tv = \lambda v$.*

Definition. *Let V be a vector space, let T be a linear operator, let $\lambda \in \mathbf{F}$ be an eigenvalue of T and let $v \in V$ so that $v \neq 0$. We say v* ***is an eigenvector of*** *T* ***corresponding to*** *λ when $Tv = \lambda v$.*

For example, let $\mathcal{D}(\mathbf{R})$ be the set of all differentiable functions from $\mathbf{R}$ to $\mathbf{R}$ and let $D : \mathcal{D}(\mathbf{R}) \to \mathcal{D}(\mathbf{R})$ be the linear map given by $Df = \frac{d}{dx}f(x)$ for all $f \in \mathcal{D}(\mathbf{R})$. Since $D(e^{2x}) =$

$2e^{2x}$, we have that $f(x) = e^{2x}$ is an eigenvector of D that corresponds to the eigenvalue $\lambda = 2$.

We undertake our study of eigenvalues and eigenvectors for linear operators in Module 8. For now, we will take a quick look back at eigenvalues and eigenvectors for 2×2 matrices.

Let $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. To find eigenvalues and eigenvectors for A, we want to find pairs (λ, v) so that $Av = \lambda v$. Recall that for any vector v, we can emulate multiplying by a scalar with multiplying by a matrix as follows:

$$\lambda v = \begin{bmatrix} \lambda v_1 \\ \lambda v_2 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda I_2 v$$

To find eigenvalues and eigenvectors for A, we want to find pairs (λ, v) so that $Av = \lambda I_2 v$. Since $\mathbf{R}^2$ is a vector space, its elements have additive inverses. Therefore the vector $-(\lambda I_2 v)$ exists and has the property that $\lambda I_2 v + -(\lambda I_2 v) = 0$. Adding this additive inverse to both sides of the equation $Av = \lambda I_2 v$ yields

$$Av + -(\lambda I_2 v) = \lambda I_2 v + -(\lambda I_2 v)$$

Simplifying yields

$$(A - \lambda I_2)v = 0$$

And so for v to be an eigenvector corresponding to the eigenvalue λ, it must be that v is an element of the null space of the matrix $A - \lambda I_2$. Recalling $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, we have

$$A - \lambda I_2 = \begin{bmatrix} \frac{1}{2} - \lambda & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} - \lambda \end{bmatrix}$$

And so we wonder, for which values of λ does this matrix have a non-trivial null space? For this matrix to have a non-trivial null space, it must be non-invertible. In other words, it must have determinant equal to 0. We have

$$\begin{aligned} det(A - \lambda I_2) &= \left(\frac{1}{2} - \lambda\right) \cdot \left(\frac{1}{2} - \lambda\right) - \frac{1}{2} \cdot \frac{1}{2} \\ &= \lambda^2 - \lambda \\ &= \lambda(\lambda - 1) \end{aligned}$$

And so, to have $det(A - \lambda I_2) = 0$ we must have $\lambda = 0$ or $\lambda = 1$. For each possible eigenvalue, 0 and 1, we seek an eigenvector.

When $\lambda = 0$ we have $A - \lambda I_2 = A$. And so any non-zero vector in the null space of A is an eigenvector corresponding to λ.

Consider now the case $\lambda = 1$. To find a corresponding eigenvector, we find $v \in \mathbf{R}^2$ so that $Av = 1v$. Rearranging, as before, we seek an element of the null space of the matrix $A - 1I_2$. We have

$$A - 1I_2 = \begin{bmatrix} \frac{1}{2} - 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} - 1 \end{bmatrix} = A - 1I_2 = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Row reducing the matrix $\left[\begin{array}{cc|c} -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{array}\right]$ yields $\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$ Therefore we have

$$null\ (A - 1I_2) = \{t(1,1) | t \in \mathbf{R}\}.$$

Choosing any $v \in null\ (A - 1I_2)$ yields an eigenvector corresponding to $\lambda = 1$. Let us choose $v = (1, 1)$. And so $A(1, 1) = 1(1, 1)$. We wonder, what does this even mean?

Let $P : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map that projects vectors on to the line spanned by $(1, 1)$.

(*In previous modules, I stuffed up the calculation of the matrix $\mathcal{M}(P)$ with respect to the standard basis. All occurrences of $\sqrt{2}$ should have been 2 – Chris*)

Recalling that A is the matrix of P with respect to the standard basis, it should be no surprise that A has an eigenvalue of $\lambda = 1$ with corresponding eigenvector $v = (1, 1)$. This linear map does not stretch or compress vectors and maps the vector $(1, 1)$ to itself.

Let us turn now to our other standing example: $R_{\frac{\pi}{2}} : \mathbf{R}^2 \to \mathbf{R}^2$. This linear map fixes no vector in $\mathbf{R}^2$ (other than the zero vector) and so perhaps we can expect a corresponding matrix to have no eigenvectors. The matrix of P with respect to the standard basis vectors is $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Proceeding as above, to find an eigenvalue for this matrix we seek values of λ so that the determinant of $\begin{bmatrix} 0 - \lambda & -1 \\ 1 & -\lambda \end{bmatrix}$ is 0. This matrix has determinant $\lambda^2 + 1$. And we see there are no real numbers λ so that $\lambda^2 + 1 = 0$.

But wait! In Module 0 we introduced complex numbers precisely for this very situation. Though the polynomial $\lambda^2 + 1 = 0$ has no solutions in the real numbers, it does have solutions in the complex numbers: namely $\lambda = i$ and $\lambda = -i$. (Recall $i^2 = -1$)

The matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is the matrix of $P : \mathbf{R}^2 \to \mathbf{R}^2$ with respect to the standard basis. From our work on isomorphism in Module 6, this matrix is a representation of P as an element of $\mathcal{M}_{2\times 2}(\mathbf{R})$, the vector space of 2×2 matrices whose entries are real numbers. However, from our work on complex numbers, we recall that every real number is a complex number whose imaginary part is equal to 0. And so instead considering our matrix as an element of $\mathcal{M}_{2\times 2}(\mathbf{R})$, instead we can consider it as an element of $\mathcal{M}_{2\times 2}(\mathbf{C})$. As an element of $\mathcal{M}_{2\times 2}(\mathbf{C})$, this matrix has two eigenvalues: i and $-i$, whereas as an element of $\mathcal{M}_{2\times 2}(\mathbf{R})$ this matrix has no eigenvalues.

We will return to these matters much more carefully in Module 8 when we consider eigenvalues and eigenvectors for linear operators of both real and complex vector spaces. To do so,

we first must spend sometime thinking about zeros of polynomials with complex coefficients.

7.1 Factoring and Division for Polynomials with Complex Coefficients

Back some time in elementary school we learned about division of positive integers. Long before we were comfortable with fractions or decimals, we learned facts like: 50 *divided by* 8 *is* 6 *with a remainder of* 2. Symbolically, we can express this as

$$50 = 8(6) + 2$$

More generally, we are relatively confident that for any pair of non-negative integers p and s with $s \neq 0$, we can divide p by s and possibly get a remainder. We express this idea formally as follows:

Theorem 7.1 (Integer Division Algorithm)**.** *For every pair of non-negative integers p, s with $s \neq 0$, there exist unique integers q and r so that*

$$p = sq + r$$

and $r < s$

In our example above we have $p = 50$, $s = 8$, $q = 6$ and $r = 2$. In saying that $r < s$, we are remembering that when we divide by 8 the remainder must be less than 8. Here, the notation q and r are chosen to remind us of the familiar words *quotient* and *remainder*.

Pairs q and s for which division yields a zero remainder (i.e., $r = 0$) bring us to our usual definition of a factor of a positive integer. For example. 5 is a factor of 50 because we we divide 50 by 5 the remainder is zero. In other words

$$50 = 5(10) + 0$$

Aside. *This theorem is traditionally called the Integer Division Algorithm, but in this context there is no actual algorithm in sight. When presented in the context of greatest common divisors, the Integer Division Algorithm is a pathway to an algorithm for computing the greatest common divisor of two integers.*

Though we may have never seen *division* written down this way, we are intimately familiar with it by way of our personal mathematics histories. Quite likely, our personal mathematics histories also include pre-calculus studies. And so we also familiar with a version of division for polynomials – rational functions.

For example, consider the rational function

$$f(x) = \frac{x^3 - x^2 + 4x - 1}{x^2 + 1}$$

If we perform this division using polynomial long division we arrive at a quotient of $x-1$ and a remainder of $3x$. In other words

$$x^3 - x^2 + 4x - 1 = (x^2+1)(x-1) + 3x$$

Notice that here rather than the $r(x) = 3x$ being less than $s(x) = x^2+1$, instead the degree of r is less than the degree of s. And so just as we can write down a formal theorem for division of integers, so too can we do so for polynomials:

Theorem 7.2 (Polynomial Division Algorithm)**.** *For every pair of polynomials $p, s \in \mathcal{P}(\mathbf{F})$ with $s \neq 0$, there exist unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ so that*

$$p = sq + r$$

and $\deg r < \deg s$.

Aside. *If you have taken MATH364, you might recall that much of the topics in that course arose from the Integer Division Algorithm. One may study corresponding topics following from the Polynomial Division Algorithm. If you are taking MATH362 this semester, you will come to recognize these structures as examples of* division rings.

Just as the case $r = 0$ in the Integer Division Algorithm leads to our definition of integer factor, so too does the case $r(x) = 0$ lead to the meaning of polynomial factor. For example, when $p(x) = x^3 - x^2 + x - 1$ and $s(x) = x^2 + 1$ we have

$$x^3 - x^2 + x - 1 = (x^2+1)(x-1)$$

Here we have $r = 0$ and so we might want to say x^2+1 is a factor of $x^3 - x^2 + x - 1$.

Definition 7.3. *Let $p, s \in \mathcal{P}(\mathbf{F})$ we say s **is a factor of** p when there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ so that $p = qs$.*

In an effort to shorten this week's readings, we'll omit the proof of the Polynomial Division Algorithm. The proof is actually quite interesting[7] and, surprisingly, depends on the Fundamental Theorem of Linear Maps.

Continuing with our previous example $p(x) = x^3 - x^2 + x - 1 = (x^2+1)(x-1)$, notice that since $x-1$ is a linear factor of $p(x)$, we have $p(1) = 0$. As expected from our experiences in pre-calculus, linear factors correspond to zeros of a polynomial.

Theorem 7.4. *Let $p \in \mathcal{P}(\mathbf{F})$ and let $\lambda \in \mathbf{F}$. We have $p(\lambda) = 0$ if and only if there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ so that $p(z) = (z-\lambda)q(z)$*

[7]...depending on your definition of interesting. See result 4.8 on page 121 of the course text for more details

Proof. Let $p \in \mathcal{P}(\mathbf{F})$ and let $\lambda \in \mathbf{F}$.

If there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ so that $p(z) = (z - \lambda)q(z)$, then

$$p(\lambda) = (\lambda - \lambda)q(\lambda) = 0q(\lambda) = 0$$

Therefore λ is a zero of p.

Assume now we have $p(\lambda) = 0$. Let $s(z) = (z - \lambda)$. By the Division Algorithm for polynomials, there exists polynomials q and r so that $\deg r < \deg s$ and

$$p = sq + r$$

Since $\deg s = 1$, it must be that $\deg r = 0$. Therefore r is a constant function. That is, $r(z) = c$ for some $c \in \mathbf{F}$. Therefore

$$p(z) = (z - \lambda)s(z) + c$$

Since $p(\lambda) = 0$ we have

$$0 = p(\lambda) = (\lambda - \lambda)s(\lambda) + c = 0s(\lambda) + c = c$$

Therefore $c = 0$ and so $p(z) = (z - \lambda)q(z)$. In other words, there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ so that $p(z) = (z - \lambda)q(z)$. □

So far the work we have done has simultaneously been for polynomials with coefficients in $\mathbf{C}$ and polynomials with coefficients in $\mathbf{R}$. However, as we saw above when we sought eigenvalues for the matrix corresponding to projection, whether we consider a polynomial as an element of $\mathcal{P}(\mathbf{R})$ or of $\mathcal{P}(\mathbf{C})$ can affect the existence of zeros. The polynomial $x^2 + 1 = 0$ has no zeros when viewed as an element of $\mathcal{P}(\mathbf{R})$, but has two zeros, when viewed as an element of $\mathcal{P}(\mathbf{C})$.

$$x^2 + 1 = (x - i)(x + i)$$

Surprisingly, all non-constant polynomials in $\mathcal{P}(\mathbf{C})$ have at least one zero.

Theorem 7.5 (The Fundamental Theorem of Algebra)**.** *If $p \in \mathcal{P}(\mathbf{C})$ and p is not a constant polynomial, then p has a zero.*

The proof this theorem is well beyond the scope of this course. It requires tools from the study of *mathematical analysis*. In the study of mathematical analysis, one takes the concepts we learn about in calculus and applies them to other collections of objects. This work is similar to the work we are doing in MATH266 in generalizing the concepts for $\mathbf{R}^n$ to arbitrary vector spaces. If you enjoyed learning about concepts in Calculus, and you have enjoyed the approach of MATH266, then you would probably enjoy MATH371 (Real Analysis I) and MATH379 (Complex Analysis).

Aside. *If we can show that the eigenvalues of a linear operator on a finite-dimensional complex vector space are exactly the same as those of a matrix corresponding to the linear operator, then by way of the characteristic equation method to compute an eigenvalue, the Fundamental Theorem of Algebra implies that every linear operator on a finite-dimensional complex vector space has a (possibly complex) eigenvalue. Our approach to proving that every complex linear operator has eigenvalue will not proceed in this way; we will prove this fact without having to grapple with the meaning of determinant for a complex linear operator.*

By combining the statement of the Fundamental Theorem of Algebra with the statement of Theorem 7.4, we arrive at the following fact:

Corollary 7.6. *For every non-constant $p \in \mathcal{P}(\mathbf{C})$ there exists $\lambda \in \mathbf{C}$ and $q \in \mathcal{P}(\mathbf{C})$ so that $p(z) = (z - \lambda)q(z)$.*

To better understand the meaning of this corollary, let us consider an example $p(x) = 2x^4 + 6x^2 + 4$. Viewed as a polynomial with real coefficients, the rational roots test tells us that this polynomial has no roots. However, we may view the coefficients of this polynomial as complex numbers:

$$\begin{aligned} 2 &= 2 + 0i \\ 6 &= 6 + 0i \\ 4 &= 4 + 0i \end{aligned}$$

And so it is possible that there exists a complex number λ so that $p(\lambda) = 0$. One can find (either with a computer, or by guessing and checking) that

$$\begin{aligned} p(i) &= 2i^4 + 6i^2 + 4 \\ &= 2(i^2)(i^2) + 6(i^2) + 4 \\ &= 2(-1)(-1) + 6(-1) + 4 \\ &= 0 \end{aligned}$$

Therefore $\lambda = i$ is a zero of p. And so by Corollary 7.6, there exists $q \in \mathcal{P}(\mathbf{C})$ so that $p(x) = (x - i)q(x)$. Without worrying too much about how one arrives at the value for q, we can find

$$q(x) = 2x^3 + 2ix^2 + 4x + 2i$$

Therefore

$$2x^4 + 6x^2 + 4 = (x - i)(2x^3 + 2ix^2 + 4x + 2i)$$

Since any zero of q will necessarily be a zero of p, to find a second zero of p it suffices to find a zero of q. We find $q(-i) = 0$. And again without worrying too much about how the computation is done, we have

$$2x^3 + 2ix^2 + 4x + 2i = (x + i)(2x^2 + 4)$$

Therefore

$$p(x) = (x - i)(x + i)(2x^2 + 4)$$

We can factor $p(x)$ further by finding a zero of $2x^2 + 4$. Using the quadratic formula we find $2x^2 + 4 = 2(x + \sqrt{2}i)(x - \sqrt{2}i)$. Therefore

$$2x^4 + 6x^2 + 4 = 2(x + i)(x - i)(x + \sqrt{2}i)(x - \sqrt{2}i)$$

When viewed as a polynomial in $\mathbf{R}$, the polynomial $2x^4+6x^2+4$ has no zeros. However, when viewed as a polynomial in $\mathbf{C}$, the polynomial as four zeros: $i, -i, \sqrt{2}i, -\sqrt{2}i$. As $2x^4+6x^2+4$ has four zeros, it factors exactly into four linear factors. This same behaviour holds for every polynomial of degree at least 1 in $\mathcal{P}(\mathbf{C})$.

Theorem 7.7. *If $p \in \mathcal{P}(\mathbf{C})$ and p is not a constant polynomial, then, up to the order of the factors, p factors as*

$$p(z) = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$$

where $c, \lambda_1, \lambda_2, \ldots \lambda_n \in \mathbf{C}$, $c \neq 0$ and $\deg p = n$.

Aside. *For polynomials that are permitted to take complex values, traditionally mathematicians choose to use z as the variable, rather than x.*

Notice that if we change $\mathbf{C}$ to $\mathbf{R}$ in the statement of this theorem it is no longer true. For example, the polynomial $x^2 + 1 \in \mathcal{P}(\mathbf{R})$ cannot be fully factored into linear factors; it has no zeros.

Aside. *Just as the existence of division of positive integers lets us define **prime number**, so too does division of polynomials let us define **prime polynomial.***

In thinking about how we might prove this theorem, let us set our sights significantly lower. Rather than proving the result for every possible degree, let us focus only the case $\deg p = 2$.

Consider $p(z) = a_2z^2 + a_1z + a_0$ with $a_0, a_1, a_2 \in \mathbf{C}$. By Corollary 7.6, there exists $\lambda \in C$ and $q \in \mathcal{P}(\mathbf{C})$ so that

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

Since $p(z)$ has degree 2 and $(z-\lambda)$ is degree 1, it must be that $q(z)$ is a degree 1 polynomial. In other words, we have

$$q(z) = c_1z + c_0$$

Therefore

$$p(z) = c_1(z + \frac{c_0}{c_1})(z - \lambda)$$

Setting $c = c_1$, $\lambda_1 = -\frac{c_0}{c_1}$ and $\lambda_2 = \lambda$ yields the required form for p:

$$p(z) = c(z - \lambda_1)(z - \lambda_2)$$

Therefore every complex polynomial of degree 2 factors exactly into 2 linear factors.

Aiming a little bit higher, let us set our sights now on the case $\deg p = 3$. Consider $p(z) = a_3z^3 + a_2z^2 + a_1z + a_0$ with $a_0, a_1, a_2, a_3 \in \mathbf{C}$. By Corollary 7.6, there exists $\lambda \in C$ and $q \in \mathcal{P}(\mathbf{C})$ so that

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

Since $p(z)$ has degree 3 and $(z-\lambda)$ is degree 1, it must be that $q(z)$ is a degree 2 polynomial. From our previous work, q factors exactly into 2 linear factors. That is,

$$q(z) = c(z - \lambda_1)(z - \lambda_2).$$

Therefore

$$p(z) = c(z - \lambda_1)(z - \lambda_2)(z - \lambda_3)$$

where $\lambda_3 = \lambda$.

Aiming even higher, let us set our sights now on the case $\deg p = 4$. Consider $p(z) = a_4x^4 + a_3z^3 + a_2z^2 + a_1z + a_0$ with $a_0, a_1, a_2, a_3, a_4 \in \mathbf{C}$. By Corollary 7.6, there exists $\lambda \in C$ and $q \in \mathcal{P}(\mathbf{C})$ so that

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

Since $p(z)$ has degree 4 and $(z-\lambda)$ is degree 1, it must be that $q(z)$ is a degree 3 polynomial. From our previous work, q factors exactly into 3 linear factors. That is,

$$q(z) = c(z - \lambda_1)(z - \lambda_2)(z - \lambda_3).$$

Therefore

$$p(z) = c(z - \lambda_1)(z - \lambda_2)(z - \lambda_3)(z - \lambda_4)$$

where $\lambda_4 = \lambda$.

Aiming even higher, let us set our sights now on the case $\deg p = 5$. Wait, no. This is a terrible idea. Repeating the same argument over again will take us literally forever to cover all possible values of $\deg p$. However, the form of this argument is ripe for the application of *proof by induction*.

Depending on which other mathematics courses we have taken, I suspect we have varying levels of comfort with mathematical induction. For those of you who have taken only MATH110 and MATH164, this is likely your first exposure to mathematical induction. Whereas, those of you who have taken MATH163 or CMPT260 are at least partial comfortable with the topic. In either case, I invite you to read the posted appendix entitled *Mathematical Induction* before proceeding.

.
.
.
.
.

.
.
.
.
.
.
.
.
.
.
.
.
.
.
.

Following our reasoning from above, we use the Principle of Mathematical Induction to prove our theorem.

Theorem (The Principle of Mathematical Induction)**.** *Let $P(n)$ be a statement that is either true or false for each integer $n \geq 1$. If the following two statements hold, then $P(n)$ is true for each integer $n \geq 1$.*

1. *$P(1)$ is true; and*
2. *for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true*

Aside 7.8. *If you have seen induction before you might find notable the lack of appearance of the phrases* base case*,* induction hypothesis *and* induction step *in what follows. As long as it is clear that the author is showing that the two statements hold, these phrases need not be used.* Base Case *refers to statement 1.* Induction hypothesis *refers to the hypothesis of statement 2.* Induction step *refers to proving that the conclusion of statement 2 holds, given that the hypothesis is true. This is likely the only time these phrases will appear in these notes.*

To use the Principle of Mathematical Induction to prove Theorem 7.7, we need to show that 1. and 2. hold in the statement of the Principle of Mathematical Induction for some statement $P(n)$ that is relevant to the statement of Theorem 7.7.

Let $P(n)$ be the statement:

Every polynomial of degree n factors as $p(z) = c(z-\lambda_1)(z-\lambda_2)\dots(z-\lambda_n)$ where $c, \lambda_1, \lambda_2, \cdots \lambda_n \in \mathbf{C}$ and $c \neq 0$.

If we can show 1. and 2. hold for this statement $P(n)$, then the Principle of Mathematical Induction tells us that $P(n)$ is true for every $n \geq 1$. In other words, Theorem 7.7 is true!

Proof of Theorem 7.7. Let $P(n)$ be the statement

$$\textit{Every polynomial of degree } n \textit{ factors as } p(z) = c(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_n) \textit{ where } c, \lambda_1, \lambda_2, \ldots \lambda_n \in \mathbf{C} \textit{ and } c \neq 0.$$

To prove our theorem, we prove that $P(n)$ is true for every $n \geq 1$. We proceed by showing that statements 1. and 2. of the Principle of Mathematical Induction hold for $P(n)$.

1. P(1) is true
Let p be a complex polynomial of degree 1. Therefore, $p(z) = c_1 z + c_0$ for $c_0, c_1 \in \mathbf{C}$. Factoring we find

$$p(z) = c_1(z + \frac{c_0}{c_1})$$

By letting $c = c_1$ and $\lambda_1 = -\frac{c_0}{c_1}$, we have

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

Therefore $P(1)$ is true.

2. for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true.
Let $k \geq 1$ be an integer so that $P(k)$ is true. Since $P(k)$ is true, we have that every complex polynomial of degree k factors as

$$c(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_k)$$

for some choice of $c, \lambda_1, \lambda_2, \ldots \lambda_n \in \mathbf{C}$ with $c \neq 0$.

Let p be a complex polynomial of degree $k+1$. By Corollary 7.6, there exists a complex polynomial q of degree k so that

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

Since $q(z)$ has degree k, it factors as

$$q(z) = c(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_k)$$

for some $c, \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbf{C}$ with $c \neq 0$. Therefore

$$p(z) = c(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_k)(z-\lambda)$$

Letting $\lambda = \lambda_{k+1}$ yields

$$p(z) = c(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_k)(z-\lambda_{k+1})$$

Therefore $P(k+1)$ is true.

Since both hypotheses of the Principle of Mathematical Induction hold, it follows that $P(n)$ is true for all $n \geq 1$. □

Aside. *As with all things in reading and writing mathematics, the level of detail in a proof depends on what background knowledge the author can expect the reader to have. For readers well experienced reading proofs, the following proof of Theorem 7.7 would suffice.*

We proceed by induction, noting that when $n = 1$ *we have* $p(z) = c_1 z + c_0 = c_1(z + c_0/c_1)$. *Assume now* $\deg p = k + 1$. *By Corollary 7.6, there exists* $q \in \mathcal{P}(\mathbf{C})$ *and* $c, \lambda \in C$ *so that* $p(z) = c(z - \lambda)q(z)$. *The result follows by induction by noticing* $\deg q = k$ *and if* $p(r) = 0$ *for* $r \neq \lambda$, *then* $q(r) = 0$.

A takeaway here is that when reading mathematics, lack of understanding is not always due to the reader. Meaning in texts can be thought to be made by the relationship between the text and the reader. Whether or not an author is aware, their text has an implied reader[8]. *If you do not possess the same background as the author's implied reader, then your lack of understanding should not be confused with a lack of ability to understand. The former need not imply the latter.*

For each module of this course the Learning Incomes *outline what knowledge the author expects their implied reader to have.*

Recall that for a polynomial, p, and a zero of p, λ, we say λ **has multiplicity** ℓ when λ appears exactly ℓ times when p is expressed as a product of linear factors. For example, if $p(z) = (x - 1)(x - 2)(x + i)(x + i)(x - (2 + 3i))$, then

- 1 has multiplicity 1;
- 2 has multiplicity 1;
- $-i$ has multiplicity 2; and
- $2 + 3i$ has multiplicity 1.

Counted according to their multiplicity, this polynomial has 5 zeros. As a complex polynomial of degree n can be factored exactly into n linear factors, such a polynomial as n zeros when we count with multiplicity.

Corollary 7.9. *Counted with multiplicity, every complex polynomial of degree* n *has* n *zeros.*

[8] see https://en.wikipedia.org/wiki/Reader-response_criticism

7.1 Test Your Understanding

1. Let $p(z) = (3i)z^5 + (3i)z^4 + (3i)z^3$ be a complex polynomial. Find all of the zeroes of p by writing p as a product of exactly 5 linear factors. Remember that a linear factor of z can be written as $(z - 0)$. What is the multiplicity of each zero of p?

2. Let $A \in \mathcal{M}_{2\times 2}(\mathbf{C})$. That is, let A be a 2×2 matrix with entries in $\mathbf{C}$. Using the Fundamental Theorem of Algebra, show A has at least one eigenvalue. (Assume that the method from MATH164 of using determinants to find eigenvalues also works for complex-valued matrices. Also assume that the procedure to find the determinant of a 2×2 complex matrix is the same as the procedure to find the determinant of a 2×2 real matrix.)

7.1 Test Your Understanding Solution

1. Factoring we have $p(z) = (3i)(z^3)(z^2 + z + 1)$. Factoring $q(z) = z^2 + z + 1$ by using the quadratic formula (and remembering $i^2 = -1$), we find

$$q(z) = \left(z - \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)\right)\left(z - \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)\right)$$

Therefore

$$p(z) = 3i(z-0)(z-0)(z-0)\left(z - \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)\right)\left(z - \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)\right)$$

And so we see p has zeroes at $z = 0$ (with multiplicity 3), $z = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$ (with multiplicity 1) and $z = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$ (with multiplicity 1).

2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a, b, c, d \in \mathbf{C}$.

Zeroes of the polynomial $p(z) = (a - z)(d - z) - bc$ are values $\lambda \in \mathbf{C}$ for which the determinant of the matrix $A - \lambda I_2 = \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix}$ is 0.

Simplifying we find $p(z) = z^2 - (a + d)z + (ad - bc)$. Since $p(z)$ is a non-constant polynomial in $\mathbf{C}$, then by the Fundamental Theorem of Algebra, p has at least one zero. Let $\lambda \in \mathbf{C}$ be a zero of p. Therefore $det(A - \lambda I_2) = 0$. Therefore the matrix $A - \lambda I_2$ is not invertible. Therefore the matrix $A - \lambda I_2$ has a non-trivial null space. Therefore there exists $v \neq 0$ so that $(A - \lambda I_2)v = 0$. Expanding and rearranging the equation $(A - \lambda I_2)v = 0$ gives $Av = \lambda v$. Therefore λ is an eigenvalue of A with corresponding eigenvector v.

Therefore every 2×2 matrix in $\mathbf{C}$ has at least one eigenvalue.

---

8 Eigenvalues and Eigenvectors of Linear Operators (LADR 5)

Learning Incomes.

- *Recall the definition of eigenvalue and eigenvector of a matrix.*
- *Recall the method of characteristic polynomial as a method to find the eigenvalues of a matrix.*
- *Understand the statement of Theorems 6.9 and 7.7*

Learning Outcomes.

- *Understand the relationship between the equations* $Tv = \lambda v$ *and* $(T - \lambda I)v = 0$.
- *Understand why the eigenvalues corresponding to* λ *are the non-zero elements of null* $T - \lambda I$
- *Understand the statement and proof of Theorem 8.4*
- *Recognize the implications of considering a matrix as an element of* $\mathcal{M}_{n\times n}(\mathbf{C})$ *as opposed to* $\mathcal{M}_{n\times n}(\mathbf{R})$
- *Be able to determine if a matrix is diagonalizable given its list of eigenvalues.*

Newly Defined Terms and Notation.

- *eigenvalue of a linear operator, eigenvector of a linear operator, eigenbasis, diagonalizable*

Representing linear maps as matrices are an essential tool in computation. That any linear map of finite-dimensional vector spaces can be represented can be represented as a matrix means that algorithms and techniques for working with matrices can be used to answer questions about linear maps.

Though algorithms and computation are not a main focus of this course, they serve as an excellent motivation for the work in this coming section. Back in Module 5 we showed that every linear map on a finite-dimensional vector space can be represented as a matrix. What's more, this representation as a matrix is not unique; we are free to choose any basis we want for the domain and the codomain to construct our matrix. It turns out that by carefully choosing a basis for the domain and codomain of a linear we can (sometimes) construct a corresponding matrix that is efficient in computation. We restrict our attention in this section to linear maps from one vector space to itself. In other words, we restrict our attention to linear operators. Further, we choose the same basis to be the same for both the domain and the codomain of the linear operator when constructing a corresponding matrix.

When we use the same basis B for the domain and codomain for the matrix of a linear operator T on V, we say $\mathcal{M}(T)$ *is the matrix of* T *with respect to* B.

We begin with an example. Let $V = \mathbf{R}^3$ and let T be the linear operator whose matrix with respect to the standard basis vectors is:

$$\mathcal{M}(T) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix}$$

We can find the image of an element of $\mathbf{R}^3$ with respect to this map by computing $\mathcal{M}(T)\mathcal{M}(v)$, where $\mathcal{M}(v)$ is constructed using the standard basis. For example,

$$T(-2,0,6) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 1(-2)+0(0)+0(6) \\ 1(-2)+2(0)+1(6) \\ 1(-2)+2(0)+1(6) \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ 4 \end{bmatrix}$$

Therefore $T(-2,0,6) = (-2,4,4)$.

However, by choosing a different bases in constructing our matrix, we can arrive at a much easier computation. Let $B_V = (0,-1,2),(-2,1,1),(0,1,1)$. By first finding the image of each element of B_V using the matrix above, we can then compute the matrix with respect to this basis. Constructing the matrix of T with respect to this basis yields

$$\mathcal{M}(T) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

To find the image of $v = (-2,0,6)$ we construct the matrix of v with respect to B_V. To do so, we write $(-2,0,6)$ as a linear combination of elements of B_V:

$$(-2,0,6) = 2(0,-1,2) + 1(-2,1,1) + 1(0,1,1)$$

Therefore with respect to B_V we have $\mathcal{M}(v) = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$.

We compute

$$\mathcal{M}(T)\mathcal{M}(v) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0(2)+0(1)+0(1) \\ 0(2)+1(1)+0(1) \\ 0(2)+0(1)+3(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

The matrix $\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$ is the matrix of a vector v with respect to B_V. We compute

$$v = 0(0,-1,2) + 1(-2,1,1) + 3(0,1,1) = (-2,4,4)$$

Therefore $T(-2,0,6) = (-2,4,4)$

Though this second calculation took a little more work than the first, this extra work paid off when we computed the product of the matrix and the vector. Having a diagonal matrix in the second calculation made for a computation with lots of multiplication by 0.

To better understand what we are seeing here, let us compute Tv for each $v \in B_V$. Using the first matrix, we find

$$T(0,-1,2) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad T(-2,1,1) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

$$T(0,1,1) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$$

In other words

$$T(0,-1,2) = 0(0,-1,2) \quad T(-2,1,1) = 1(-2,1,1) \quad T(0,1,1) = 3(0,1,1)$$

The diagonal entries of $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ are eigenvalues of T. And, their corresponding eigenvectors are the elements of B_V!

As briefly discussed at the start of Module 7, our ideas of eigenvalue and eigenvector of a matrix extend easy to linear maps. As we have seen above, matrices formed with respect to a basis of eigenvectors give rise to representations of linear maps that are efficient in computation.

8.1 Eigenvalues of Linear Operators (LADR 5A)

We begin with two definitions. These definitions generalize the corresponding definitions for matrices from MATH164.

Definition 8.1. *Let V be a vector space, let T be a linear operator on V and let $\lambda \in \mathbf{F}$. We say λ* ***is an eigenvalue of*** *T when there exists $v \in V$ with $v \neq 0$ so that $Tv = \lambda v$.*

Definition 8.2. *Let V be a vector space, let T be a linear operator, let $\lambda \in \mathbf{F}$ be an eigenvalue of T and let $v \in V$ so that $v \neq 0$. We say v* ***is an eigenvector of*** *T* ***corresponding to*** *λ when $Tv = \lambda v$.*

For example, let $\mathcal{D}(\mathbf{R})$ be the set of all differentiable functions from $\mathbf{R}$ to $\mathbf{R}$ and let $D : \mathcal{D}(\mathbf{R}) \to \mathcal{D}(\mathbf{R})$ be the linear map given by $Df = \frac{d}{dx}f(x)$ for all $f \in \mathcal{D}(\mathbf{R})$. Since $D(e^{2x}) = 2e^{2x}$, we have that $f(x) = e^{2x}$ is an eigenvector of D that corresponds to the eigenvalue $\lambda = 2$.

From our work in MATH164, perhaps we recall that our method to find an eigenvalue by finding the zeros of a characteristic polynomial found its first step is expressing a scalar multiplication in $\mathbf{R}^n$ as the product of a scalar, a matrix and a vector:

$$\lambda v = \lambda I_n v$$

We use the analogue of the identity matrix, the identity function, to achieve the same end for linear maps:

$$\lambda v = \lambda I v$$

where $I : V \to V$ is the linear map for which $Iu = u$ for every $u \in V$. Since V is a vector space and $\lambda I v \in V$, this vector has an additive inverse: $-(\lambda I v)$. And so

$$\begin{aligned} Tv &= \lambda v \\ Tv &= \lambda I v \\ Tv + -(\lambda I v) &= \lambda I v + -(\lambda I v) \\ (T - \lambda I)v &= 0 \end{aligned}$$

On the left, the notation $T - \lambda I$ refers to a linear map: $T - \lambda I : V \to V$ so that

$$(T - \lambda I)u = Tu - \lambda(Iu) = Tu - \lambda u$$

for all $u \in U$.

If λ is an eigenvalue of T and v is a corresponding eigenvector, then $(T - \lambda I)v = 0$. In other words, v is in the null space of the linear map $(T - \lambda I)$. Since $v \neq 0$, it then follows that $\textit{null}\ (T - \lambda I) \neq \{0\}$. By Theorem 4.9, the linear map $(T - \lambda I)$ is not injective.

Similarly, if $(T - \lambda I)$ is not injective for some $\lambda \in \mathbf{F}$, then $(T - \lambda I)$ has a non-trivial null space. That is, there exists $v \in null\ (T - \lambda I)$ so that $v \neq 0$. Working backwards though our algebraic manipulations above implies

$$Tv = \lambda v$$

Therefore λ is an eigenvalue of T and v is an eigenvector corresponding to λ.

Combining these two arguments yields the following result.

Lemma 8.3. *Let V a finite dimensional vector space and let $\lambda \in \mathbf{F}$. The linear map $T - \lambda I$ is not injective if and only if λ is an eigenvalue of T.*

Lemma 8.3 tells us non-injectivity is equivalent to λ being an eigenvalue. And so we can add a statement about eigenvalues to our list of equivalent statements in Theorem 6.9.

Theorem 8.4. *Let V be a finite dimensional vector space, let T be a linear operator on V and let $\lambda \in \mathbf{F}$. The following four statements are equivalent*

i. *λ is an eigenvalue of T.*

ii. *$T - \lambda I$ is not injective.*

iii. *$T - \lambda I$ is not invertible.*

iv. *$T - \lambda I$ is not surjective.*

Proof. Let V be a finite dimensional vector space, let T be a linear operator on V and let $\lambda \in \mathbf{F}$.

Statements *i.* and *ii.* are equivalent by Lemma 8.3. Statements *ii.*, *iii.* and *iv.* are equivalent by Theorem 6.9. Therefore all four statements are equivalent. □

From our work in MATH164, we recall that not every matrix admits an eigenvalue. For example, using the characteristic equation method we would find that the matrix $A = \begin{bmatrix} 5 & 9 \\ -3 & -5 \end{bmatrix}$ has characteristic equation $\lambda^2 + 2$. This polynomial has no zeros and thus A has no eigenvalues. However, viewing A as an element of $\mathcal{M}_{2\times 2}(\mathbf{C})$ rather than as an element of $\mathcal{M}_{2\times 2}(\mathbf{R})$ yields the following factorization of the polynomial above.

$$\lambda^2 + 2 = (\lambda - \sqrt{2}i)(\lambda + \sqrt{2}i)$$

By the Fundamental Theorem of Algebra, perhaps we are then convinced that every matrix in $\mathcal{M}_{2\times 2}(\mathbf{C})$ has an eigenvalue. Using the interpretation of complex linear operators as complex valued matrices, perhaps then we can conclude that every linear operator on a complex vector space of dimension 2 has an eigenvalue.

This seems like a lot of work! To do this, we would have to prove that our ideas for determinants extend to complex matrices. And then we would have to prove that the eigenvalues of a linear map are the same as the eigenvalues of any representation of the linear operator as a matrix. Rather than this approach, we take a more direct approach to proving that every complex linear operator has an eigenvalue.

Let V be a finite-dimensional vector space and let $\mathcal{L}(V, V)$ be the set of all linear operators on V. As $\mathcal{L}(V, V)$ is a vector space, we can add and scalar multiply linear operators. If S and T are linear operators on V, then $S+T$ is the linear operator on V so that $(S+T)v = Sv+Tv$ for each $v \in V$. If $\lambda \in \mathbf{F}$, then λS is the linear operator on V so that $(\lambda S)v = \lambda(Sv)$ for each $v \in V$.

Aside. *The piece of notation: $(S+T)v = Sv+Tv$ is not an application of the distributivity property of vector spaces. In this context, v is not a scalar; it is an element of V, not an element of $\mathbf{F}$. This piece of notation defines for us what we mean when we write $S+T$. For $S, T \in \mathcal{L}(V, V)$, the sum of S and T is the element of $\mathcal{L}(V, V)$ so that $(S+T)v = Sv+Tv$.*

In Module 6 we considered function composition as an analogue for matrix multiplication. Recall, the linear operator $S \circ T$ is defined to be the linear operator on V so that $(S \circ T)v = S(T(v))$ for each $v \in V$. If we denote $T \circ T$ as T^2, then we can make meaning of a expression like:

$$T^2 + 8T + 15I$$

This expression is a sum of linear operators. Since the set of linear operators is a vector space, it is closed with respect to addition, $T^2 + 8T + 15I \in \mathcal{L}(V, V)$. By remembering the meaning of each of these pieces of notation we have

$$\begin{aligned}(T^2 + 8T + 15I)v &= T^2v + 8Tv + 15Iv \\ &= (T \circ T)v + 8Tv + 15v\end{aligned}$$

Even though there are no variables in sight, expressed as $T^2+8T+15I$, the expression looks a lot like a polynomial! It turns out that factoring this expression as if it were a polynomial leads to something meaningful

$$T^2 + 8T + 15I = (T + 3I)(T + 5I)$$

Looking at this we wonder why we have written $T+3I$ rather than $T+3$. Since T is a linear operator and 3 is a scalar adding them together carries no meaning. However, we can add T to the linear operator $3I$.

Remember that what appears to be multiplication is actually function composition in disguise And so we have

$$(T + 3I)(T + 5I) = (T + 3I) \circ (T + 5I)$$

Consider the equation

$$(T + 3I)(T + 5I)v = 0$$

To have equality here, it must be that v is an element of the null space of the linear map $T^2 + 8T + 15I$. If v is not in the null space of the linear map $(T + 5I)$, then $(T + 5I)v \neq 0$. Let $w = (T + 5I)v$. Remembering the meaning of function composition, we have

$$0 = (T + 3I) \circ (T + 5I)v = (T + 3I)\left((T + 5I)(v)\right) = (T + 3I)w$$

And so if v is not in the null space of $(T + 5I)$, then w is in the null space of the linear map $(T + 3I)$.

If $v \neq 0$, then the null space of $T^2 + 8T + 15I$ is non trivial. From our work above, having $v \neq 0$ further implies that at least one of $(T + 3I)$ or $(T + 5I)$ has a non-trivial null space. Which in turn implies that at least one of $(T+3I)$ or $(T+5I)$ is not injective. From Theorem 8.4, it then follows that at least one of -3 or -5 is an eigenvalue of T.

Let us generalize this line of reasoning. As above, we can make meaning of an expression like:

$$a_0Iv + a_1Tv + a_2T^2v + a_3T^3v + \cdots + a_nT^nv$$

For $v \in V$ and For scalars $a_0, a_1, \ldots, a_n$, this expression results in an element of V.

Using the definition of addition for linear maps, we have

$$a_0Iv + a_1Tv + a_2T^2v + a_3T^3v + \cdots + a_nT^nv = (a_0I + a_1T + a_2T^2 + a_3T^3 + \cdots + a_nT^n)v$$

Squinting a little and forgetting that T is a linear map, the expression

$$a_0I + a_1T + a_2T^2 + a_3T^3 + \cdots + a_nT^n$$

looks like a polynomial!

If $a_0, a_1, \ldots, a_n \in \mathbf{C}$, then Theorem 7.7 implies the existence of the following factorization:

$$a_0I + a_1T + a_2T^2 + a_3T^3 + \cdots + a_nT^n = c(T - \lambda_1 I)(T - \lambda_2 I)\cdots(T - \lambda_n I)$$

for $c, \lambda_1, \lambda_2, \ldots, \lambda_n \in C$ with $c \neq 0$.

Following the line of reasoning from our work above, if $a_0I + a_1T + a_2T^2 + a_3T^3 + \cdots + a_nT^n$ has a non-trivial null space, then there exists $j \in \{0, 1, 2, \ldots, n\}$ so that $(T - \lambda_j I)$ is injective. From Theorem 8.4, this implies that λ_j is an eigenvalue of T. We use these ideas to prove that every complex linear operator admits an eigenvalue.

Theorem 8.5. *Let V be a finite-dimensional complex vector space so that $V \neq \{0\}$. Every linear operator on V has an eigenvalue.*

Proof. Let V be a finite-dimensional complex vector space so that $V \neq \{0\}$. Let $n = \dim V$ Let $v \in V$ so that $v \neq 0$. Let $T : V \to V$ be a linear map.

Consider the following list of elements of V:

$$(Iv, Tv, T^2v, T^3v, \ldots, T^nv)$$

Since $\dim V = n$ and this list has $n + 1$ entries, this list is linearly dependent. Therefore there exist complex scalars $a_0, a_1, \ldots, a_n$ so that

$$a_0Iv + a_1Tv + \cdots + a_nT^nv = 0$$

and at least one scalar is non-zero.

Recalling the meaning of addition in $\mathcal{L}(V, V)$, we have

$$a_0v + a_1Tv + \cdots + a_nT^nv = (a_0I + a_1T + a_2T^2 + \cdots + a_nT^n)v$$

Therefore

$$(a_0I + a_1T + a_2T^2 + \cdots + a_nT^n)v = 0$$

By Theorem 7.7 this polynomial in T factors into exactly n linear factors. That is, there exists $c, \lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbf{C}$ so that $c \neq 0$ and

$$(a_0I + a_1T + a_2T^2 + \cdots + a_nT^n)v = (c(T - \lambda_1I)(T - \lambda_2I) \cdots (T - \lambda_nI))v = 0$$

Remembering that the operation that looks like multiplication here is actually function composition, we have:

$$0 = ((T - \lambda_1I) \circ (T - \lambda_2I) \circ \cdots \circ (T - \lambda_nI))\, v$$

As we evaluate from the right, eventually we reach an index j so that $(T - \lambda_jI)$ evaluated at a non-zero vector yields 0. Therefore $(T - \lambda_jI)$ has a non-trivial null space. Therefore $(T - \lambda_jI)$ is not injective. And so by Theorem 8.4, λ_j is an eigenvalue of T. □

8.1 Test Your Understanding

1. Let V be a finite-dimensional vector space and let T be a linear operator on V. Show T is invertible if and only if 0 is not an eigenvalue of T.

2. Let V be a complex vector space of dimension 2 and let $v \in V$ so that $v \neq 0$. Let T be a linear operator on V so that $T^2v + 4Iv = 0$. Show that T has at least one eigenvalue whose imaginary part is non-zero.

8.1 Test Your Understanding Solution

1. Let V be a finite-dimensional vector space and let T be a linear operator on V. Let $\lambda = 0$. By Theorem 8.4 the statements λ *is not an eigenvalue of* T and $T - \lambda I$ *is invertible* are equivalent. For $\lambda = 0$, we have $T - \lambda I = T$. Therefore the statements 0 *is not an eigenvalue of* T and T *is invertible* are equivalent. In other words T is invertible if and only if 0 is not an eigenvalue of T.

2. Using the definition of addition for linear operators we have $T^2v+4Iv = (T^2+4I)v = 0$. Factoring using the quadratic formula we find $T^2 + 4I = (T + 2iI)(T - 2iI)$. Let $w = (T - 2iI)v$.

   If $w = 0$, then *null* $(T - 2iI)$ is non-trivial. In other words $(T - 2iI)$ is not injective. By Theorem 8.4, $\lambda = 2i$ is an eigenvalue of T.

   Otherwise if $w \neq 0$, then $(T + 2iI)w = 0$. A similar argument shows $\lambda = -2i$ is an eigenvalue of T.

   Therefore T has at least one eigenvalue whose imaginary part is non-zero.

8.2 Eigenvectors and Diagonal Matrices

Following our work introducing eigenvalues in the previous section, we now work towards answering the question that was raised in the introduction – which linear operators admit representative matrices that have lots of zeros? From our work in the introduction, we considered a basis that was formed from eigenvectors of a matrix. We begin by proving that such a basis can even exist by proving a statement about linear independence of eigenvectors.

Theorem 8.6. *Let V be a finite-dimensional vector space and let T be a linear operator on T. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of T with corresponding eigenvectors $v_1, v_2, \ldots, v_m$. The list $(v_1, v_2, \ldots, v_m)$ is linearly independent.*

Proof. Let V be a finite-dimensional vector space and let T be a linear operator on T. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of T with corresponding eigenvectors $v_1, v_2, \ldots, v_m$ We proceed to show the list $(v_1, v_2, \ldots, v_m)$ is linearly independent using a proof by contradiction.

Assume that the list $(v_1, v_2, \ldots, v_m)$ is not linearly independent. Since the list (v_1) is linearly independent and the list $(v_1, v_2, \ldots, v_m)$ is linearly dependent, there is a smallest integer k such that the list $(v_1, v_2, \ldots, v_{k-1})$ is linearly independent, but the list $(v_1, v_2, \ldots, v_{k-1}, v_k)$ is linearly dependent.

Since $(v_1, v_2, \ldots, v_{k-1})$ is linearly independent and $(v_1, v_2, \ldots, v_{k-1}, v_k)$ is linearly dependent the vector v_k can be written as a linear combination of elements of the list $(v_1, v_2, \ldots, v_{k-1})$. In other words, $v_k \in span(v_1, v_2, \ldots, v_{k-1})$. Therefore there exists scalars $a_1, a_2, \ldots, a_{k-1}$ so that

$$v_k = a_1 v_1 + a_2 v_2 + \cdots + a_{k-1} v_{k-1} \tag{9}$$

As the definition of eigenvector requires an eigenvector to be non-zero, we have $v_k \neq 0$. Therefore at least one of these scalars is non zero.

Consider applying T to both sides of this equation:

$$Tv_k = T(a_1 v_1 + a_2 v_2 + \cdots + a_{k-1} v_{k-1})$$

Since T is linear we have

$$Tv_k = a_1 Tv_1 + a_2 Tv_2 + \cdots + a_{k-1} Tv_{k-1}$$

Since each of $v_1, v_2, \ldots, v_k$ are eigenvectors of T we have

$$\lambda_k v_k = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1} \tag{10}$$

Multiplying equation (9) by λ_k and subtracting it from equation (10) yields

$$0 = a_1(\lambda_k - \lambda_1)v_1 + a_2(\lambda_k - \lambda_2)v_2 + \ldots a_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1}$$

Since the list $(v_1, v_2, \ldots, v_{k-1})$ is linearly independent the only linear combination of the list that yields the zero vector is the one with all zero coefficients. Therefore $a_1(\lambda_k - \lambda_1) = 0, a_2(\lambda_k - \lambda_2) = 0, \ldots, a_{k-1}(\lambda_k - \lambda_{k-1}) = 0$.

By hypothesis $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues. Therefore $\lambda_k - \lambda_1 \neq 0, \lambda_k - \lambda_2 \neq 0, \ldots, \lambda_k - \lambda_{k-1} \neq 0$. Therefore $a_1 = a_2 = \cdots = a_{k-1} = 0$. This is a contradiction because above we concluded that at least one of these scalars is non-zero.

Therefore the list $(v_1, v_2, \ldots, v_m)$ is linearly independent. □

A linearly independent list in a vector space of dimension n cannot have more than n entries, which implies the following:

Corollary 8.7. *A linear operator on a vector space of dimension n has at most n distinct eigenvalues.*

A linearly independent list of length n in a vector space of dimension n is necessarily a basis, which implies the following:

Corollary 8.8. *Let V be a finite-dimensional vector space and let T be a linear operator. If T has n distinct eigenvalues, then a list of corresponding eigenvectors is a basis for V.*

Unfortunately, this is not an *if and only if* statement. The converse of the statement of Corollary 8.8 is not true. That is, it is false to say:

If V has a basis consisting of eigenvectors, then T has n distinct eigenvalues

We examine this further on Assignment 4.

From our work in the introduction, a basis consisting of eigenvectors seems particularly useful in constructing matrices with lots of zeros. And so, we give a special name for such a basis:

Definition 8.9. *Let V be a finite-dimensional vector space, let T be a linear operator on V and let B_V be a basis of V. We say B_V* ***is an eigenbasis of*** *V* ***with respect to*** *T when every vector in B_V is an eigenvector of T.*

Let V be a finite-dimensional vector space of dimension $n = 4$, let T be a linear operator and let $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ be distinct eigenvalues of T. Let $E_V = (v_1, v_2, v_3, v_4)$ where for each $i \in \{1, 2, 3, 4\}$, the vector v_i is an eigenvector corresponding to λ_i. By Corollary 8.8, E_V is a basis for T. Therefore E_V is an eigenbasis of V with respect to T.

As T is a linear operator, it can represented as a matrix with respect to E_V. To do this, we first compute Tv_i for each $i \in \{1, 2, 3, 4\}$. Since v_i is an eigenvector corresponding to λ_i for each $i \in \{1, 2, 3, 4\}$, we have

$$\begin{aligned} Tv_1 &= \lambda_1 v_1 \\ Tv_2 &= \lambda_2 v_2 \\ Tv_3 &= \lambda_3 v_3 \\ Tv_4 &= \lambda_4 v_4 \end{aligned}$$

To construct $\mathcal{M}(T)$ with respect to the basis E_V we now express each Tv_i as a linear combination of elements of E_V.

$$\begin{aligned} Tv_1 &= \lambda_1(v_1) + 0(v_2) + 0(v_3) + 0(v_4) \\ Tv_2 &= 0(v_1) + \lambda_2(v_2) + 0(v_3) + 0(v_4) \\ Tv_3 &= 0(v_1) + 0(v_2) + \lambda_3(v_3) + 0(v_4) \\ Tv_4 &= 0(v_1) + 0(v_2) + 0(v_3) + \lambda_4(v_4) \end{aligned}$$

Therefore

$$\mathcal{M}(T) = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$$

The matrix of T with respect to a basis consisting of eigenvectors is a diagonal matrix! Computation with such a matrix is particularly efficient!

Aside. *Representing a linear operator with a diagonal matrix is particularly useful when we want to repeatedly apply the operator. This idea is particularly important in the study of Markov chains. If we can find a diagonal matrix D to use instead of our transition matrix, then determining the state of our probabilistic system after k time-steps amounts to computing D^k; an easy task for a diagonal matrix.*

Unfortunately, learning about Markov chains and transition matrices does not require students to know very much linear algebra. This is why techniques in studying Markov chains can feel like magic *when introduced purely from an applied context. Markov chain methods make a lot more sense what we recognize that a Markov process is a linear operator.*

Conversely, let V be a vector space of dimension $n = 4$ and let S be a linear operator of V. Let $B_V = (w_1, w_2, w_3, w_4)$ be a basis for V so that the following matrix is the matrix of S

with respect to B_V

$$\mathcal{M}(S) = \begin{bmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ 0 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & \alpha_4 \end{bmatrix}$$

By construction we have

$$\begin{aligned} Sw_1 &= \alpha_1(w_1) + 0(w_2) + 0(w_3) + 0(w_4) \\ Sw_2 &= 0(w_1) + \alpha_2(w_2) + 0(w_3) + 0(w_4) \\ Sw_3 &= 0(w_1) + 0(w_2) + \alpha_3(w_3) + 0(w_4) \\ Sw_4 &= 0(w_1) + 0(w_2) + 0(w_3) + \alpha_4(w_4) \end{aligned}$$

Simplifying we find

$$\begin{aligned} Sw_1 &= \alpha_1 w_1 \\ Sw_2 &= \alpha_2 w_2 \\ Sw_3 &= \alpha_3 w_3 \\ Sw_4 &= \alpha_4 w_4 \end{aligned}$$

In other words, each of $\alpha_1, \alpha_2, \alpha_3$, and α_4 are eigenvalues of S with corresponding eigenvectors w_1, w_2, w_3, w_4. Therefore B_V is an eigenbasis of V with respect to T.

From our work in MATH164, perhaps we recognize that a matrix that is zero everywhere except possibly on the main diagonal is called **diagonal matrix**. With this terminology in mind, we define a similar notion for a linear operator.

Definition 8.10. *Let V be a finite-dimensional vector space and let T be a linear operator on V. We say T **is diagonalizable** when there exists a basis B_V of V so that the matrix of T with respect to B_V is a diagonal matrix.*

From our work above, we have the following result.

Theorem 8.11. *Let V be a finite-dimensional vector space and let T be a linear operator on V. The linear map T is diagonalizable if and only of V has an eigenbasis with respect to T.*

To contextualize, let us dive into an example. Let $T : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_2(\mathbf{R})$ so that $T(1) = 1$, $T(x) = x$ and $T(x^2) = 1 + x$. With respect to the basis $B_V = (1, x, x^2)$, this linear map has

the following matrix

$$\mathcal{M}(T) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

This is not a diagonal matrix and so B_V is not an eigenbasis with respect to T. To attempt to find a diagonal matrix for T, we need to find an eigenbasis for V with respect to T. That is, we need to find three linearly independent eigenvectors of T. By observation, $v = 1$ and $v = x$ are eigenvectors corresponding to $\lambda = 1$.

To discover more possible eigenvalue of T, we find the eigenvalues of the matrix $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Aside. *We haven't proven that the eigenvalues of the linear operator are the same as the eigenvalues of a corresponding matrix. We will do this on an upcoming assessment.*

We compute the characteristic polynomial of A:

$$det(A - I\lambda) = (1 - \lambda)^2(-\lambda)$$

This polynomial has zeros at $\lambda = 0$ and $\lambda = 1$.

To construct an eigenbasis for $\mathcal{P}_2(\mathbf{R})$ with respect to T, we want to find a linearly independent list of length 3 consisting of eigenvectors. Each of these three eigenvectors correspond to either $\lambda = 0$ or $\lambda = 1$. Let us begin by finding eigenvectors corresponding to $\lambda = 0$.

For $\lambda = 0$, a corresponding eigenvector $v \neq 0$ satisfies

$$Tv = 0v = 0$$

Therefore any non-zero element of the null space of T is an eigenvector corresponding to $\lambda = 0$. An element $v = a_0 + a_1x + a_2x^2$ is in the nullspace of T when $Tv = 0$. In other words

$$\begin{aligned} 0 + 0x + 0x^2 &= T(a_0 + a_1x + a_2x^2) \\ &= a_0T1 + a_1Tx + a_2Tx^2 \\ &= a_0 + a_1(x) + a_2(1 + x) \\ &= (a_0 + a_2) + (a_1 + a_2)x \end{aligned}$$

These manipulations give rise to the following system of equations

$$\begin{array}{rcrcl} a_0 & & + & a_2 & = & 0 \\ & a_1 & + & a_2 & = & 0 \end{array}$$

The set of solutions to this system is given by $(a_0, a_1, a_2) = t(-1, -1, 1)$ for $t \in \mathbf{R}$. Each of these solutions corresponds to a polynomial $a_0 + a_1x + a_2x^2$ in the null space of T

Therefore

$$null\ T = \{a_0 + a_1x + a_2x^2 = -t - xt + tx^2 | t \in \mathbf{R}\}$$

And so every element of $\mathcal{P}_2(\mathbf{R})$ of the form $-t - tx + tx^2$ is an eigenvector corresponding to the eigenvalue $\lambda = 0$.

For $\lambda = 1$, a corresponding eigenvector $v \neq 0$ satisfies

$$Tv = 1v$$

Rearranging, we find $(T - I)v = 0$. Therefore any non-zero element of the null space of $T - I$ is an eigenvector corresponding to $\lambda = 1$. An element $v = a_0 + a_1x + a_2x^2$ is in the nullspace of $T - I$ when $Tv - Iv = 0$. In other words

$$\begin{aligned} 0 + 0x + 0x^2 &= T(a_0 + a_1x + a_2x^2) - I(a_0 + a_1x + a_2x^2) \\ &= a_0T1 + a_1Tx + a_2Tx^2 - a_0I1 - a_1Ix - a_2Ix^2 \\ &= a_0 + a_1(x) + a_2(1 + x) - a_0 - a_1x - a_2x^2 \\ &= a_2 + a_2x - a_2x^2 \end{aligned}$$

These manipulations give rise to the following system of equations

$$\begin{aligned} a_2 &= 0 \\ a_2 &= 0 \\ a_2 &= 0 \end{aligned}$$

The set of solutions to this system is given by $(a_0, a_1, a_2) = s(1, 0, 0) + t(0, 1, 0)$ for $s, t \in \mathbf{R}$. Each of these solutions corresponds to a polynomial $a_0 + a_1x + a_2x^2$ in the null space of $T - I$ Therefore

$$null\ T - I = \{a_0 + a_1x + a_2x^2 = s + tx \mid s, t \in \mathbf{R}\}$$

And so every element of $\mathcal{P}_2(\mathbf{R})$ of the form $s + tx$ is an eigenvector corresponding to the eigenvalue $\lambda = 1$.

Back to the problem at hand – we seek a list of three linearly independent elements of $\mathcal{P}_2(\mathbf{R})$, each of which are eigenvectors of T. From our work above, every eigenvector of T is of the form $-t - tx + tx^2$ or of the form $s + tx$. By observation each element of the the list $(1, x, 1 + x - x^2)$ is one of the two required forms. Further, this list is linearly independent. Since this list has three elements, it is an eigenbasis of $\mathcal{P}_2(\mathbf{R})$ with respect to T.

Since $T1 = 1$, $Tx = 1x$ and $T(1 + x - x^2) = 0$, the matrix of T with respect to this basis is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

And so we see T is diagonalizable.

8.2 Test Your Understanding

1. Let V be a vector space of dimension 3. Let T be a linear operator on V so that $\lambda_1 = 1$, $\lambda_2 = 1 + 3i$, $\lambda_3 = 1 - 3i$ are the eigenvalues of T.

   (a) Explain how you know T is diagonalizable.

   (b) Let $E = (v_1, v_2, v_3)$ be an eigenbasis of V, whose vectors correspond respectively to the eigenvalues $\lambda_1 = 1, \lambda_2 = 1 + 3i$ and $\lambda_3 = 1 - 3i$. Find the matrix of T with respect to E.

2. Let V be a vector space of dimension 3. Let T be a linear operator on V. Let λ_1 be the only eigenvalues of T. Show that if $\dim null\ (T - I\lambda_1) = 1$, then T is not diagonalizable.

8.2 Test Your Understanding Solution

1. (a) Since T has three distinct eigenvalues, then by Corollary 8.8 T is diagonalizable.

   (b) To find the matrix of T with respect to E we first compute the image of each element of E, and then write them as a linear combination of elements of E.

$$Tv_1 = \lambda_1 v_1 = \lambda_1(v_1) + 0(v_2) + 0(v_3)$$
$$Tv_2 = \lambda_2 v_2 = 0(v_1) + \lambda_2(v_2) + 0(v_3)$$
$$Tv_3 = \lambda_3 v_3 = 0(v_1) + 0(v_2) + \lambda_3(v_3)$$

   Substituting our values for each of λ_1, λ_2 and λ_3 and constructing the matrix we find:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1+3i & 0 \\ 0 & 0 & 1-3i \end{bmatrix}$$

2. Let V be a vector space of dimension 3. Let T be a linear operator on V. Let λ_1 eigenvalues of T. We proceed by contradiction. If T is diagonalizable, then by Theorem 8.11 V has an eigenbasis of consisting of eigenvectors of V. Let $E = (u, v, w)$ be this eigenbasis. Since each of u, v, w is an eigenvector and T has only one eigenvalues, each of these eigenvectors corresponds to λ. Therefore $u, v, w \in null\ (T - I\lambda)$ Since $\dim null\ (T - I\lambda_1) = 1$, then the list (u, v, w) is linearly dependent. This contradicts that E is a basis. Therefore T is not diagonalizable.

---

9 Inner Product Spaces Part I (LADR 6)

Learning Incomes.

- *Recall identities related to dot product from MATH164*

Learning Outcomes.

- *Understand the motivation for the definition of dot product in* $\mathbf{C}^n$
- *Be able to compute an inner product in* $\mathbf{C}^n$*.*
- *Understand the parts of the definition of inner product.*
- *Understand the proofs of various properties of inner products and norms.*
- *Be able to determine if a pair of vectors are orthogonal in a given inner product space*
- *Be able to compute the projection of a vector onto another in a given inner product space.*

Newly Defined Terms and Notation.

- *complex conjugate, magnitude, inner product,* $\langle u, v\rangle$*, inner product space, norm,* $\|v\|$

Throughout Modules 4-7 we studied linear maps from one vector space to another. For finite-dimensional vector spaces U and V, a bijective linear map $\phi : U \to V$ is an isomorphism. In other words, U and V are the same vector space, but with possibly different labels. For the case $V = \mathbf{R}^n$, Theorem 6.4, tells us that every real vector space of dimension n is isomorphic to $\mathbf{R}^n$.

For a pair of vectors $u, v \in \mathbf{R}^2$, recall the following formula

$$u \cdot v = \|u\|\|v\| \cos\theta$$

Here, $\|u\|$ and $\|v\|$ respectively denote the length of the vectors u and v, the angle θ is the angle between u and v and $u \cdot v$ is the dot product of u and v. Notice that these ideas; length, angle and dot product; do not appear as part of the definition of vector space. Therefore $\mathbf{R}^2$ has more structure than just the structure we get from thinking about $\mathbf{R}^2$ as a vector space. That is to say, we can do more with vectors in $\mathbf{R}^2$ than just add and scalar multiply.

Our ideas for angle and length in $\mathbf{R}^2$ extend easily to $\mathbf{R}^3$. As in $\mathbf{R}^2$ we can draw pictures and physically measure angles to confirm that the formula above holds when $u, v \in \mathbf{R}^3$. However, when we consider these concepts for $\mathbf{R}^n$ for $n \geq 4$ we use the equation above to define what angles mean in $\mathbf{R}^n$. By defining the dot product and length of a vector in $\mathbf{R}^n$ and then asserting the above formula to be true, we can make sense of the meaning of the idea of angle in $\mathbf{R}^n$ even without an easy to visualize physical interpretation. If we are not

too unsettled by thinking about angles in, say, $\mathbf{R}^7$, then perhaps we can be satisfied that we can think about angles in any vector space.

Looking at our formula above, to talk about angles in a vector space we need to be able to talk about dot products and lengths in a vector space. Recall that in $\mathbf{R}^n$ we have $\|v\| = \sqrt{v \cdot v}$. In $\mathbf{R}^n$ length and dot product are related. And so to make sense of the meaning of length in an arbitrary vector space, we first need to know about the dot product. As the definition of vector space does not mention the existence of a dot product, once we equip our vector space with a notion of dot product it is no longer just a vector space. We call these mathematical structures *inner product spaces.*

Returning back to our formula above, rearranging yields

$$\cos\theta = \frac{u \cdot v}{\|u\|\|v\|}$$

Since $\cos(\theta) \in [-1, 1]$, for this definition of angle to extend to other vector spaces, we must have $\frac{u \cdot v}{\|u\|\|v\|} \in [-1, 1]$. In other words we require $|u \cdot v| \leq \|u\|\|v\|$.

Our work over the next two modules proceeds as follows. First we make meaning of the notation $u \cdot v$ for arbitrary vector spaces. We then use this new operation in a vector space to make meaning of the notation $\|u\|$ and $\|v\|$. Finally, with the help of some of our intuition from Euclidean geometry, we prove $|u \cdot v| \leq \|u\|\|v\|$. As these concepts will be meaningful in both real and complex vector spaces, we first introduce/review some further facts about complex numbers.

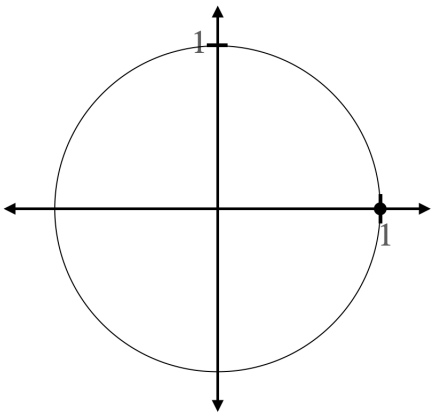
Aside. *If you have taken MATH163, you may recall that we can use relation to impose structure on a set. Here we doing something similar. The definition of vector space imposes an additive and scalar multiplicative structure on to a set. By adding a notion of dot product we are adding additional structure to the set.*

With this week being abbreviated due the public holiday on Friday, these notes are shorter than normal. Despite the page count, there is less material here than in previous modules. Depending on your mathematics background, Section 9.1 may be review. And throughout 9.1 and 9.2 there are many figures, which serve to inflate the page count.

Depending on your personal mathematics history, the material that follows may fall anywhere between being completely new and a complete review. In you are comfortable with the complex plane and magnitudes and conjugates of complex numbers, then feel free to fully skip the material in this section. If you do choose to skip this section, do recognize that as this is part of the material in MATH266, it is fair for me to ask about any of the material in this section as part of an assessment.

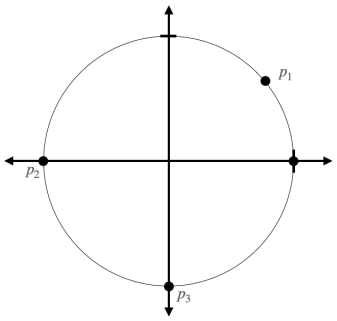
9.1 A Geometric Interpretation of Complex Numbers

Consider a circle of radius 1 centred at the origin of the xy plane:



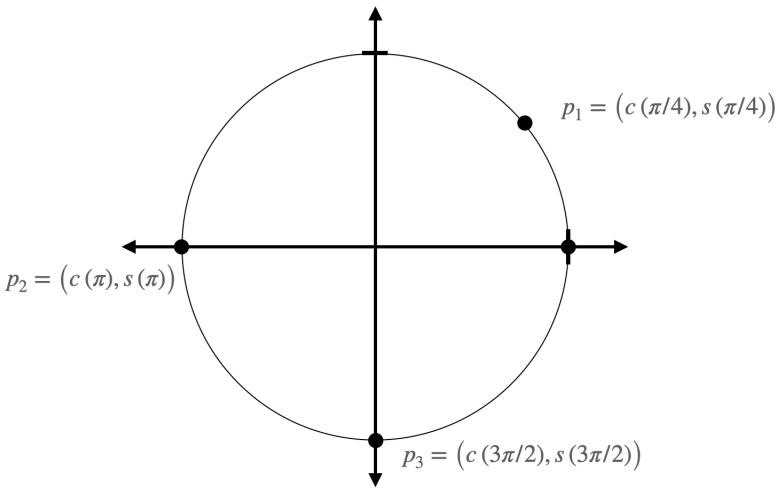
Imagine travelling around the circle anti-clockwise starting from the the indicated point, $(1, 0)$. As you travel around the circle you stop periodically and record how far you have travelled. Since the circle has radius 1, you calculate its circumference to be 2π. And so, for example, when you have travelled one-eighth of the way around the circle you have travelled a total distance of $\frac{1}{8}(2\pi) = \frac{\pi}{4}$. The table below gives the distances d you have travelled to reach the marked points

p	d
p_1	$\frac{\pi}{4}$
p_2	π
p_3	$\frac{3\pi}{2}$



As you travel around the circle you realize that you can associate the coordinates of points on the circle with distance you must travel to get there. Let $c(d)$ denote the the x-coordinate

of the point we reach after travelling distance d. And let $s(d)$ denote the y-coordinate of the point we reach after travelling distance d. For example

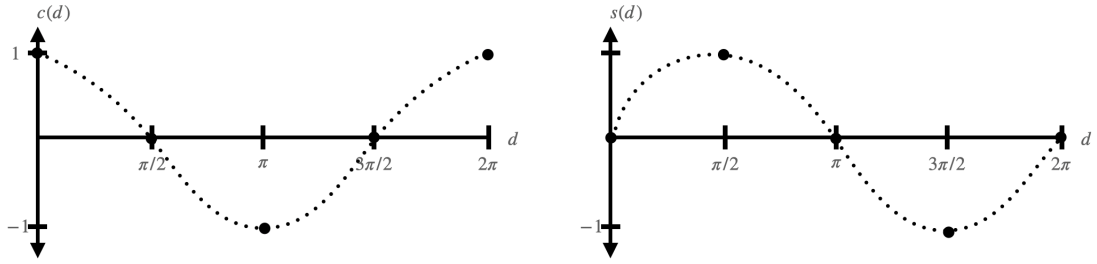


Let us take a moment to investigate these two functions c and s. Computing some values we find

d	$c(d)$	$s(d)$
0	1	0
$\pi/2$	0	1
π	-1	0
$3\pi/2$	0	-1
2π	1	0

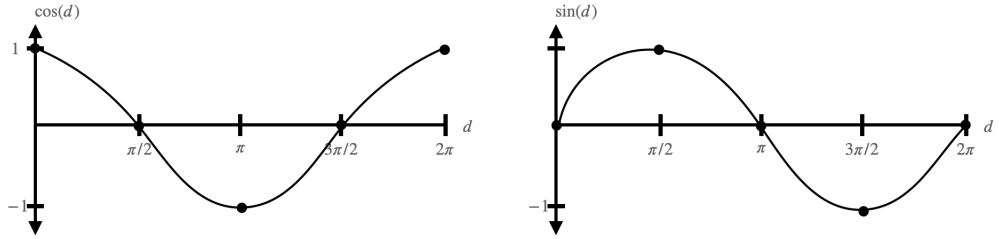
For example, when we have travelled a distance of π anti-clockwise around the circle starting at $(1, 0)$ we have reached $(-1, 0)$.

Eye-balling some curves based on our calculations we have

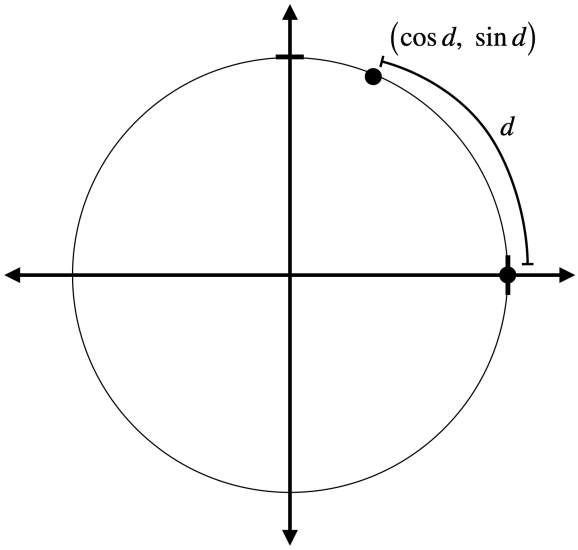


These functions look awfully familiar. In fact, they are exactly the functions cosine and sine

in the domain $[0, 2\pi]$.



Definition 9.1. *For $d \in [0, 2\pi]$, let $\cos d$ denote the x-coordinate of the point on unit circle reached by travelling distance d anti-clockwise around the circle starting at $(1, 0)$. For $d \in [0, 2\pi]$, let $\sin d$ denote the y-coordinate of the point on unit circle reached by travelling distance d anti-clockwise around the circle starting at $(1, 0)$.*



Aside 9.2. *Yep, sine and cosine have a definition that has nothing to do with right-angled triangles. One may also define* $\tan d$ *using this circle without resorting to its definition as a ratio.*

The word sine *is derived from a Sanskrit word. Its path to the English language comes by way of transliteration of a word into Arabic followed by a subsequent translation of a similar sounding Arabic word to English. The word sine appears in English texts from over 400 years ago.*

The study of trigonometric functions is ancient and universal. For example, their study

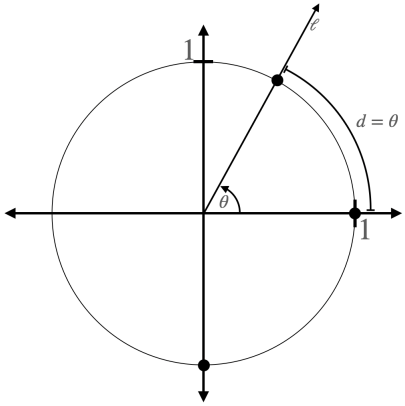
appears in work in Greece in the 1st century, in India in the 4th century and in work during the early Golden Age of Islam in the 9th century.

"*But wait!*", we cry. "*Aren't cosine and sine related somehow to angles? How can we define them in terms of the distance around the circle from the point* $(1, 0)$*?*

For any point (x, y) on the unit circle, consider the line ℓ from the origin through (x, y). Let θ denote the angle that this line makes with the positive x-axis.

Just as we can associate each point on the circle with a distance in the range $[0, 2\pi)$, we can associate this angle θ with this same distance. In fact, we can use this distance as the label for our angle.

Definition 9.3. *Let ℓ be a line with one end at the origin. We denote the angle formed between ℓ and the positive x-axis as θ where $d = \theta$ is the distance around the unit circle from $(1, 0)$ to the point of intersection between ℓ and the unit circle. We call θ the* ***radian angle****.*



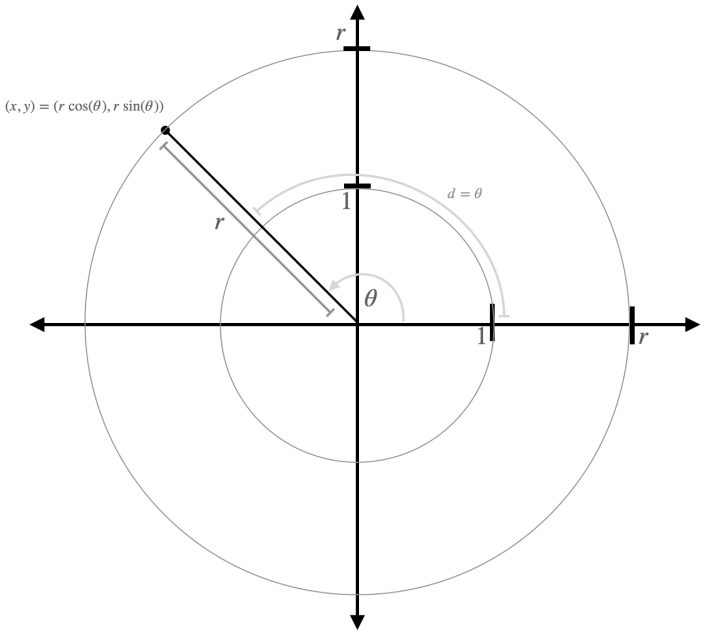
Aside. *Think of a radian angle as multiplying a fraction by 2π. This fraction tells us the distance we must travel around the unit circle to reach the corresponding point. This is a great example of how the mindless "rule" of always reducing fractions their lowest terms can lead to confusion. We write $\pi/4$ to talk about the angle formed walking $1/8$ of the way around the circle. Surely it is more clear to express this angle as $\frac{1}{8}(2\pi)$ rather than $\pi/4$?*

Just as we denote any point on the unit circle using cosine, sine and radian angle, we may in fact denote any point in $\mathbf{R}^2$ this way.

Theorem 9.4. *For $r > 0$, let C_r be the circle of radius r centred at the origin. And let ℓ be*

a line through the origin so that the radian angle between the positive x-axis and ℓ is θ. The point of intersection of ℓ and C_r is given by $(r\cos(\theta), r\sin(\theta))$.

One may obtain a proof of this theorem by reasoning about similar triangles in the image below and in various related images where the point of intersection lies in different quadrants.

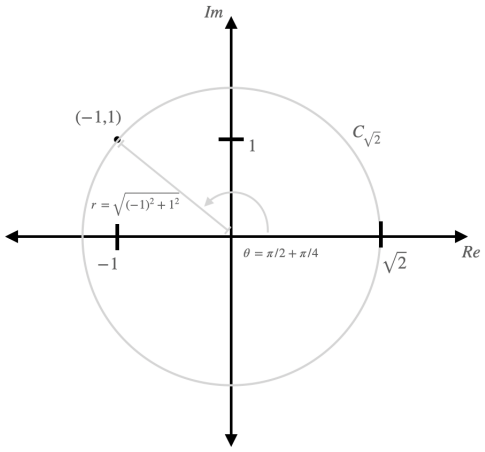


For any point $(x, y) \in \mathbf{R}^2$ there exists a circle centred at the origin that passes through (x, y). And so we have:

Corollary 9.5. *For every $(x, y) \in \mathbf{R}^2$, there exists $r \geq 0$ and $\theta \in [0, 2\pi)$ so that*

$$(x, y) = (r\cos(\theta), r\sin(\theta)).$$

For example, we can express the point $(-1, 1)$ as $(\sqrt{2}\cos(3\pi/4), \sqrt{2}\sin(3\pi/4))$.



Aside 9.6. *We have managed to define radian angles without having to invoke degrees. Using* $90°$ *to denote a quarter rotation relies on us having to* remember*, without any intuition, that a full rotation corresponds to* $360°$*. We have chosen as a society to define the full rotation as* $360°$*. That there are* $360°$ *in a full rotation is only* correct *insofar that we have all decided to agree on this convention. The choice of* 360 *is in fact a good one – there are lots of divisors of* 360*, which makes it easy to express exactly lots of different angles. This is the same reason that* 60 *is a good choice for number of minutes in an hour. Choosing to define there to be* 24 *hours in a day is, however, patently absurd. We can blame ancient Egypt for that.*

On the other hand we can reason *that* $\pi/8$ *denotes a quarter rotation by remembering only that the ratio between the diameter of a circle and its circumference is denoted as* π*. The only convention that we are required to agree upon to reach a* correct *answer is that we use the symbol* π *to denote this ratio. (We don't even need to know the value of* π*!)*

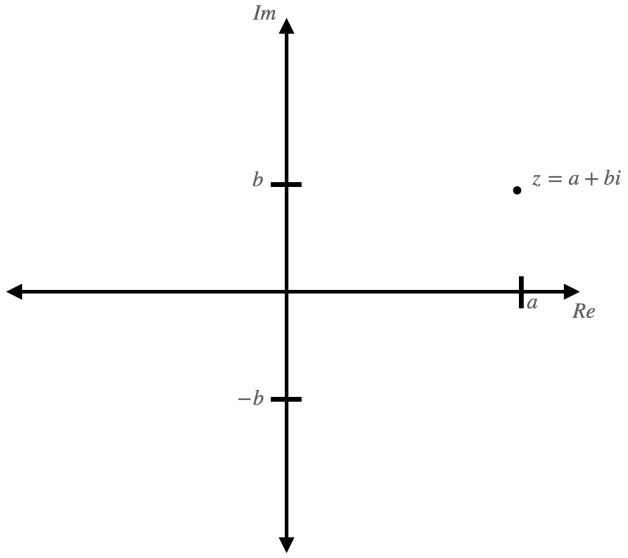
What does any of this have to do with complex numbers? Recall the definition of the set of complex numbers:

$$\mathbf{C} = \{a + bi \mid a, b \in \mathbf{R}\}$$

Just as we can situate elements of $\mathbf{R}$ geometrically on the real line, so too can we find a geometric interpretation for elements of $\mathbf{C}$.

The **Complex Plane** consists of two axes, the real axis, which is horizontal, and the imaginary axis, which is vertical. Each axis is indexed by $\mathbf{R}$ according to the usual order of elements on this set. The point of intersection of the vertical line through a on the real axis

and the horizontal line though b on the imaginary axis is denoted as $a + bi$.



Just as the magnitude of a vector in $\mathbf{R}^2$ refers to the distance from the origin in $\mathbf{R}^2$, so too do we use this language to talk about the distance from $0 + 0i$ in $\mathbf{C}$.

Definition 9.7. *Let $z = a + bi$ be a complex number. The magnitude of z is the length of the line from $0 + 0i$ to $a + bi$ in the complex plane. We denote this quantity as $|z|$. This quantity is given by the formula*

$$|z| = \sqrt{a^2 + b^2}$$

For example, when $z = 3 + 4i$ we have.

$$|z| = \sqrt{3^2 + 4^2} = 5$$

By seemingly pure coincidence, one can express the magnitude of $z = a + bi$ in a slightly different way. Notice

$$(a + bi)(a - bi) = (a^2 + b^2) + (ab - ab) = a^2 + b^2.$$

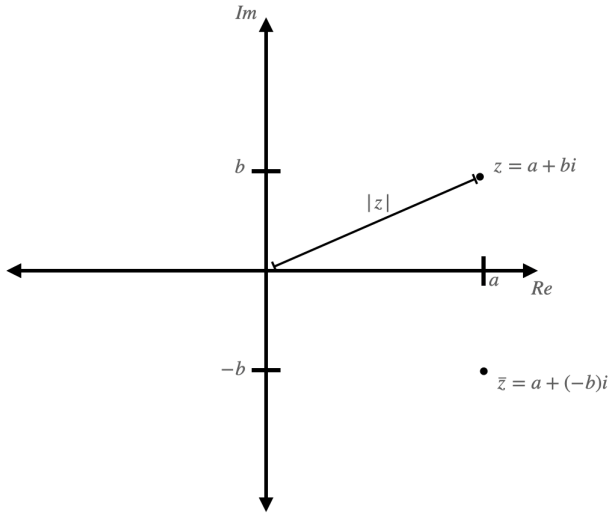
Therefore

$$|z| = \sqrt{(a + bi)(a - bi)}$$

For reasons that will become clear in 9.2, expressing the magnitude of a complex number in this way is very convenient. For this end, we define the following.

Definition 9.8. *Let $z = a + bi$ be a complex number. The complex conjugate of z is the complex number $a - bi$. We denote this quantity as $\overline{z}$. That is, we define*

$$\overline{z} = a - bi$$



Just as we have seen throughout many places in this course, complex conjugates have a scent of linearity about them. That is for all $\lambda, z_1, z_2 \in \mathbf{C}$ we have

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \qquad \text{and} \qquad \overline{\lambda z_1} = \overline{\lambda}\,\overline{z_1}$$

We note this only as a curiousity for now, but we return to this fact in the following section.

Putting together magnitude and conjugate we arrive at the following, perhaps familiar, identity. Except now, we have the knowledge and experience to prove this fact, rather than memorize a meaningless equation.

Theorem 9.9. *For every $z \in \mathbf{C}$ we have*

$$|z|^2 = z\,\overline{z}$$

Proof. Let $z = a + bi$ be a complex number. By definition, $\overline{z} = a - bi$. We compute

$$z\,\overline{z} = (a + bi) \cdot (a - bi) = a^2 + b^2 + 0i = a^2 + b^2$$

Recall $|z| = \sqrt{a^2 + b^2}$. Therefore $z\,\overline{z} = |z|^2$. □

Aside 9.10. *Since z is a complex number, it has a multiplicative inverse. Multiplying the equation $|z|^2 = z\,\overline{z}$ on both sides by this inverse and then simplifying gives an unintuitive expression for the multiplicative inverse: $z^{-1} = \frac{1}{|z|^2}\overline{z}$.*

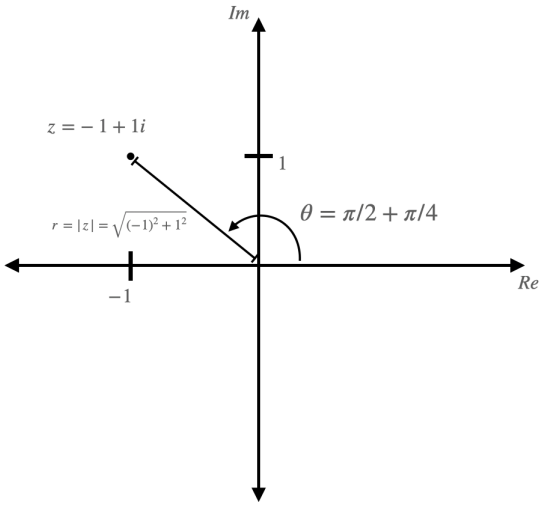
Above, we spent some time reviewing some trigonometric concepts in $\mathbf{R}^2$. Given that the Complex Plane seems to only differ from the Cartesian Plane in the names of the axes and how we label the points, our work in defining sine, cosine and radians extends to the complex plane.

Since every $(a, b) \in \mathbf{R}^2$ can be expressed as $(r\cos(\theta), r\sin(\theta))$ so too can we express a complex number in this way.

Theorem 9.11. *For every $z \in \mathbf{C}$, there exists $r \geq 0$ and $\theta \in [0, 2\pi)$ so that*

$$z = r\cos(\theta) + i\,(r\sin(\theta))$$

By way of example, let $z = -1 + 1i$. We find r and θ so that $z = r\cos(\theta) + i\,(r\sin(\theta))$. With a quick sketch we see $r = |z| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$ and $\theta = \pi/2 + \pi/4 = 3\pi/4$.



And so

$$z = \sqrt{2}\cos(3\pi/4) + i\left(\sqrt{2}\sin(3\pi/4)\right)$$

Aside 9.12. *Notice that we have written $i\,(r\sin(\theta))$ rather than writing $(r\sin(\theta))\,i$ for the imaginary part of z in the statement of this theorem. The only reason to do this is convention. Writing $(r\sin(\theta))\,i$ in the statement of this theorem would carry the same meaning.*

Theorem 9.11 tells us that for each complex number $z = a + bi$ there is a pair (r, θ) so that $z = r\cos(\theta) + i\,(r\sin(\theta))$. And so instead of denoting a complex number by reference to its position in the complex plane, which is what we do when we write $z = a + bi$, we could instead denote a complex number with the pair (r, θ). Such a coordinate system is called *Polar Coordinates.* For example, we could refer to $z = -1 + 1i$ with the pair $(\sqrt{2}, 3\pi/4)$. If you take further courses in calculus beyond the 100-level, such coordinates will likely make an appearance.

9.1 Test Your Understanding

1. Let $z = 6 + 8i$
   (a) Compute $|z|^2$
   (b) Compute $\frac{1}{|z|^2}\overline{z}$
   (c) Show $z^{-1} = \frac{1}{|z|^2}\overline{z}$
2. Let $z = -4 - 4i$. Find r and θ so that $z = r\cos\theta + ir\sin\theta$

9.1 Test Your Understanding Solution

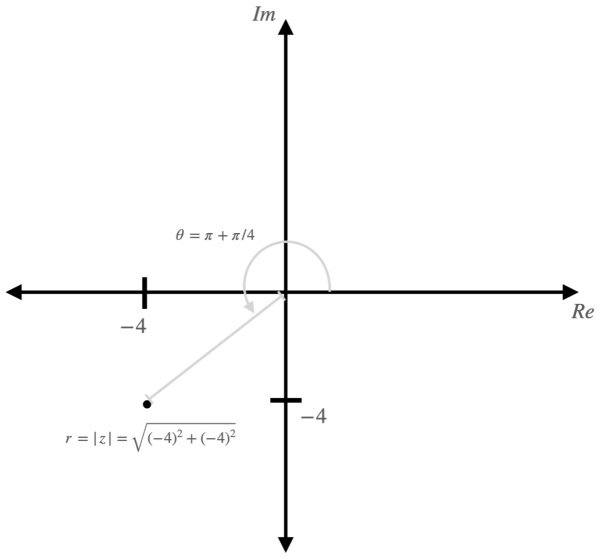
1. (a) We compute: $|6+8i|^2 = 6^2+8^2 = 100$

   (b) We compute:
   $$\frac{1}{|z|^2}\overline{z} = \frac{1}{100}(6-8i) = \frac{3}{50} - \frac{2}{25}i$$

   (c) To show $z^{-1} = \frac{1}{|z|^2}\overline{z}$ we show $z\left(\frac{1}{|z|^2}\overline{z}\right) = 1$. We compute
   $$\begin{aligned} z\left(\frac{1}{|z|^2}\overline{z}\right) &= \frac{1}{|z|^2}z\overline{z} \\ &= \frac{1}{100}(6+8i)(6-8i) \\ &= \frac{1}{100}(6^2+8^2) \\ &= \frac{1}{100}(100) \\ &= 1 \end{aligned}$$

   Therefore $z^{-1} = \frac{1}{|z|^2}\overline{z}$.

2. We sketch:



And so we see $r = |z| = \sqrt{32} = 4\sqrt{2}$ and $\theta = \pi + \pi/4 = 5\pi/4$. Therefore

$z = 4\sqrt{2}\cos(5\pi/4) + i4\sqrt{2}\sin(5\pi/4)$

9.2 The Dot Product for Vector Spaces: Real and Complex Inner Product Spaces

Just as our formula for the length of a vector in $\mathbf{R}^2$ generalizes to give us a formula for length of a vector in $\mathbf{R}^n$, so too does our formula for length in $\mathbf{C}$ give rise to a formula for length of a vector in $\mathbf{C}^n$. Conveniently, we can define these both at the same time.

Definition 9.13. *For $v \in \mathbf{F}^n$, the **norm of** v, denoted $\|v\|$, is given by*

$$\|v\| = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}$$

Aside. *For the life of me, I have no idea why the word* norm *is used to talk about distances in mathematics. It is a good idea that we use a word other than* distance*, as our intuition for distance is very Euclidean. If you proceed further in higher mathematics courses, you will see lots of different types of norms. For example, the* Taxi-Cab *norm measures distances in $\mathbf{Z}^2$ by analogy to the distance one would have to travel in a city laid out as a grid. Another common example of norm, from computer science, is Hamming Weight.*

As this definition permits $v \in \mathbf{F}^n$, this definition is valid for elements of $\mathbf{C}^n$. For example in $\mathbf{C}^2$ we can compute

$$\begin{aligned}\|(1+i, 0+3i))\| &= \sqrt{|1+i|^2 + |3i|^2} \\ &= \sqrt{(1+i)(1-i) + (3i)(-3i)} \\ &= \sqrt{2+9} \\ &= \sqrt{11}\end{aligned}$$

Notice here that despite the fact that our vector in $\mathbf{C}^2$, the norm of the vector is in $\mathbf{R}$ and is non-negative.

From MATH164, for $v \in \mathbf{R}^n$ we recall

$$v \cdot v = \|v\|^2$$

We define the dot product in $\mathbf{C}^n$ so that this equality remains true. From our work in the previous section and the definition above, for $v \in \mathbf{C}^n$ we have

$$\|v\|^2 = |v_1|^2 + |v_2|^2 + \cdots + |v_n|^2 = v_1\overline{v_1} + v_2\overline{v_2} + \cdots + v_n\overline{v_n}$$

And so to have $v \cdot v = \|v\|^2$ we require the following definition for dot product in $\mathbf{C}^n$.

Definition 9.14. *Let $u, v \in \mathbf{C}^n$, the dot product of u and v, denoted $u \cdot v$ is given by*

$$u \cdot v = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n}$$

For example

$$\begin{aligned}(1+i,0+3i)\cdot(3+2i,-1+4i) &= (1+i)(3-2i)+(3i)(-1-4i)\\ &= (5+i)+(12-3i)\\ &= 17-2i\end{aligned}$$

Here are vectors are in $\mathbf{C}^2$ and their dot product is an element of $\mathbf{C}$. Notice here that unlike the dot product in $\mathbf{R}^n$, we may have $u\cdot v \neq v\cdot u$. In our example above when we reverse the order of the products we get

$$(3+2i,-1+4i)\cdot(1+i,0+3i) = 17+2i = \overline{17-2i} = \overline{(1+i,0+3i)\cdot(3+2i,-1+4i)}$$

This property holds in general. That is, for all $u, v \in \mathbf{C}^n$ we have $u\cdot v = \overline{v\cdot u}$.

As every finite-dimensional vector space is isomorphic to either $\mathbf{C}^n$ or $\mathbf{R}^n$, perhaps then we can define a dot product, and subsequently length (and maybe angle) for other vector spaces. To do so, we make a slight change in notation. As the notation $x\cdot y$ reminds us of multiplication of numbers, we denote our dot product on vector spaces with $\langle x, y\rangle$ and call it an *inner product*. For example, in $\mathbf{R}^3$ we have

$$\langle u, v\rangle = u_1v_1 + u_2v_2 + u_3v_3$$

And in $\mathbf{C}^3$ we have

$$\langle u, v\rangle = u_1\overline{v_1} + u_2\overline{v_2} + u_3\overline{v_3}$$

Using our definition of dot product from MATH164, we are able to prove a number of facts about the dot product that are not obvious from the definition. For example, we can prove $\langle v, v\rangle = 0$ if and only if $v = 0$. We can also prove $\langle v, v\rangle \geq 0$ for all $v \in V$.

In a vector space by definition we have a notion of distributivity. It turns out that our usual dot product also admits a notions of distributivity.

For example, in $\mathbf{R}^2$ we have

$$\langle u+v, w\rangle = (u_1+v_1)w_1 + (u_2+v_2)w_1 = (u_1w_1)+(v_1w_1)+(u_2w_2)+(v_2w_2) = \langle u, w\rangle + \langle v, w\rangle$$

We describe this property using the word **additive**. The inner product in $\mathbf{R}^2$ is additive in the first entry. We also get an distributivity looking property with scalar multiples

$$\langle \lambda u, v\rangle = \lambda u_1v_1 + \lambda u_2v_2 = \lambda(u_1v_1 + u_2v_2) = \lambda\langle u, v\rangle$$

We define inner product for an arbitrary vector space so that it emulates these properties.

Definition 9.15. *Let V be a vector space. An **inner product** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ so that*

1. $\langle v, v\rangle \in \mathbf{R}$ *and* $\langle v, v\rangle \geq 0$ *for all* $v \in V$.
2. $\langle v, v\rangle = 0$ *if and only if* $v = 0$.
3. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ *for all* $u, v, w \in V$
4. $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ *for all* $u, v \in V$ *and all* $\lambda \in \mathbf{F}$
5. $\langle u, v\rangle = \overline{\langle v, u\rangle}$ *for all* $u, v \in V$

So that our definition makes sense for both real and complex vector spaces, recall that for $r \in \mathbf{R}$ we can consider r as a complex number by writing it as $r + 0i$. And so, we have $\overline{r} = r - 0i = r$. With this in mind, when V is a real vector space, the last requirement in the definition of inner product requires that we have $\langle u, v\rangle = \langle v, u\rangle$.

Aside. *The usual example of an inner product on a vector space other than* $\mathbf{F}^n$ *is as follows. Consider* $\mathcal{C}_{[-1,1]}(\mathbf{R})$*, the real vector space of continuous functions on the interval* $[-1, 1]$*. For* $f, g \in \mathcal{C}_{[-1,1]}(\mathbf{R})$*, let*

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\, dx$$

One can verify, though perhaps not easily, that $\langle f, g\rangle$ *satisfies the five properties above.*

Just like we saw when we defined vector spaces, these properties imply further properties of inner products that are not explicitly stated in the definition. For example, 3. tells us inner products are necessarily additive in the first entry. We can prove that inner products are also additive in the second entry. That is, we show

$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle \text{ for all } u, v, w \in V$$

Theorem 9.16. *Let* V *be an inner product space. For all* $u, v, w \in V$ *we have* $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$

Proof. Let V be vector space with an inner product and let $u, v, w \in V$. By the definition of inner product, we have

$$\begin{aligned}
\langle u, v + w\rangle &= \overline{\langle v + w, u\rangle} \\
&= \overline{\langle v, u\rangle + \langle w, v\rangle} \\
&= \overline{\langle v, u\rangle} + \overline{\langle w, v\rangle} \\
&= \langle u, v\rangle + \langle u, w\rangle
\end{aligned}$$

□

However, 4. changes slightly when we change from the first entry to the second entry.

Theorem 9.17. *Let V be an inner product space. For all $u, v \in V$ and all $\lambda \in \mathbf{F}$ we have*

$$\langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle$$

Proof. Let V be vector space with an inner product, let $u, v \in V$ and let $\lambda \in \mathbf{F}$. By the definition of inner product and we have

$$\begin{aligned} \langle u, \lambda v \rangle &= \overline{\langle \lambda v, u \rangle} \\ &= \overline{\lambda \langle v, u \rangle} \\ &= \overline{\lambda} \overline{\langle v, u \rangle} \\ &= \overline{\lambda} \langle u, v \rangle \end{aligned}$$

□

Our definition of vector space requires that we have a set of objects together with a notion of addition and scalar multiplication for those objects that satisfy some set of properties. We define an *inner product space* in the same way. It is a vector space for which there is an inner product for the elements.

Definition 9.18. *Let V be a vector space. We call V an **inner product space** when we have defined an inner product on V.*

Aside. *Other than $\mathbf{R}^n$ and $\mathbf{C}^n$, one may find another example of an inner product by looking at random variables. The set of random variables of an experiment forms an vector space. Defining $\langle X, Y \rangle = E(XY)$ turns this vector space into an inner product space.*

Let V be an inner product space. And let $v \in V$. By analogy to the formula for length in $\mathbf{R}^n$, we arrive at the following definition of norm in an inner product space.

Definition 9.19. *Let V be an inner product space and let $v \in V$. The **norm of** v, denoted $\|v\|$, is given by*

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

In $\mathbf{R}^n$ multiplying a vector by a scalar has the effect of multiplying the length of the vector by the scalar. If this scalar is negative, then the resulting length comes from multiplying the absolute value of the scalar by the length of the vector. That is

$$\|\lambda v\| = |\lambda| \, \|v\|$$

This same fact holds true in any inner product space. In this case, however, the scalar, λ, may be a complex number.

Theorem 9.20. *Let V be an inner product space. For all $v \in V$ we have $\|\lambda v\| = |\lambda| \, \|v\|$*

Proof. Let V be an inner product space and let $v \in V$. By definition of norm and the properties of inner products, we have

$$\begin{aligned} \|\lambda v\|^2 &= \langle \lambda v, \lambda v \rangle \\ &= \lambda \langle u, \lambda v \rangle \\ &= \lambda \overline{\lambda} \langle v, v \rangle \\ &= |\lambda|^2 \langle v, v \rangle \\ &= |\lambda|^2 \|v\|^2 \end{aligned}$$

Taking square roots on both sides yields the required result. □

With these initial pieces in place, we are now in a position to return to our original motivation for our study; angles in vector spaces. Restated using inner products, consider the equality

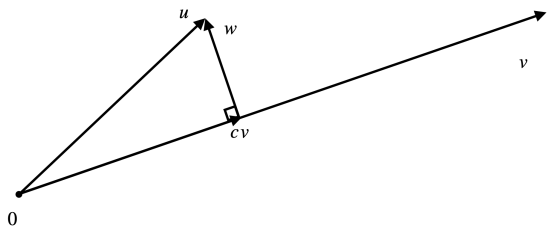
$$\langle u, v \rangle = \|u\| \|v\| \cos \theta$$

Recall from MATH164, that in $\mathbf{R}^n$ the quantity $\frac{\langle u,v \rangle}{\|u\|\|v\|}$ gives us information about the angle between u and v. When this quantity is 1 or -1, u is a scalar multiple of v. And so the angle between them is 0 or π. On the other hand, when this quantity is 0, the angle between u and v is $\pi/2$. In other words, u and v are orthogonal.

With this in motivation in mind, we define orthogonality in an inner product space analogously.

Definition 9.21. *Let V be an inner product space and let $u, v \in V$. We say u **and** v **are orthogonal** when $\langle u, v \rangle = 0$*

Now that we have defined orthogonality in vector spaces, topics like orthogonal projection all of a sudden have meaning. In $\mathbf{R}^2$, projecting a vector u on to v results in a scalar multiple of v, say cv, so that v is orthogonal to $u - cv$. Letting $w = u - cv$ yields the following picture.



We find c so that v is orthogonal to $u - cv$. In other words, we find c so that

$$0 = \langle u - cv, v \rangle$$

By the definition of inner product we have

$$\begin{aligned} 0 &= \langle u - cv, v \rangle \\ &= \langle u, v \rangle - \langle cv, v \rangle \\ &= \langle u, v \rangle - c\langle v, v \rangle \\ &= \langle u, v \rangle - c\|v\|^2 \end{aligned}$$

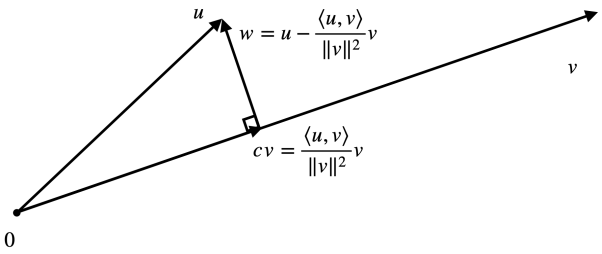
Rearranging we find

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

and

$$\begin{aligned} w &= u - cv \\ &= u - \frac{\langle u, v \rangle}{\|v\|^2} v \end{aligned}$$

Returning to our figure, we find:



Since we can make sense of scalar multiples, vector addition and orthogonality in any inner product space, so too can we consider projections in any inner product space. Though we were guided by a picture in $\mathbf{R}^2$ all of our work above is the same in any inner product space. We appealed only to properties of an inner product, never specifically to the usual dot product in $\mathbf{R}^2$. In other words, if V is an inner product space and $u, v \in V$ with $v \neq 0$, then the projection of u onto v is a scalar multiple of v, say cv, so that v and $u - cv$ are orthogonal. From our work above, we have $c = \frac{\langle u,v \rangle}{\|v\|^2}$ and $u = \frac{\langle u,v \rangle}{\|v\|^2} v + w$. And so, the projection of u onto v is given by $\frac{\langle u,v \rangle}{\|v\|^2} v$.

Returning back to our thoughts from the introductory remarks in 9.0, to make sense of angles in an inner product space, we were interested in proving $|\langle u, v \rangle| \leq \|u\|\|v\|$. It turns out that the work we have done here will be essential in this effort. We return to this in Module 10.

Aside. *In our study of vector spaces, we have taken our scalars to be elements of either* $\mathbf{R}$ *or* $\mathbf{C}$*. These are not the only choices one can make for scalars when thinking about vector spaces. Let* $\mathbf{Z}_2 = \{0, 1\}$*. Define addition and multiplication with the following tables.*

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

Taking binary strings of a fixed length as vectors (e.g., nibbles, bytes, etc...), one can form a vector space with scalars in $\mathbf{Z}_2$*. This vector space is the underlying mathematical object for digital transmission of data. With wifi network and cell phone networks propagating digital transmissions all around us, it is not an overstatement to say that we inhabit this vector space in our daily lives. One can define an inner product on this vector space so that the resulting norm is the* ***Hamming Weight*** *of a bit string. The study of this vector space makes possible the transmission of data in noisy channels. For more information, do an internet search for* `Hamming Codes`*.*

9.2 Test Your Understanding

1. Is matrix multiplication an inner product for the vector space of 2×2 matrices? That is, if we let $\langle A, B \rangle = AB$ does $\langle A, B \rangle$ satisfy the definition of inner product?
2. Let $u = (1 + i, 2 + i)$ and $v = (1 + 2i, 3 + 0i)$
   (a) Compute $\|v\|^2$.
   (b) Determine if u and v are orthogonal.
   (c) Compute the projection of u onto v.
3. Let V be an inner product space and let u and v be elements of V. Using the definition of norm and the fact that inner products are additive in both the first and second entry, show $\langle u + v, u + v \rangle = \|u\|^2 + \|v\|^2 + \langle u, v \rangle + \overline{\langle u, v \rangle}$

9.2 Test Your Understanding Solution

1. No. If $A, B \in \mathcal{M}_{n\times n}(\mathbf{F})$, then $AB \in \mathcal{M}_{n\times n}(\mathbf{F})$. Therefore $\langle A, B\rangle \notin \mathbf{F}$. Therefore $\langle A, B\rangle$ is not an inner product.

2. Let $u = (1+i, 2+i)$ and $v = (1+2i, 3+0i)$.

   (a) By definition we have $\|v\| = \langle v, v\rangle$. Therefore

   $$\begin{aligned}\|v\| &= \langle (1+2i, 3+0i), (1+2i, 3+0i)\rangle \\ &= (1+2i)(\overline{1+2i}) + (3+0i)(\overline{3+0i}) \\ &= (1+2i)(1-2i) + (3+0i)(3-0i) \\ &= 14\end{aligned}$$

   (b) To determine if u and v are orthogonal, we $\langle u, v\rangle$.

   $$\begin{aligned}\langle u, v\rangle &= \langle (1+i, 2+i), (1+2i, 3+0i)\rangle \\ &= (1+i)(\overline{1+2i}) + (2+i)(\overline{3+0i}) \\ &= (1+i)(1-2i) + (2+i)(3-0i) \\ &= 9+2i\end{aligned}$$

   Since $\langle u, v\rangle \neq 0$, u and v are not orthogonal.

   (c) The projection of u onto v is given by $\frac{\langle u,v\rangle}{\|v\|^2}v$. We compute

   $$\begin{aligned}\frac{\langle u, v\rangle}{\|v\|^2}v &= \frac{9+2i}{14}(1+2i, 3+0i) \\ &= \frac{1}{14}(9+2i)(1+2i, 3+0i) \\ &= \left(\frac{9}{14} + \frac{2}{14}i\right)(1+2i, 3+0i) \\ &= \left(\frac{5}{14} + \frac{10}{7}i, \frac{27}{14} + \frac{6}{14}i\right)\end{aligned}$$

3. Let V be an inner product space and let u and v be elements of V. Since inner products are additive in the first entry, we have

   $$\langle u+v, u+v\rangle = \langle u, u+v\rangle + \langle v, u+v\rangle$$

   Since inner products are additive in the second entry we have

   $$\langle u, u+v\rangle + \langle v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, v\rangle + \langle v, u\rangle$$

   By the definition of norm we have

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

By property 5 in the definition of inner product we have

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

Therefore $\langle u + v, u + v\rangle = \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \overline{\langle u, v\rangle}$

---

10 Inner Product Spaces Part II + Change of Basis (LADR 6,10)

Learning Incomes.

- *Recall definitions and notation for inner product spaces and complex numbers from Module 9*
- *Recall the Triangle Inequality and Pythagoras' Theorem from secondary school geometry*
- *Understand the relationship between eigenbases and diagonalization.*
- *Be able to construct the matrix of a linear operator with respect to a pair of bases.*

Learning Outcomes.

- *Understand the statement and proof of Pythagoras' Theorem for Inner Product Spaces*
- *Understand the statement and proof of the Triangle Inequality for Inner Product Spaces*
- *Understand the statement and proof of the Cauchy-Schwarz Inequality.*
- *Be able to construct a change of basis matrix for a finite-dimensional vector space*
- *Be able to change from the representation of the matrix of a vector with respect to one basis to the matrix of the vector with respect to a different basis.*

Newly Defined Terms and Notation.

- *Change of basis matrix*

10.1 The Cauchy-Schwarz Inequality

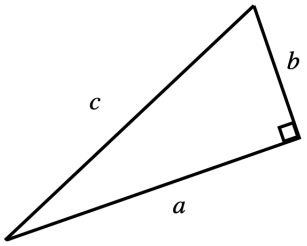
Let u and v be vectors in $\mathbf{R}^n$ and let θ be the angle between u and v. In $\mathbf{R}^n$ the inner product, norm and θ satisfy the following equality

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|\|v\|}$$

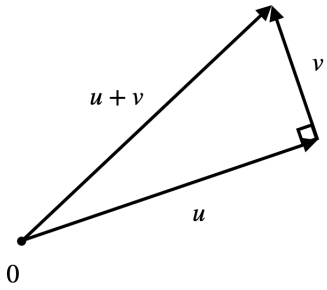
In $\mathbf{R}^2$ and $\mathbf{R}^3$ we can visually inspect angles and draw some diagrams to convince ourselves that this equality holds. However for $\mathbf{R}^n$ with $n \geq 4$ we take this equality as the definition of angle between a pair of vectors. We can use this formula to give meaning to angles in real vector spaces that don't have any obvious sense of geometry. However, to do so, we must be slightly careful.

Since $\cos(\theta) \in [-1, 1]$, for this definition of angle to extend to arbitrary real inner product spaces, we must have $\frac{\langle u,v\rangle}{\|u\|\|v\|} \in [-1, 1]$. In other words we require $|\langle u, v\rangle| \leq \|u\|\|v\|$. To prove $|\langle u, v\rangle| \leq \|u\|\|v\|$ in any inner product space, we first cast our minds back to a topic in geometry that we are well familiar with: Pythagoras' Theorem.

Theorem 10.1 (Pythagoras' Theorem in $\mathbf{R}^2$)**.** *If a, b, c are the lengths of a right-angled triangle and c is the length of the hypotenuse, then $a^2 + b^2 = c^2$*



If we situate the non-right angle of our triangle at the origin in $\mathbf{R}^2$ and label the vectors appropriately we have



Expressed using norms in $\mathbf{R}^2$, we can write Pythagoras' Theorem as $\|u\|^2 + \|v\|^2 = \|u+v\|^2$. Surprisingly, this statement is true for any pair of orthogonal vectors in an inner product space.

Theorem 10.2 (Pythagoras' Theorem for Inner Product Spaces)**.** *Let V be an inner product space and let $u, v \in V$. If u and v are orthogonal, then $\|u\|^2 + \|v\|^2 = \|u+v\|^2$.*

Proof. By definition of norm, we have $\|u+v\|^2 = \langle u+v, u+v\rangle$ Since inner products are additive in both the first and second entry, we have

$$\langle u+v, u+v\rangle = \langle u, u+v\rangle + \langle v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle$$

Since u and v are orthogonal, we have $\langle u, v\rangle = 0$ and $\langle v, u\rangle = 0$. Therefore

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

□

Using Pythagoras' Theorem for Inner Product Spaces and projections in inner product spaces, we prove $|\langle u, v\rangle| \leq \|u\|\|v\|$.

Aside. *This result is called the Cauchy-Schwarz Inequality, after Augustin-Louis Cauchy and Hermann Schwarz, both mathematicians working in Europe during the 17th century. In the early 1800s Cauchy provided a proof for an inequality related to sums and products. His student, Viktor Bunyakovsky generalized Cauchy's work for integrals. It wasn't until the late 1800s that Schwarz provided a general proof.*

This inequality comes up in a number of surprising places, from probability theory to calculus. It should strike us as astounding that this proof was cutting-edge mathematics less than 150 years ago and now it is taught *in a second year (!) mathematics course. Truly we stand on the shoulders of giants.*

Theorem 10.3 (The Cauchy-Schwarz Inequality)**.** *Let V be an inner product space. For every $u, v \in V$ we have $|\langle u, v\rangle| \leq \|u\|\|v\|$.*

Proof. Let V be an inner product space and let $u, v \in V$.

If one of u or v is zero, then $\langle u, v\rangle = 0$ and $\|u\|\|v\| = 0$. In this case the inequality holds.

Otherwise, let us assume that we have $u \neq 0$ and $v \neq 0$. Consider the projection of u onto v.

By our work in the previous section, the result of projecting u onto v is a vector cv so that $c = \frac{\langle u, v\rangle}{\|v\|^2}$. Let $w \in V$ so that $u = cv + w$. Since cv and w are orthogonal, by Pythagoras'

Theorem for Inner Product Spaces, the definition of inner product and Theorem 9.9, we have

$$\begin{aligned}\|u\|^2 &= \|cv\|^2 + \|w\|^2 \\ &= |c|^2\|v\|^2 + \|w\|^2\end{aligned}$$

Substituting $c = \frac{\langle u,v\rangle}{\|v\|^2}$ and simplifying we have

$$\begin{aligned}\|u\|^2 &= \left|\frac{\langle u,v\rangle}{\|v\|^2}\right|^2 \|v\|^2 + \|w\|^2 \\ &= \frac{|\langle u,v\rangle|^2}{\|v\|^4}\|v\|^2 + \|w\|^2 \\ &= \frac{|\langle u,v\rangle|^2}{\|v\|^2} + \|w\|^2\end{aligned}$$

Since $\|w\|^2 \geq 0$ we have

$$\|u\|^2 \geq \frac{|\langle u,v\rangle|^2}{\|v\|^2}$$

Multiplying both sides of the equality by $\|v\|^2$ and taking the square root of both sides yields $\|u\|\|v\| \geq |\langle u,v\rangle|$. □

Aside. *In $\mathbf{R}^2$, the Cauchy-Schwarz Inequality becomes $|u_1v_1 + u_2v_2| \leq \sqrt{u_1^2+u_2^2}\sqrt{v_1^2+v_2^2}$. Squaring both sides yields the following inequality*

$$(u_1v_1 + u_2v_2)^2 \leq (u_1^2+u_2^2)(v_1^2+v_2^2)$$

In the inner product space of random variables where $\langle X,Y\rangle = E(XY)$, the Cauchy-Schwarz Inequality gives $E(XY)^2 \leq E(X^2)E(Y^2)$.

In the inner product space of continuous functions on the domain $[-1,1]$, where $\langle f,g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$, Cauchy-Schwarz Inequality gives $\left(\int_{-1}^{1} f(x)g(x)\,dx\right)^2 \leq \left(\int_{-1}^{1} f(x)^2\,dx\right)\left(\int_{-1}^{1} g(x)^2\,dx\right)$

Just as a version of Pythagoras' Theorem is true in every inner product space, so too is a version of another staple of secondary-school geometry: the Triangle Inequality.

Aside. *Yes, the triangle inequality has a proof! It is not a mathematical fact that we are required to believe just because someone told us it was true. Just about every mathematical fact in the secondary school curriculum that isn't just the definition of a piece of notation has a proof that we are capable of understanding. And even those things that are definitions of pieces of notation usually have some justification. Faith in the correctness of the mathematics curriculum is not a pre-requisite for teaching.*

Theorem 10.4 (The Triangle Inequality for Inner Product Spaces)**.** *Let V be an inner product space. For all $u, v \in V$ we have*

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

Proof. Let V be an inner product space and let $u, v \in V$. Using the properties of the inner product, we compute

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u+v\rangle + \langle v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \end{aligned}$$

Consider the sum $\langle u, v\rangle + \overline{\langle u, v\rangle}$. Since $\langle u, v\rangle \in \mathbf{F}$, there exists $a, b \in \mathbf{R}$ so that $\langle u, v\rangle = a + bi$. (If $\mathbf{F} = \mathbf{R}$, then $b = 0$). Therefore $\langle u, v\rangle + \overline{\langle u, v\rangle} = a + bi + a - bi = 2a = 2Re\langle u, v\rangle$. And so

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

By definition, we have $|\langle u, v\rangle| = \sqrt{a^2 + b^2}$. Since $\sqrt{a^2 + b^2} \geq a$ we have $2|\langle u, v\rangle| \geq 2Re\langle u, v\rangle$. And so

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

By Cauchy-Schwarz inequality, we have $|\langle u, v\rangle| \leq \|u\|\|v\|$. Therefore

$$\begin{aligned} \|u+v\|^2 &\leq \langle u, u\rangle + \langle v, v\rangle + 2\|u\|\|v\| \\ &\leq (\|u\| + \|v\|)^2 \end{aligned}$$

Taking the square root of both side yields the desired inequality. □

10.1 Test Your Understanding

1. The proof of Pythagoras' Theorem for Inner Product Spaces more or less consists of a single sequence of equalities:

$$\begin{aligned}\langle u+v, u+v\rangle &= \langle u, u+v\rangle + \langle v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle\end{aligned}$$

   Using the definition of inner product and any relevant results from Module 9 explain how each equality in this proof follows.

2. The proof of the Cauchy-Schwarz inequality contains the following equality:

$$\|cv\|^2 + \|w\|^2 = |c|^2\|v\|^2 + \|w\|^2$$

   Using relevant results and definitions from Module 9, explain why this equality holds.

10.1 Test Your Understanding Solution

1. The equality $\langle u+v, u+v\rangle = \langle u, u+v\rangle + \langle v, u+v\rangle$ holds because by definition, inner products are additive in the first entry. The equality $\langle u, u+v\rangle + \langle v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle$ holds because by Theorem 9.16, inner products are additive in the second entry.

2. Consider $\|cv\|^2$. By the definition of norm we have $\|cv\|^2 = \langle cv, cv\rangle$. By property 4 in the definition of inner product we have $\langle cv, cv\rangle = c\langle v, cv\rangle$. By Theorem 9.17 we have $c\langle v, cv\rangle = c\bar{c}\langle v, v\rangle$. By Theorem 9.9 we have $c\bar{c}\langle v, v\rangle = |c|^2\langle v, v\rangle$. By the definition of norm we have $|c|^2\langle v, v\rangle = |c|^2\|v\|^2$. Therefore $\|cv\|^2 + \|w\|^2 = |c|^2\|v\|^2 + \|w\|^2$.

---

11 Change of Basis

Recall the linear map from Part II of the midterm:

Let $T : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_2(\mathbf{R})$ be a linear map. Let $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$ be the matrix of T with respect to the bases $(1,\ x,\ x^2)$ and $(1,\ 2x,\ x^2)$.

Consider the following computation:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 11 \\ 0 \end{bmatrix}$$

This matrix computation gives a corresponding computation of T. The basis of the domain of the linear map is $(1,\ x,\ x^2)$. Therefore the vector $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ corresponds to the element $1(1) + 2(x) + 3(x^2) = 1 + 2x + 3x^2$ in $\mathcal{P}_2(\mathbf{R})$. The basis of the codomain of the linear map is $(1,\ 2x,\ x^2)$. Therefore the vector $\begin{bmatrix} 2 \\ 11 \\ 0 \end{bmatrix}$ corresponds to the element $2(1) + 11(2x) + 0(x^3) = 1 + 22x$. In other words, $T(1 + 2x + 3x^2) = 1 + 22x$.

Consider now the linear map $S : \mathcal{P}_2(\mathbf{R}) \to \mathcal{P}_2(\mathbf{R})$ so that the matrix of S with respect to the bases $(1,\ x,\ x^2)$ and $(1,\ 2x,\ x^2)$ is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Performing the same computation as above, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

As the basis of the codomain is $(1,\ 2x,\ x^2)$, the vector $\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$ corresponds to the element $1(1) + 1(2x) + 3(x^2) = 1 + 2x + 3x^2$. In other words $S(1 + 2x + 3x^2) = 1 + 2x + 3x^2$.

Hmm. The linear operator doesn't change the element $f(x) = 1 + 2x + 3x^2$, but the matrix operation changes the representation of f in one basis, $B = (1, x, x^2)$, to a representation in a different basis $B' = (1, 2x, x^2)$.

Consider now $g(x) + a + bx + cx^2$. The matrix of g with respect to the basis $(1, x, x^2)$ is $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$

Computing we find:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a \\ b/2 \\ c \end{bmatrix}$$

As the basis of the codomain is $(1,\ 2x,\ x^2)$, the vector $\begin{bmatrix} a \\ b/2 \\ c \end{bmatrix}$ corresponds to the element $1(a) + \frac{b}{2}(2x) + c(x^2) = a + bx + cx^2$. In other words $S(a + bx + cx^2) = a + bx + cx^2$. The linear operator S is the identity function! That is, $Sv = v$ for all $v \in \mathcal{P}_2(\mathbf{R})$. What's more, the matrix of S converts a matrix of a vector with respect to the basis $(1, x, x^2)$ into the matrix of the same vector with respect to the basis $(1, 2x, x^2)$.

Let us generalize. Let V be a finite-dimensional vector space and B and B' be bases of V. Let us think a little about a matrix of identity linear operator with respect to B and B'.

Recalling our method of construction of $\mathcal{M}(I)$, we first express Iv as a linear combination of the elements of B' for each $v \in B$. As I is the identity linear map, we have $Iv = v$ for all $v \in B$. Therefore the columns of the matrix $\mathcal{M}(I)$ are the coefficients with each element of the basis B is expressed as a linear combination of the basis B'.

As I is the identity linear operator, we have $Iw = w$ for all $w \in V$. From our example in above we saw

$$\mathcal{M}(I)\mathcal{M}_B(w) = \mathcal{M}_{B'}(w)$$

where $\mathcal{M}(I)$ is the matrix of I with respect to with respect to B and B', $\mathcal{M}_B(w)$ is the matrix of w with respect to B and $\mathcal{M}_{B'}(w)$ s the matrix of w with respect to B'. In other words, multiplying the matrix of I by the matrix of a element of V with respect to the basis B gives the matrix of that same element with respect to the basis B'.

We proceed with an example in $\mathbf{R}^2$. Let $B = (1, 1), (1, 0)$ and let $B' = (-1, 0), (-1, 2)$. To find the matrix of I with respect to with respect to B and B' we first express Iv as a linear combination of elements of B' for each $v \in V$

$$I(1, 1) = (1, 1) = -\frac{3}{2}(-1, 0) + \frac{1}{2}(-1, 2)$$
$$I(1, 0) = (1, 0) = -1(-1, 0) + 0(-1, 2)$$

Therefore the matrix of I with respect to with respect to B and B' is

$$\begin{bmatrix} -3/2 & -1 \\ 1/2 & 0 \end{bmatrix}$$

Using this matrix we can convert between representations of elements of $\mathbf{R}^2$ using B to representations using B'. For example, consider the vector $u = (3, -4)$. To find the matrix of u with respect to B we express u as a linear combination of elements of B.

$$v = -4(1, 1) + 7(1, 0)$$

Therefore the matrix of u with respect to B is

$$\begin{bmatrix} -4 \\ 7 \end{bmatrix}$$

Consider the following product:

$$\begin{bmatrix} -3/2 & -1 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} -4 \\ 7 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

Interpreting $\begin{bmatrix} -1 \\ -2 \end{bmatrix}$ as the matrix of a vector with respect to B', we find that this matrix corresponds to:

$$-1(-1, 0) + 2(-1, 2) = (3, -4) = u$$

For any $v \in \mathbf{R}^2$ we can construct the matrix of u with respect to B. Let x be this matrix. Multiplying the matrix of I with respect to with respect to B and B' by x gives the matrix of u with respect to B'.

We state this as a theorem:

Theorem 11.1. *Let V be a finite-dimensional vector space, let B and B' be bases of V and let $v \in V$ If A is the matrix of the identity linear operator with respect to B and B' and x is the matrix of v with respect to B, then Ax is the matrix of v with respect to B'.*

Theorem 11.1 tells us that we can use a matrix to convert between the representation of a vector as a matrix with respect to one basis to a representation of the same vector as a matrix with respect to a second basis. This work leads to the following definition.

Definition 11.2. *Let V be a finite dimensional vector space and let B and B' be bases of V. Let P be the matrix of the identity operator on V with respect to B and B'. We say P* ***is the change of basis matrix in V with respect to B and B'.***

Surprisingly, if P is the change of basis matrix in V with respect to B and B', then P^{-1} is the change of basis matrix in V with respect to B' and B. From our example above, we have From our example we have

$$\begin{bmatrix} -3/2 & -1 \\ 1/2 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 2 \\ -1 & -3 \end{bmatrix}$$

We can check that the matrix $\begin{bmatrix} 0 & 2 \\ -1 & -3 \end{bmatrix}$ is the change of basis matrix in $\mathbf{R}^2$ with respect to B' and B. (See Test Your Understanding 10.2)

Aside. *At the start of term, some students asked if we would cover the Discrete Fourier Transform (DFT) this semester. The setup for thinking about DFT takes a fair amount of background knowledge. In short, DFT is an identity linear map that transforms a vector which represents a sample with respect to time into a vector that represents that same sample as a linear combination of frequencies. In essence, this is a change of basis transformation. The sample does not change, but the way the sample is represented changes.*

Returning back to our work in eigenvalues and eigenvectors, let us consider a computationally useful application of change of basis.

The region of Eigenville is neatly divided into city, suburban and rural communities[9]. Each year the proportion of the population in each region changes as follows:

- 75% of people living in the city stay in the city; 10% of people living in the city move to the suburbs; and 15% of people living in the city move to a rural community.
- 60% of people living in the suburbs stay in the suburbs; 10% of people living in the suburbs move to the city; and 30% of people living in the suburbs move to a rural community.
- 70% of people living in a rural community stay in a rural community; 10% of people living in a rural community move to the city; and 20% of people living in a rural community move to suburbs.

Using this information we can study the long-term behaviour of the distribution of the population in Eigenville.

Let us think about the change in population distribution as a linear operator on $\mathbf{R}^3$. Our coordinates in $\mathbf{R}^3$ represent the proportion of people living in each region. For example, $p = \left(\frac{1}{2}, \frac{1}{8}, \frac{3}{8}\right)$ represents a population distribution where $\frac{1}{2}$ of the population lives in the city, $\frac{1}{8}$ of the population lives in the suburbs and $\frac{3}{8}$ of the population live in a rural community.

If p was the population distribution at the start of the year, then we can use the information above to find the distribution at the end of the year. Of the $\frac{1}{2}$ of the population that was in the city, three-quarters of them would stay in the city. Of the $\frac{1}{8}$ of the population that was in the suburbs, one-tenth of them would move to the city. Of the $\frac{3}{8}$ of the population that was in a rural community, one-tenth of them would move to the city. Therefore, at the end

[9]example courtesy of https://www.math.wustl.edu/∼freiwald/309markov.pdf. Though I am quite certain that this author pinched the example from somewhere else without attribution.

of the year the proportion of the population in the city would be:

$$\frac{3}{4}\left(\frac{1}{2}\right)+\frac{1}{10}\left(\frac{1}{8}\right)+\frac{1}{10}\left(\frac{3}{8}\right)=\frac{33}{80}=41.25\%$$

By repeating the computation above for the suburbs and the rural community, we find that if p was the population distribution at the start of a year, then at the end of the year the population distribution would be $p'=(0.4125,0.2375,0.35)$

Aside. *This might seem a little odd. There are plenty of elements of* $\mathbf{R}^3$ *that cannot be interpreted as a probability distribution. This turns out not to matter! We focus only on those elements of* $\mathbf{R}^3$ *that do correspond to probability distributions in Eigenville.*

Let T be the linear operator on $\mathbf{R}^3$ so that if v was the population distribution at the start of the year, then Tv is the population distribution at the end of the year. From our work above we have $T(0.5,0.125,0.375)=(0.4125,0.2375,0.35)$. Using our thinking from above, let us see if we can determine the population distribution of Eigenville after 20 years if the initial population distribution is given by p.

Since Tp is the population distribution after one year, evaluating $T(Tp)$ gives the population distribution after two years. And so to find the population distribution of Eigenville after 20 years we want to compute $(\underbrace{T\circ T\circ\cdots\circ T\circ T\circ T}_{20\text{ times}})(v)$. Using our notation for repeated composition, we want to compute $T^{20}p$.

The matrix of T with respect to the standard basis vectors is

$$A=\begin{bmatrix}3/4 & 1/10 & 1/10\\ 1/10 & 3/5 & 1/5\\ 3/20 & 3/10 & 7/10\end{bmatrix}$$

The matrix of p with respect to the standard basis vectors is

$$\begin{bmatrix}1/2\\ 1/8\\ 3/8\end{bmatrix}$$

Therefore $Tp=\begin{bmatrix}3/4 & 1/10 & 1/10\\ 1/10 & 3/5 & 1/5\\ 3/20 & 3/10 & 7/10\end{bmatrix}\begin{bmatrix}1/2\\ 1/8\\ 3/8\end{bmatrix}$ And so to find $T^{20}p$ we can compute

$$\begin{bmatrix}3/4 & 1/10 & 1/10\\ 1/10 & 3/5 & 1/5\\ 3/20 & 3/10 & 7/10\end{bmatrix}^{20}\begin{bmatrix}1/2\\ 1/8\\ 3/8\end{bmatrix}$$

Yuck! What a tedious calculation this would be! Certainly this computation would be easier if we could represent T using a diagonal matrix – finding the 20th power of a diagonal matrix amounts to finding the 20th power of each of the diagonal entries.

Recall the following result from Module 8.

Theorem. *Let V be a finite-dimensional vector space and let T be a linear operator on V. The linear map T is diagonalizable if and only if V has an eigenbasis with respect to T.*

Using the matrix of T with respect to the standard basis vectors we find that T has eigenvalues $\lambda_1 = \frac{13}{20}, \lambda_2 = 1$ and $\lambda_3 = \frac{2}{5}$. (We know from Assignment 4 that the eigenvalues of the linear operator and the eigenvalues of the matrix are the same.) By Corollary 8.8, V has an eigenbasis, $E_{\mathbf{R}^3}$. The matrix of T with respect to this eigenbasis is

$$D = \begin{bmatrix} 13/20 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2/5 \end{bmatrix}$$

Since D is the matrix of T with respect to our eigenbasis, to compute with D we must express our population distribution as linear combinations of the eigenbasis, as opposed to the standard basis in $\mathbf{R}^3$. From our work above, we can do this by computing P, the change of basis matrix in $\mathbf{R}^3$ with respect to the standard basis and $E_{\mathbf{R}^3}$.

If P is the change of basis matrix that changes representations of elements of $\mathbf{R}^3$ with respect to the standard basis to representations with respect to the basis $E_{\mathbf{R}^3}$, then Pp is the matrix of p with respect to the basis $E_{\mathbf{R}^3}$. Therefore $D(Pp)$ gives the population distribution of Eigenville after one year, expressed with respect to the basis $E_{\mathbf{R}^3}$. And so $P^{-1}(DPp)$ gives the population distribution of Eigenville after one year, expressed with respect to the standard basis. In other words $Ap = P^{-1}DPp$

Therefore

$$A^{20}p = (P^{-1}DP)^{20}p$$

It is not immediately clear yet why this is helpful. Let us start expanding on the right side and see what we find:

$$(P^{-1}DP)^{20}p = (P^{-1}DP)^{18}(P^{-1}DP)(P^{-1}DP)p$$

Notice the product PP^{-1} appearing above on the right. (Thank goodness matrix multiplication is associative!) By definition PP^{-1} is the identity matrix, therefore

$$(P^{-1}DP)^{20}p = (P^{-1}DP)^{18}(P^{-1}D^2P)p$$

Repeating this process again:

$$\begin{aligned}(P^{-1}DP)^{18}(P^{-1}D^2P)p &= (P^{-1}DP)^{17}(P^{-1}DP)(P^{-1}D^2P)p \\ &= (P^{-1}DP)^{17}(P^{-1}D^3P)p\end{aligned}$$

And again...

$$\begin{aligned}(P^{-1}DP)^{17}(P^{-1}D^3P)p &= (P^{-1}DP)^{16}(P^{-1}DP)(P^{-1}D^4P)p \\ &= (P^{-1}DP)^{16}(P^{-1}D^4P)p\end{aligned}$$

And again...

$$\begin{aligned}(P^{-1}DP)^{16}(P^{-1}D^4P)p &= (P^{-1}DP)^{15}(P^{-1}DP)(P^{-1}D^5P)p \\ &= (P^{-1}DP)^{15}(P^{-1}D^5P)p\end{aligned}$$

And after another fifteen times, we find

$$A^{20}p = (P^{-1}D^{20}P)p$$

In general we have the following result:

Theorem 11.3. *Let A, D and P be matrices so that $A = P^{-1}DP$. If D is a diagonal matrix, then for all $k \geq 1$ we have $A^k = P^{-1}D^kP$.*

We omit a general proof of this result, but note that one may use the Principle of Mathematical Induction to prove this statement is true for all integers $k \geq 1$.

From our work above, we have $T^{20}p = (P^{-1}D^{20}P)p$. That is

$$T^{20}p = P^{-1}\begin{bmatrix}\left(\frac{13}{20}\right)^{20} & 0 & 0 \\ 0 & 1^{20} & 0 \\ 0 & 0 & \left(\frac{2}{5}\right)^{20}\end{bmatrix}P\begin{bmatrix}1/2 \\ 1/8 \\ 3/8\end{bmatrix}$$

Once we have computed P and P^{-1}, finding the population distribution after a different number of years or with respect to a different initial distribution takes almost no work at all. We can replace 20 with any number of years that we want, and replace p with any initial population distribution we want.

---

11 Test Your Understanding

1. Let $B = (1,1),(1,0)$ and let $B' = (-1,0),(-1,2)$. Confirm $\begin{bmatrix} 0 & 2 \\ -1 & -3 \end{bmatrix}$ is the change of basis matrix in $\mathbf{R}^2$ with respect to B' and B by computing the matrix of the identity linear operator in $\mathbf{R}^3$ with respect to B' and B.
2. In the Eigenville example, we can find that $E_{\mathbf{R}^3} = (-5/3, 2/3, 1), (2, 2, 3), (0, -1, 1)$ is an eigenbasis whose vectors respectively correspond to the eigenvalues $\lambda_1 = 13/20$, $\lambda_2 = 1$ and $\lambda_3 = 2/5$. Compute P and P^{-1}.

---

11 Test Your Understanding Solution

1. We find

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

Therefore the change of basis matrix in $\mathbf{R}^2$ with respect to B' and B is $\begin{bmatrix} 0 & 2 \\ -1 & -3 \end{bmatrix}$

2. By definition, P is the matrix of the identity linear operator on $\mathbf{R}^3$ with respect to the standard basis and the basis $E_{\mathbf{R}^3} = (-5/3, 2/3, 1), (2, 2, 3), (0, -1, 1)$. Expressing the standard basis vectors as a linear combination of the eigenbasis, we find

$$(1,0,0) = -3/7(-5/3, 2/3, 1) + 1/7(2,2,3) + 0(0,-1,1)$$
$$(0,1,0) = 6/35(-5/3, 2/3, 1) + 1/7(2,2,3) + -3/5(0,-1,1)$$
$$(0,0,1) = 6/35(-5/3, 2/3, 1) + 1/7(2,2,3) + 2/5(0,-1,1)$$

Therefore

$$P = \begin{bmatrix} -3/7 & 6/35 & 6/35 \\ 1/7 & 1/7 & 1/7 \\ 0 & -3/5 & 2/5 \end{bmatrix}$$

We compute:

$$P^{-1} = \begin{bmatrix} -5/3 & 2 & 0 \\ 2/3 & 2 & -1 \\ 1 & 3 & 1 \end{bmatrix}$$

Appendices

A Mathematical Induction

Like *proof by contradiction*, *proof by induction* is a proof technique we can use to justify the truth of some mathematical facts. We begin with an example that has nothing to do with linear algebra.

Let n be a positive integer. Consider the following sum:

$$s(n) = 1 + \sum_{i=1}^{n} 2^i = 1 + 2^1 + 2^2 + \cdots + 2^n$$

When n is small it is easy to compute values of $s(n)$:

$$\begin{aligned} s(1) &= 3 \\ s(2) &= 7 \\ s(3) &= 15 \\ s(4) &= 31 \end{aligned}$$

You may notice each of these values is one fewer than a power of 2.

$$\begin{aligned} s(1) &= 2^2 - 1 \\ s(2) &= 2^3 - 1 \\ s(3) &= 2^4 - 1 \\ s(4) &= 2^5 - 1 \end{aligned}$$

We wonder does this pattern hold in general. That is, does $s(n) = 2^{n+1} - 1$ for each positive integer n? Let us consider the case $n = 5$, but rather than compute directly, let us try a different approach:

$$s(5) = 1 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5$$

Notice $1 + 2^1 + 2^2 + 2^3 + 2^4 = s(4)$. We have already confirmed $s(4) = 2^5 - 1$. Thus

$$s(5) = \left(1 + 2^1 + 2^2 + 2^3 + 2^4\right) + 2^5 = \left(2^5 - 1\right) + 2^5$$

Simplifying, we notice $2^5 + 2^5 = 2(2^5) = 2^6$. Therefore

$$s(5) = 2^6 - 1$$

Using this same technique, we can verify $s(6) = 2^7 - 1$. And using $s(6) = 2^7 - 1$ we can use the same technique to verify $s(7) = 2^8 - 1$. (If this is not clear to you, take a moment to do the calculation. Compute $s(6)$ by noticing that the sum of the first 6 terms in $s(6)$ is equal to $s(5)$)

Using this reasoning, once we have verified that our formula holds for any particular positive integer k, then we can then verify that our formula holds for $s(k+1)$. This idea is the central component of mathematical induction. Let us formalize.

Let $P(n)$ denote a statement that is either true or false for each integer $n \geq 1$. For example, if $P(n)$ is the statement *"n is odd"*, then $P(2)$ is false and $P(1)$ is true.

If $Q(n)$ is the statement

$$1 + \sum_{i=1}^{n} 2^i = 2^{n+1} - 1$$

then above we have confirmed that each of $Q(1), Q(2), Q(3), Q(4)$ and $Q(5)$ is true. In asking if

$$1 + \sum_{i=1}^{n} 2^i = 2^{n+1} - 1$$

holds for each integer $n \geq 1$, we are asking if $Q(n)$ is true for each $n \geq 1$.

Imagine you are provided with the following information about a statement $P(n)$ that is either true or false for each integer $n \geq 1$.

1. $P(1)$ is true; and
2. for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true

Can we conclude $P(2)$ is true?

Statement 2 tells us that knowing $P(1)$ is true tells us $P(2)$ is true. Thus, to know if $P(2)$ is true, it is enough to know $P(1)$ is true. Statement 1 tells us that $P(1)$ is true, thus we can conclude $P(2)$ is true.

Can we conclude $P(3)$ is true?

Statement 2 tells us that knowing $P(2)$ is true tells us $P(3)$ is true. Thus, to know if $P(3)$ is true, it is enough to know $P(2)$ is true. We have already concluded that $P(2)$ is true, thus we can conclude $P(3)$ is true.

A similar line of reasoning tells $P(4)$ is true. Once we have concluded $P(4)$ is true, then we can conclude $P(5)$ is true, and so on. Thus, knowing that statements 1. and 2. hold is enough to convince us that $P(n)$ is true for each $n \geq 1$.

We express this idea with the following theorem.

Theorem (The Principle of Mathematical Induction)**.** *Let $P(n)$ be a statement that is either true or false for each integer $n \geq 1$. If the following two statements hold, then $P(n)$ is true for each integer $n \geq 1$.*

1. *$P(1)$ is true; and*
2. *for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true*

The Principle of Mathematical Induction is a theorem. It has a hypothesis and a conclusion. The hypothesis is:

1. *$P(1)$ is true; and*
2. *for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true*

The conclusion is:
$P(n)$ is true for each integer $n \geq 1$.

Given some particular statement $P(n)$ we can prove $P(n)$ is true for all $n \geq 1$ by showing that $P(n)$ satisfies the hypotheses of the The Principle of Mathematical Induction.

Let us consider applying The Principle of Mathematical Induction to study $s(n)$ above. Let $P(n)$ be the statement

$$1 + \sum_{i=1}^{n} 2^i = 2^{n+1} - 1$$

We can use the Theorem of Mathematical Induction, to prove $P(n)$ is true for each integer $n \geq 1$ by verifying that 1. and 2. hold for this $P(n)$.

1. $P(1)$ is true

$P(1)$ is the statement

$$1 + \sum_{i=1}^{1} 2^i = 2^2 - 1$$

We compute $1 + \sum_{i=1}^{1} 2^i = 1 + 2^1 = 3$ and $2^2 - 1 = 3$. Therefore $P(1)$ is true. And so statement 1. holds in the Principle of Mathematical Induction

2. *for each $k \geq 1$, if $P(k)$ is true, then $P(k+1)$ is true*
To show this statement is true, we assume the hypothesis is true and we use that to show that the conclusion is true. Let $k \geq 1$ be an integer so that $P(k)$ is true. Thus the following statement is true:

$$1 + \sum_{i=1}^{k} 2^i = 2^{k+1} - 1$$

We want to show $P(k+1)$ is true. That is, we want to show

$$1 + \sum_{i=1}^{k+1} 2^i = 2^{k+2} - 1$$

is true.

Consider the sum

$$1 + \sum_{i=1}^{k+1} 2^i$$

The last term of this sum is 2^{k+1}. Therefore

$$1 + \sum_{i=1}^{k+1} 2^i = \left(1 + \sum_{i=1}^{k} 2^i\right) + 2^{k+1}$$

By hypothesis $P(k)$ is true. And so $1 + \sum_{i=1}^{k} 2^i = 2^{k+1} - 1$. Thus

$$1 + \sum_{i=1}^{k+1} 2^i = \left(1 + \sum_{i=1}^{k} 2^i\right) + 2^{k+1} = \left(2^{k+1} - 1\right) + 2^{k+1} = 2^{k+2} - 1$$

Therefore $P(k+1)$ is true. And so we see statement 2 holds the Principle of Mathematical Induction.

Since 1. and 2. hold in the Principle of Mathematical Induction, so must the conclusion. Therefore $P(n)$ is true for each $n \geq 1$. That is,

$$1 + \sum_{i=1}^{n} 2^i = 2^{n+1} - 1$$

for each $n \geq 1$.

Aside. *If you have seen induction before you might find notable the lack of appearance of the phrases* base case*,* induction hypothesis *and* induction step*. As long as it is clear that the author is showing that the two statements hold, these phrases need not be used.* Base Case *refers to statement 1.* Induction hypothesis *refers to the hypothesis of statement 2.* Induction step *refers to proving that the conclusion of statement 2 holds, given that the hypothesis is true. This is likely the only time these phrases will appear in these notes. This author abhors them and suspects their use contributes to students thinking that induction is* magic*.*

The application of the Theorem of Mathematical Induction is so ubiquitous in mathematics, that one need only mention the word *induction* in the proof for the reader to know what to expect. For example, if the author knows that their reader is well-experienced with induction the following proof would suffice.

Theorem. *If n is a positive integer, then*

$$\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$$

Proof. We proceed by induction on n, noting $1 + 2^1 = 3$.

Consider $n = k + 1$. By induction, $1 + \sum_{i=1}^{k} 2^i = 2^{k+1} - 1$.

Thus,

$$1 + \sum_{i=1}^{k+1} 2^i = \left(1 + \sum_{i=1}^{k} 2^i\right) + 2^{k+1} = \left(2^{k+1} - 1\right) + 2^{k+1} = 2^{k+2} - 1.$$

□

Aside. *The Principle of Mathematical Induction is a theorem. This suggests that one can write down a proof of the Principle of Mathematical Induction. This is true, but not immediately relevant to us. We will not concern ourselves with the proof of the Principle of Mathematical Induction.*

Let us consider now an application of the Principle of Mathematical Induction germane to our study of polynomials in Module 7. From our work in pre-calculus, perhaps we recall the following result.

Theorem A.1. *For every $n \geq 1$, a polynomial of degree n has at most n zeroes.*

Let $P(n)$ be the statement

Every polynomial of degree n has at most n zeroes.

To prove our theorem, we want to show $P(n)$ is true for every $n \geq 1$.

Consider the statement $P(1)$:

Every polynomial of degree 1 has at most 1 zero.

A polynomial of degree 1 is a polynomial of the form $p(x) = ax + b$. Such a polynomial crosses the x-axis at $x = -\frac{b}{a}$. Therefore $p\left(-\frac{b}{a}\right) = 0$. And so p has at least 1 zero. Therefore $P(1)$ is true.

We apply the Principle of Mathematical Induction to show $P(n)$ is true for every $n \geq 1$. This then provides a proof of our theorem.

We have already shown that 1. holds in the statement of the Principle of Mathematical Induction for the statement $P(n)$:

Every polynomial of degree n has at most n zeroes.

And so now we need only show that 2. holds in the statement of the Principle of Mathematical Induction.

Let $k \geq 1$ be an integer so that $P(k)$ is true. Let p be a polynomial of degree $k+1$. Polynomials of degree $k+1$ can be partitioned in to those that cross that cross the x-axis and those that don't.

If p does not cross the x-axis, then p has no zeroes. If p has no zeroes, then p has at most $k+1$ zeroes as $0 \leq k+1$.

Otherwise, assume p crosses the x-axis at $x = \lambda$. Therefore $(x - \lambda)$ is a linear factor of p. Therefore dividing p by the factor $(x - \lambda)$ yields no remainder. In other words, there exists a polynomial q of degree k so that

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

Since $P(k)$ is true, q has at most k zeroes.

To show p has at most $k+1$ zeroes, we show that every zero of p is either λ or a zero of q. Since q has at most k zeroes, it would then follow that p has at most $k+1$ zeroes – all of the zeroes of q and also λ. Let $r \neq \lambda$ be a zero of p.

Since $r \neq \lambda$,

$$0 = p(r) = (r - \lambda)q(r)$$

Since $r \neq \lambda$ we have $r - \lambda \neq 0$. Since $0 = (r - \lambda)q(r)$ and $r - \lambda \neq 0$ we have that $q(r) = 0$. Therefore every zero of p that is not equal to λ must be a zero of q.

Since q has at most k zeroes, then p has at most $k+1$ zeroes. Therefore every polynomial of degree at most $k+1$ has $k+1$ zeroes. In other words, $P(k+1)$ is true.

Since statement 1 and statement 2 hold in the Principle of Mathematical Induction, so must the conclusion. Therefore $P(n)$ is true for each $n \geq 1$. In other words, for every $n \geq 1$ a polynomial of degree n has at most n zeroes.

Index

Table of Notation