Math 225 Linear Algebra

Haim Grebnev

1 Introduction

- My name is Haim Grebnev, I'm a postdoctoral scholar. My job is to conduct research and to teach. My field of research is inverse problems with a focus on geometric analysis. Inverse problems is a field that studies math arising from various imaging techniques such as CT scans, sonar sensing, electric impedance tomography, etc. Currently I work on a generalization of the equations that arise in X-ray imaging that is used in polarimetric neutron tomography called the non-Abelian X-ray transform.
- In this course we study linear algebra – a field no less important than calculus which embeds itself in nearly every area of math, physics, computer science, and more. Applications range from plotting the optimal route between two points on Google/Bing maps to tuning parameters in AI models that are learning to perform various tasks. This course also will also be an introduction mathematical rigor, where you will learn to justify mathematical ideas beyond doubt (also known as "writing proofs").
- Class structure.

2 Mathematics

- Most likely in all your previous math courses you didn't prove any of the theorems or equations that you learned (if you have: I'm glad to hear it). Why prove things? The answer should be obvious: how do you know that the theorems or equations you learned are correct. Let me point out that theorems and equations didn't fall out of the heavens, they were discovered by humans. And we sure hope that our predecessors discovered them correctly! That's why with every theorem and equation we provide a proof for why it's correct and hence learning them is a vital tradition.
- Another important reason for learning proofs is that they give insight into the mechanisms that make existing theorems and equations work. This is essential for innovation because knowing how theorems and equations work in existing contexts gives you a good starting point for figuring out how they work in new contexts, or how to correctly formulate new theorems and equations that build off old ideas. For instance, in flat Euclidean space the angles of a triangle always add up to 180°. But what about on a curved space, such as the surface of a sphere?
- The last important reason that I'll mention for learning proofs is that many important mathematical techniques are embedded in them. Theorem statements and equations only represent a surface level quantity of useful mathematical techniques. Hence working through proofs is in essence a way to work through examples of successful problem solving whose goal is to arrive at new mathematical results.

3 Mathematical Statements

- To prove things, we need statements, or else how do we know what we want to prove? A **proposition** (or **statement** or **assertion**) is a statement that can objectively either be true or false. Some examples are (some here are true and some are false):
    - The number 2 is smaller than the number 1.
    - The angles of a triangle in Euclidean space add up to 180°.
    - For any number x, $x^2 + 4 \geq 4x$
    - All cats have two eyes.
    - A planet of mass m accelerates towards a planet of mass M with acceleration $\frac{GM}{r^2}$ where r is the distance between them (G is a fixed constant you can look up in a textbook).

  These examples come with caveats; the first being the fourth one which delves a little bit into philosophy. In particular we need to consider the following question: what is a cat, how do you define a cat? Do you consider it a large conglomerate of molecules? But in that case it will be difficult to define its distinction from a great deal of things, such as my sandwich. Is it a four-legged creature? But then my neighbor's dog fits that description. If you think about it: it's hard to rigorously define a cat. Biologists have endeavored to give a relatively quantitative way to think about this in terms of genera and species.

  The 5th question suffers the same phenomenon: what is a planet? Furthermore, acceleration is defined using the continuous axis of the real numbers. How do we know that objects in our universe travel in such a medium: they may be jumping on discrete distances on extremely small scales.

- Hence mathematics tries to stray away from dealing with physical objects and only deal with abstract objects called sets (more on that later) such as numbers. Yet still, the second question posed above is also problematic because we have to consider the following question: how do we define angles? Before we can prove the second statement, we must defines angles, and even triangles! But that can be done very precisely.

- What I described in the previous paragraph is a rather modern point of view of math. Back in Euclid's day, they proposed to work on geometry by declaring axioms about points and lines without defining the latter two, and then work from there. Today we've shifted the undefined objects and axioms into deep set theory and to define things using sets. We also point out that modern mathematics has started to consider related, but different, objects of study called "categories," but we won't get into that.

- There is a field of math called mathematical logic, which allows us to quantitively talk about statements, implications, and so on. If p is a statement and q is a statement, then $p \implies q$ stands for "p implies q." In this case, we always assume that the items before the "$\implies$" are true. For example

- $p =$ "x is a number," $q =$ "it holds that $x^2 + 4 \geq 4x$" then $p \implies q$ is our third proposition above:

  "If x is a number, then it holds that $x^2 + 4 \geq 4x$"

  (alternative) "x being a number implies that $x^2 + 4 \geq 4x$"

Notice that I can write this using if and then/implies. This is why $p \implies q$ is called an **if…then…(implies) statement**.

- You can also put the logical operators "**and**" (a.k.a. **conjunction**) and "**or**" (a.k.a. **disjunction**) into statements, which are denoted by "$\wedge$" and "$\vee$" respectively. For instance, $(p \wedge q) \implies r$ means "p and q imply r." Examples include.

  - $p =$ "T is a right triangle," "$q =$ "a, b, c are the length of the sides of T with c being the biggest", $r =$ "$a^2 + b^2 = c^2$." Then $(p \wedge q) \implies r$ is the famous Pythagorean theorem:

    "If T is a right triangle *and* a, b, c are the length of the sides of T with c being the biggest, then $a^2 + b^2 = c^2$."

    You can also rewrite this using "implies."

  - $p =$ "$x < 0$", $q =$ "$y < 0$," $r =$ "$xy < 0$." Then $(p \vee q) \implies r$ is the statement

    "If $x < 0$ or $y < 0$, then $xy < 0$"

    This is an example of a false statement, because you can take the counterexample $x = -2$, $y = -3$, but $xy = 6$ which is not less than zero.

- If p is a statement and q is a statement, then $p \Leftrightarrow q$ stands for "p implies q and q implies p." This can alternatively be written as "p is true if and only if q is true" (think about why this is true!). This is why "$\Leftrightarrow$" is called **if and only if** or **iff** for short. An example includes

  - $p =$ "x is even," $q =$ "$x + 2$" is even. Then "$p \Leftrightarrow q$" is

    "x is even if and only if $x + 2$ is even."

  Although not often used, we mention that "$p \Leftarrow q$" means q implies p. We typically write this however as "$q \implies p$."

- **Definition 3.1:** Suppose that p and q are statements and consider the statement "$p \implies q$." The **converse** of the statement "$p \implies q$" is the statement "$q \implies p$" (or equivalently "$p \Leftarrow q$")

- As a remark regarding the above definition, we point out that $p \Leftrightarrow q$ can then be reformulated as that $p \implies q$ *and* its "converse" $q \implies p$ are true.

4 Sets

- Modern mathematics is the study of phenomenon surrounding sets and maps between them.[1] We won't define sets, because that's a question too deep for this course and is an entire fascinating field in of itself. From our point of view, a **set** is a collection of objects such as a collection of points in the plane (e.g. triangle, lines), numbers, sets of functions, etc. Individual items are called **elements** in them. Examples are
    - The set of even numbers between 2 and 8: $\{2,4,6,8\}$
    - The line in two-dimensional space passing through zero with slope 2: $\{(x,y) \in \mathbb{R}^2 : y = 2x\}$.
    - My breakfast this morning (not really a mathematical example): {Bread, Cheese, Milk, Tea}
- Some famous set whose notation is standard are:

  $\mathbb{N} =$ the set of all positive whole numbers (does not include zero). Informally $\mathbb{N} = \{0,1,2,\dots\}$.

  $\mathbb{Z} =$ the set of all whole numbers. Informally $\mathbb{Z} = \{\dots,-2,-1,0,1,2,\dots\}$.

  $\mathbb{Q} = \left\{\frac{a}{b} : a,b \in \mathbb{N} \text{ and } b \neq 0\right\}$. This is the set of all rational numbers.

  $\mathbb{R} =$ The set of all real numbers.

  $\mathbb{C} = \left\{a + b\sqrt{-1} : a,b \in \mathbb{R}\right\}$. This is the set of all complex numbers.
- **Notation 4.1:** If S is a set and a is an element in S, then we write $a \in S$. If a is not in S, we write $a \notin S$.
- **Notation 4.2:** We denote the **empty set** (i.e. the set with nothing in it) as $\emptyset$.
- **Definition 4.3:** Suppose that S and Q are sets. We say that S is a **subset** of Q (denote by $S \subseteq Q$) if every element in S is in Q (draw a picture!). Although not used as often, an equivalent way of saying this is that Q is a **superset** of S (denoted by $Q \supseteq S$). It's easy to see that two sets are equal $S = Q$ if and only if both $S \subseteq Q$ and $Q \subseteq S$.
- **Definition 4.4:** Suppose that A and B are sets. The **intersection** of A and B is the set of all elements that are contained both in A and B:

  $$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

  The **union** of A and B is the set of all elements that are either in A or B:

  $$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

[1] Technically maps between sets are also sets themselves.

5 Quantifiers

- Mathematicians make use of the following shorthand notation due to their widespread use in statements
    - $\forall$= "For all" or equivalently "for any"
    - $\exists$= "there exists"

  To help you remember them, the "$\forall$" resembles an upside down "A" (as in All/Any) and "$\exists$" resembles a backwards "E" (as in Exists).

  Amusing comment: if later you're having trouble remembering if it is supposed to be an upside down or backwards "A" or an upside down or backwards "E," then note that two of these give A and E back again and hence can't be the right one!

- **Example 5.1:** The statement "For any real number x that is also a rational number, there exist integers a and $b \neq 0$ such that $x = a/b$" can be rewritten as follows:

$$\forall x \in \mathbb{R} : x \in \mathbb{Q},\ \ \exists a \in \mathbb{Z}\ \ \exists b \in \mathbb{Z} : b \neq 0 \text{ such that } x = \frac{a}{b}.$$

6 Proofs

6.1 Direct proofs

- There are three ways of proving things: proving something directly, by contradiction, and by induction. There is another method called "proof by contrapositive." We begin by demonstrating a **direct proof**, which proves something directly. Suppose a prior paper published and proved the following correct proposition:

- **Proposition 6.1:** For any real number x, $x^2 \geq 0$.

- **Proof:** (suppose the proof was provided in another paper) ■

- Let us build off of this to prove the following proposition:

- **Proposition 6.2:** Suppose that x is a number. Then $x^2 + 4 \geq 4x$.

- **Proof:** Fix any number x. We have that $x - 2$ is also number. By Proposition 6.1 above, we have that $(x-2)^2 \geq 0$. Expanding the left-hand side gives

$$x^2 - 4x + 4 \geq 0$$

  Taking $4x$ to the other side gives

$$x^2 + 4 \geq 4x$$

■

  (the symbol "■" means "end of proof.")

6.2 Proof by Contradiction/Contrapositive

- Next we discuss proof by contradiction and proof by contrapositive. We start with the former. Every statement p has what's called a **negation**, denote by $\neg p$, which is the statement referring to the condition when p is *not* true. Out loud "$\neg p$" is read as "not p." Some examples are
  - $p =$ "x is divisible by 2 (i.e. x is even)", $\neg p =$ "x is not divisible by 2 (i.e. x is odd)."
  - $p =$ "if n is even, then n^2 is even," $\neg p =$ "if n is even, then n^2 is not even"

  Note that $\neg(\neg p) = p$.
- Proofs by contradiction work as follows. Suppose we want to prove a statement p (i.e. that p is true). Proving it directly may be difficult, so we can try the following trick instead. Let's suppose for the moment that p is instead false and see what happens. In other words, we assume $\neg p$. If by clever and ingenious arguments we can show that assuming that p is false leads us to, or more precisely "implies," some sort of logical contradiction, then we conclude that p could not have been false in the first place and hence must be true. This is **proof by contradiction**!
- **Theorem 6.3:** Suppose that $n \in \mathbb{Z}$ is an integer. If n^2 is odd, then n is also odd.

  Proof: Suppose not: suppose there exists an $n \in \mathbb{Z}$ such that n^2 is odd but n is not odd!

  > *Remark:* This last sentence can also be written as "We will prove this by contradiction: suppose…" or "For the sake of contradiction, suppose…"

  Then n is divisible by 2, or in other words $m = n/2$ is also an integer. Then $n = 2m$, and hence

  $$n^2 = (2m)^2 = 4m^2.$$

  Thus n^2 is divisible by 2 since

  $$\frac{n^2}{2} = \frac{4m^2}{2} = 2m^2$$

  is also an integer. But that is a contradiction since we assumed that n^2 is odd and hence is not divisible by 2. Thus n must indeed be odd.

  ■

- Often in math we want to prove that one statement implies the other (i.e. $p \implies q$). An important result in logic is the following:
- **Theorem 6.4:** If p and q are statements

  $$(p \implies q) \quad \Leftrightarrow \quad (\neg q \implies \neg p).$$

  In words, "p implies q" is equivalent to "not q implies not p."

  Remark: Note that $p \implies q$ is not equivalent to $\neg p \implies \neg q$, which is a commonly made error.

  Proof: Let's start by proving the easier direction:

$$(p \Longrightarrow q) \quad \Longleftarrow \quad (\neg q \Longrightarrow \neg p).$$

In other words, we assume $\neg q \Longrightarrow \neg p$ and want to show that $p \Longrightarrow q$. To show that $p \Longrightarrow q$, suppose that p is true. We want to show that this implies that q is true. We do this by contradiction: suppose q is not true. But then by $\neg q \Longrightarrow \neg p$ we get that p is not true, which contradicts the fact that we assumed that p was true. Hence q must be true.

To prove the other direction:

$$(p \Longrightarrow q) \quad \Longrightarrow \quad (\neg q \Longrightarrow \neg p),$$

observe that by what we just proved,

$$(p \Longrightarrow q) = \big(\neg(\neg p) \Longrightarrow \neg(\neg q)\big) \quad \Longrightarrow \quad \big((\neg q) \Longrightarrow (\neg p)\big) = (\neg q \Longrightarrow \neg p)$$

(we just removed the unnecessary parentheses around $(\neg q)$ and $(\neg p)$ in the very last step).

■

- As an illustration of the above theorem and remark, try setting $p =$ "It is raining" and $q =$ "I bring an umbrella to school" and convincing yourself that "If it is raining then I bring an umbrella to school" and "If I didn't bring an umbrella to school then it isn't raining" are equivalent. However, neither statements imply or are implied by "If it isn't raining then I don't bring an umbrella to school" since if the former two hold, it may so happen that I simply bring an umbrella to school every day no matter what!

- **Proof by contrapositive** is the method of proving $p \Longrightarrow q$ by instead proving $\neg q \Longrightarrow \neg p$. The reason this is justified is that in the previous theorem we proved that these are equivalent. For the statement $p \Longrightarrow q$, the equivalent statement $\neg q \Longrightarrow \neg p$ is called its **contrapositive**. This is why this is called proof by contrapositive!

  We point out that

  > "Whenever you can perform a proof by contrapositive, you can also do a proof by contradiction."

  This is done as follows. Assume that p is true, show that $\neg q \Longrightarrow \neg p$ (which you would do for a proof by contrapositive), and then arrive at the contradiction that you have $\neg p$ but you assumed p (i.e. that p is true). Hence q must be true (i.e. you've showed that $p \Longrightarrow q$).

  It so happens that the proof of Theorem 6.3 can also be written as a proof by contrapositive. There $p =$ "n^2 is odd" and $q =$ "n is odd," and we want to show $p \Longrightarrow q$. The sentence "Suppose that $n \in \mathbb{Z}$ is an integer" is simply a supporting statement that provides context.

  Contrapositive proof of Theorem 6.3: We prove this by contrapositive. Suppose that n is even (i.e. $\neg q$). Then n is divisible by 2, or in other words $m = n/2$ is also an integer. Then $n = 2m$, and hence

  $$n^2 = (2m)^2 = 4m^2.$$

Thus n^2 is divisible by 2 since

$$\frac{n^2}{2} = \frac{4m^2}{2} = 2m^2$$

(i.e. we showed $\neg p$).

■

- We end with subsection with a few notes on negations:
- **Note 6.5:** If you have statements p and q, then the negation of "p is true and q is true" is that "either p is false or q is false." Mathematically this is written as

$$\neg(p \wedge q) \quad \Leftrightarrow \quad (\neg p) \vee (\neg q). \tag{6.6}$$

The negation of "p is true or q is true" is "p is false and q is false." Mathematically,

$$\neg(p \vee q) \quad \Leftrightarrow \quad (\neg p) \wedge (\neg q). \tag{6.7}$$

We note that something like "p is false" can be treated as a statement itself and so by (6.6) the negation of "p is false and q is true" is "p is true or q is false," or mathematically

$$\neg\big((\neg p) \wedge q\big) \quad \Leftrightarrow \quad \big(\neg(\neg p)\big) \vee (\neg q) \quad = \quad p \vee (\neg q).$$

- **Note 6.8:** If you have a sequence of statements using quantifiers and a final conclusion, it's actually easy to negate it. Consider the following statement

$$\forall m \in \mathbb{Z} : \underbrace{m \text{ is odd}}_{\text{condition}} \quad \forall n \in \mathbb{Z} : \underbrace{n \text{ is odd}}_{\text{condition}} \quad \exists k \in \mathbb{Z}, \quad \underbrace{(m+n) = 2k}_{\text{conclusion}}. \tag{6.9}$$

This turns out to be a true statement (it's a good exercise to prove this). Its negation is obtained by turning all "∀" to "∃" and "∃" into "∀" without changing the conditions, and change the conclusion to the opposite:

$$\exists m \in \mathbb{Z} : \underbrace{m \text{ is odd}}_{\text{condition same}} \quad \exists n \in \mathbb{Z} : \underbrace{n \text{ is odd}}_{\text{condition same}} \quad \forall k \in \mathbb{Z}, \quad \underbrace{(m+n) \neq 2k}_{\text{opposite conclusion}}. \tag{6.10}$$

This turns out to be a false statement (it has to be since the previous statement was correct). You might wonder why we would want to form false statements. Well the answer is simple: suppose you want to prove (6.9) by contradiction. Then you have to assume that it's false, or in other words assume its negation which is (6.10). Then go from there and show that this leads to a contradiction. This will then imply that (6.9) had to be correct in the first place.

Important: Since we didn't state this is a formal theorem, please be careful when applying this rule of thumb since sometimes statement using "∀" and "∃" are written in nonstandard ways. In particular, it may be hard to figure out what exactly the conclusion is. For instance, some authors may place them in front of the sentence!

6.3 Proof by Induction

- The third way of proving something is by (mathematical) induction and its variant "strong induction." We begin with **(mathematical) induction**. It's a very simple idea: suppose you have a sequence of statements $p_0, p_1, p_2, ...$ (it doesn't have to start at zero: it could be labeled for instance $p_2, p_3, p_4, ...$) and you want to show that they are all correct. Then you do the following:

  1. Prove that the first statement p_0 is correct.
  2. Prove that for arbitrary n, $p_n \implies p_{n+1}$.

  Then you get that all statements $p_0, p_1, p_2, ...$ are correct because of the chain reaction $p_0 \implies p_1 \implies p_2 \implies \cdots$.

  Strong induction is similar:

  3. Prove that the first statement p_0 is correct.
  4. Prove that for arbitrary n, $p_0, \dots, p_n \implies p_{n+1}$.

  Let's try an example of ordinary induction:[2]

- **Theorem 6.11:** For any integer $n \geq 0$, $n^3 + 2n$ is divisible by 3.

  Proof: We prove this by induction on n.

  > *Remark:* In other words, the statements we want to prove are $p_0, p_1, p_2, ...$ where each $p_n =$ "$n^3 + 2n$ is divisible by 3."

  The base case is $n = 0$ (i.e. p_0). If we plug in $n = 0$ we get that $n^3 + 2n = 0$ which is indeed divisible by 3. Next suppose that $n^3 + 2n$ is divisible by 3 (i.e. assume p_n is true), we will show that $(n+1)^3 + 2(n+1)$ is also divisible by 3 (i.e. we will show that p_{n+1} is also true and hence $p_n \implies p_{n+1}$). We have that expanding and distributing $(n+1)^3 + 2(n+1)$ gives

  $$(n+1)^3 + 2(n+1) = n^3 + \underbrace{3n^2} + \underbrace{3n} + \underbrace{1} + 2n + \underbrace{2} = (n^2 + 2n) + 3\left(\underbrace{n^2 + n + 1}\right).$$

  By the inductive hypothesis (i.e. that p_n is true) we know that $n^2 + 2n$ is divisible by 3 and hence can be written in the form $n^2 + 2n = 3m$ for some integer m. Thus the above number can be written as

  $$= 3(m + n^2 + n + 1).$$

  Thus $(n+1)^3 + 2(n+1)$ is indeed divisible by 3, which completes the induction.

  ■

- **Note 6.12:** A disadvantage of proof by induction is that it often doesn't tell you how someone came up with the theorem. The latter could come from a good guess, numerical trials, genius, etc.

[2] Example taken from https://tutors.com/lesson/mathematical-induction-proof-examples

Sometimes, but not often, a formula that is proven by induction can be derived in the first place by mimicking a proof by induction that would be used to prove it.

7 Functions/Maps

- **Definition 7.1:** Given two sets A and B, a **function** (or **map**) $f : A \to B$ from A to B is a rule that takes an element $a \in A$ and ouputs an element $f(b)$ in B.[3] The set A is called the **domain** of f, which is denote by $\mathrm{dom}\, f$. The **range** (or **image**) of f, denote by $\mathrm{range}\, f$, $\mathrm{ran}\, f$, or $\mathrm{Im}\, f$ is the set of all elements "hit" by f:

$$\mathrm{range}\, f = \{b \in B : \exists a \in A \text{ such that } b = f(a)\}.$$

Remark: Some people assign terminology for the set B in the above definition and some don't. Hence we refrain from calling it anything special.

- **Example 7.2:** The exponential function is a map of the form $\exp : \mathbb{R} \to \mathbb{R}$. In particular, for every $x \in \mathbb{R}$ this function outputs $\exp x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$. The domain of this function is $\mathbb{R}$ and the range is $\{y \in \mathbb{R} : y > 0\}$.

- **Example 7.3:** Consider the sets $S = \{1,2,3,\dots,10\}$ and $Q = \{1,2,3,4,5\}$ and the map $f : S \to Q$

$$f(n) = \begin{cases} (n+1)/2 & \text{if } n \text{ is odd} \\ n/2 & \text{if } n \text{ is even} \end{cases}$$

(draw it out!). In this case the domain of f is S and the range is all of Q.

- **Example 7.4:** Consider the function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = n + 1$ (draw it out!). Both the domain of f is $\mathbb{Z}$ and the range of f is $\mathbb{Z}$.

- **Definition 7.5:** Suppose that $f : A \to B$ is a function.

  1. We say that f is **injective** (or **one-to-one**) if $a_1 \neq a_2$ implies that $f(a_1) \neq f(a_2)$. By Theorem 6.4 this is equivalent to its contrapositive:

$$f(a_1) = f(a_2) \quad \Longrightarrow \quad a_1 = a_2.$$

     Loosely, this is described as no two distinct elements get mapped to the same thing.

  - We say that the function f is **surjective** (or **onto**) if $\mathrm{range}\, f = B$. Since $\mathrm{range}\, f \subseteq B$ is automatic, this is equivalent to $B \subseteq \mathrm{range}\, f$ or in other words

$$\forall b \in B,\ b \in \mathrm{range}\, f \quad \Leftrightarrow \quad \underbrace{\forall b \in B \quad \exists a \in A,\ b = f(a)}_{\text{most useful formulation}}.$$

[3] Strictly speaking, functions can be defined on a more fundamental level as subsets of the Cartesian product $A \times B$, but we won't take this approach in our course.

Loosely, this is described as "f hits everything."

- We say that f is **bijective** if it is both injective and surjective.

- The function in Example 7.2 is injective but not surjective. The function in Example 7.3 is not injective but is surjective (why?). Thus neither are bijective. However the function in Example 7.4 is both injective and surjective and thus bijective.

  As an illustration, we prove the last sentence. First we prove that the f in Example 7.4 is injective. Take $n_1, n_2 \in \mathbb{Z}$ such that $n_1 \neq n_2$. Then $n_1 + 1 \neq n_2 + 1$. Since $f(n_1) = n_1 + 1$ and $f(n_2) = n_2 + 1$, we have that $f(n_1) \neq f(n_2)$ and so f is indeed injective. Next let's prove that f is surjective. Take any $m \in \mathbb{Z}$. We need to show that there exists an n such that $m = f(n)$. Let $n = m - 1$. Then $f(n) = (m - 1) + 1 = m$ and so f is indeed surjective. Hence f is both injective and surjective, and thus bijective.

- The importance of the question of whether a function is injective, surjective, and/or bijective arises for instance when one wishes to know if it's possible to construct an inverse for the function, and if so what type? Before we can talk about inverses, we need to discuss composition:

- **Definition 7.6:** If one has functions $f : A \to B$ and $g : B \to C$, then the **composition** $g \circ f : A \to C$ is defined as the map

$$g \circ f(a) = g\big(f(a)\big).$$

- **Definition 7.7:** Suppose we have a function $f : A \to B$.

  1. A map $g : B \to A$ is called a **left inverse** of f if
  $$g \circ f(x) = x.$$
  2. A map $g : B \to A$ is called a **right inverse** of f if
  $$f \circ g(x) = x.$$
  3. A map $g : B \to A$ is called an **inverse** of f if it is both a left inverse and a right inverse:
  $$g \circ f(x) = x \quad \text{and} \quad f \circ g(x) = x.$$
  If such an inverse g exists for f, then we write $g = f^{-1}$ and say that f is **invertible** (note that f is automatically the inverse of g as well). We emphasize that an inverse is automatically both a left inverse and a right inverse.

  Remark: A function may not have any of the inverses listed above, or only a left inverse or only a right inverse. The following theorem gives a convenient way to check when and which inverses exist:

- **Theorem 7.8:** Suppose we have a function $f : A \to B$.

  a) f has a left inverse if and only if f is injective.

b) f has a right inverse if and only if f is surjective

c) f has an inverse if and only if f is bijective.

Proof: Let's start with a). First suppose that f has a left inverse $g : B \to A$. Then

$$f(a_1) = f(a_2) \quad \Longrightarrow \quad g\big(f(a_1)\big) = g\big(f(a_2)\big) \quad \Longrightarrow \quad a_1 = a_2$$

and hence f is injective. Now assume that f is injective. We will construct a left inverse $g : B \to A$ as follows. Take any $b \in \text{range}\, f$ and let $a \in A$ be such that $b = f(a)$. Note that such an a is unique since f is injective. Set $g(b) = a$. For all other $b \notin \text{range}\, f$, set $g(b)$ to be anything. Then by definition, for any $a \in A$

$$g\big(f(a)\big) = \big(\text{unique element in } A \text{ that } f \text{ maps to } f(a)\big) = a.$$

Next let's prove b). First suppose that f has a right inverse $g : B \to A$. Take any $b \in B$. We need to show that there exists an $a \in A$ such that $b = f(a)$. Note that $a = g(b)$ works since

$$f(a) = f\big(g(b)\big) = b.$$

Now suppose that f is surjective. We will construct a right inverse $g : B \to A$ as follows. For any $b \in B$, let $a \in A$ be any element such that $f(a) = b$, which is possible since f is surjective. Set $g(b) = a$. Then by definition for any $b \in B$

$$f\big(g(b)\big) = f(\text{element that gets mapped to } b \text{ by } f) = b.$$

Finally let's prove c). If f has an inverse $g : B \to A$, then as noted in Definition 7.7 g is both a left inverse and a right inverse and hence by parts a) and b) f is both injective and surjective and hence bijective. Now suppose that f is bijective. Hence it is injective and so we can construct the map $g : B \to A$ as we did in the proof of a) to get a left inverse. It is also a right-inverse in this case since for any $b \in B$

$$f\big(g(b)\big) = f(\text{unique element in } A \text{ that } f \text{ maps to } b) = b.$$

■

- **Note 7.9:** We remark that if a function $f : A \to B$ is not surjective, then we can easily modify it to become surjective by considering the function $f : A \to \text{range}\, f$. Note that we use the same letter "f" for both despite the fact that they're technically different functions: we use context to differentiate between the two (typically this is not an issue).

  We remark that this is done all the time. For instance, we cannot construct an inverse to the exponential function $\exp : \mathbb{R} \to \mathbb{R}$ because it is not surjective. But if we consider $\exp : \mathbb{R} \to \{y \in \mathbb{R} : y > 0\}$, then this function becomes bijective and so we can construct an inverse for it. In this, case that inverse is called the "natural logarithm:" $\ln : \{y \in \mathbb{R} : y > 0\} \to \mathbb{R}$.

8 Fields

- We now discuss fields, which are generalizations of usual arithmetic that you know with real numbers.
- **Definition 8.1:** A **field** is a set F paired with two operations "$+$" and "$\cdot$", called **addition** and **multiplication** respectively, that satisfy the following properties: for any elements $a, b, c \in F$
    1. (Commutativity) $a + b = b + c$ and $a \cdot b = b \cdot a$.
    2. (Associativity) $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
    3. (Identities) There exists distinct (i.e. different) elements denoted 0 and 1 (called **zero** and **one** respectively) in F satisfying that
    $$0 + a = a \quad \text{and} \quad 1 \cdot a = a.$$
    Both 0 and 1 are called **identity elements**.
    4. (Inverses) For any $a \in F$ and any nonzero $b \in F$ (i.e. $b \neq 0$) there exists an $\tilde{a} \in F$ and a $\tilde{b} \in F$ such that
    $$a + \tilde{a} = 0 \quad \text{and} \quad b \cdot \tilde{b} = 1.$$
    Here $\tilde{a}$ is called an **additive inverse** of a and $\tilde{b}$ is called a **multiplicative inverse** of b.
    5. (Distributivity) $a \cdot (b + c) = a \cdot b + a \cdot c$
- **Remark 8.2:** A few things to keep in mind:
    - A technical point: both "$+$" and "$\cdot$" are what's called **operations**, which are maps of the form $f : F \times F \to F$ (don't worry about the notation for now: it means that it takes two elements in F and outputs an element in F). When constructing examples of fields, you need to make sure that your output is an element in F: this phenomenon is called the "operations '$+$' and '$\cdot$' are **closed**."

      An illustrative example of when this fails is the set {irrational numbers} with the usual "$+$" and "$\cdot$" because for instance $\sqrt{2}$ is irrational but $\sqrt{2} \cdot \sqrt{2} = 2$ which is not irrational!
    - We often just write ab instead of $a \cdot b$.
- The reason for studying fields is that a wide variety of sets with operations that appear in math happen to be fields and so we study their properties together by simply studying fields themselves. Fields also happen to be the natural multipliers in linear algebra, as we'll see later.
- **Example 8.3:** The sets $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{C}$ with the usual operations "$+$" and "$\cdot$" are fields.
- **Example 8.4:** The set $\mathbb{Z}_5 = \{0,1,2,3,4\}$ where the operations "$+$" and "$\cdot$" work as follows
$$a + b = \text{remainder}\big((a + b) \div 5\big) \quad (\text{e.g. } 3 + 4 = \text{remainder}(7 \div 5) = 2)$$

$$ab = \text{remainder}\big((ab) \div 5\big) \quad (\text{e.g. } 2 \cdot 4 = \text{remainder}(8 \div 5) = 3)$$

With some effort, one can show that $\mathbb{Z}_5$ is a field. This example can be generalized to $\mathbb{Z}_k = \{0,1,2,\dots,k-1\}$ and a theorem from number theory says that $\mathbb{Z}_k$ is a field if and only if k is prime.

- Fields satisfy many of the natural arithmetic properties that you already know from the real numbers. For instance, the following theorem says that you can cancel things from both sides of an equation.

- **Theorem 8.5 (Cancelation Laws):** Suppose that F is a field. For any elements $a, b, c \in F$,

  1. If $a + b = a + c$ then $b = c$.
  2. If $a \cdot b = a \cdot c$ and $a \neq 0$, then $b = c$.

  Proof: We only prove 1) since 2) is proved very similarly and thus is left as an exercise. Suppose $a + b = a + c$. Let $\tilde{a}$ be an additive inverse for a. Then adding $\tilde{a}$ to both sides on the left gives

$$\tilde{a} + (a + b) = \tilde{a} + (a + c)$$
$$(\tilde{a} + a) + b = (\tilde{a} + a) + c$$
$$0 + b = 0 + c$$
$$b = c.$$

  ■

- **Theorem 8.6:** For a field F, the elements 0 and 1 are unique. The inverse $\tilde{a}$ and $\tilde{b}$ in Definition 8.1 Part 4) are also unique.

  Proof: This quickly follows from the cancelation law: you can read the proof in Appendix C in the book. ■

- **Notation 8.7:** Due to the uniqueness of $\tilde{a}$ and $\tilde{b}$ in Definition 8.1 Part 4) proved in Theorem 8.6, people write $-a$ and b^{-1} for them instead respectively. Thus the equations in Definition 8.1 Part 4) become

$$a + (-a) = 0 \quad \text{and} \quad b \cdot b^{-1}.$$

- **Definition 8.8:** For a field F, we define the operations **subtraction** "$-$" and **division** "$/$" as

$$a - b = a + (-b) \quad \text{and} \quad \frac{a}{b} = a \cdot b^{-1}.$$

- **Theorem 8.9:** Suppose that F is a field. For any elements $a, b \in F$,

  a. $a \cdot 0 = 0$
  b. $-1 \cdot a = -a$

c. 0 does not have a multiplicative inverse (i.e. 0^{-1} does not exist).

Proof: The proof of a) can be found in Appendix C in the book and c) follows immediately from a) (hint: try contradiction). So we will simply prove b). We have that

$$-1+1=0 \implies (-1+1)\cdot a = 0\cdot a = 0 \implies (-1)\cdot a + 1\cdot a = 0$$
$$\implies (-1)\cdot a + a = 0.$$

Hence $(-1)\cdot a$ is an additive inverse of a. Since by Theorem 8.6 the additive inverse is unique, $(-1)\cdot a = -a$.

■

9 Vector Spaces

9.1 Definition

- Next we discuss "vector spaces" which generalize the concept of vectors in Euclidean space that you worked with in calculus. This is analogous to how "fields" generalized the arithmetic of real numbers $\mathbb{R}$.
- **Definition 9.1:** A **vector space** V over a field F is a set that comes with an addition operation:

$$x+y \quad \text{where} \quad x,y\in V$$

(x and y are called **vectors**) and a scalar multiplication operation:

$$ax \quad \text{where} \quad a\in F \quad x\in V$$

(the a is called a **scalar**) that satisfies the following properties: for any $a,b\in F$ and any $x,y,z\in V$

1. (Commutativity) $x+y=y+x$ for any $x,y\in V$,
2. (Associativity) $(x+y)+z=x+(y+z)$ for any $x,y,z\in V$,
3. There exists an element denoted $0\in V$ such that $x+0=x$ (for all $x\in V$),
4. For any $x\in V$ there is an $\tilde{x}\in V$ such that $x+\tilde{x}=0$,
5. $1x=x$,
6. $(ab)x=a(bx)$,
7. $a(x+y)=ax+ay$,
8. $(a+b)x=ax+bx$.

- **Remark 9.2:** In the above definition, the "+" operation for V is different from the "+" operation for F, though we use the same symbol. Similarly, the $0 \in V$ and $0 \in F$ are also different though we use the same symbol (the former is a vector and the latter is a scalar).

- **Example 9.3:** The set of three-dimensional vectors $\mathbb{R}^3$ that you know from calculus is a vector space over $\mathbb{R}$. In this case $V = \mathbb{R}^3$ and $F = \mathbb{R}$ where addition and scalar multiplication is defined component wise as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} \quad \text{and} \quad a\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \\ ax_3 \end{pmatrix}.$$

  The rigorous definition of $\mathbb{R}^3$ is the set of 3-tuples of real numbers:

$$\mathbb{R}^3 = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \mathbb{R}\}.$$

  This example can of course be generalized to $V = \mathbb{R}^n$ and $F = \mathbb{R}$ where $\mathbb{R}^n$ is the set of n-tuples of real numbers:

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R}\}.$$

- **Example 9.4:** More examples include the following where addition and scalar multiplication is defined in the classic way from calculus:
  - $V = \mathbb{C}^n = \{(z_1, \dots, z_n) : z_1, \dots, z_n \in \mathbb{C}\}$ and $F = \mathbb{C}$.
  - $V = \mathbb{C}^n$ and $F = \mathbb{R}$.
  - Take any field F. Then F^n is a vector space over F.
  - $V = \{\text{functions } f : \mathbb{R} \to \mathbb{R}\}$ and $F = \mathbb{R}$.
  - $V = P_n(F) = \{a_n x^n + \cdots + a_1 x + a_0 : \text{each } a_i \in F\}$ where each x^k is defined as a separate object and we add these and perform scalar multiplication coefficient wise.
  - $V = P(F) = \{a_n x^n + \cdots + a_1 x + a_0 : n \geq 0 \text{ is an integer and each } a_i \in F\}$ (i.e. no restriction on the size of n).

- **Example 9.5:** The set of all 2×3 (read "2 by 3") real matrices over $\mathbb{R}$:

$$V = \left\{ \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} : a_{1,1}, \dots, a_{2,3} \in \mathbb{R} \right\} \quad \text{and} \quad F = \mathbb{R}.$$

  In this case addition and scalar multiplication is defined component wise as in Example 9.3. This can of course be generalized to $m \times n$ matrices over any field F:

$$V = \left\{ m\overbrace{\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}}^{n} : a_{1,1}, \dots, a_{m,n} \in F \right\}.$$

- **Remark 9.6:** Since the examples in Example 9.3, Example 9.4, and Example 9.5 are so famous, we often don't write the F involved in each. But you should on the homework!
- **Nonexample 9.7:** Some items that are not vector spaces include
  - Setting $V = \mathbb{R}^n$ and $F = \mathbb{C}$ is not a vector space.
  - Setting $V = \{x \in \mathbb{R}^3 : \text{Length of } x \leq 1\}$ and $F = \mathbb{R}$ is not a vector space.
- Many, but not all, of the properties of fields that we discussed earlier have analogs for vector spaces with similar proofs. We list them here:
- **Theorem 9.8:** Suppose we have a vector space V over a field F. For any $x, y, z \in V$

  a) (Cancellation Law) If $x + y = x + z$ then $y = z$,

  b) The vector 0 is unique,

  c) For any $x \in V$ its additive inverse $\tilde{x}$ is unique. For this reason, we write "$-x$" for the additive inverse rather than "$\tilde{x}$,"

  d) $0x = 0$,

  e) $(-1)x = -x$.

  Proof: Left as an exercise. ■

9.2 Subspaces

- Subspaces are subsets of vector spaces that are also vector spaces themselves. As we'll prove later, for intuition you may use the fact that in Euclidean space $\mathbb{R}^n$ subspaces are planes and lines that pass through zero. Here is the rigorous definition:
- **Definition 9.9:** Suppose that V is a vector space over a field F. A subset W of V (i.e. $W \subseteq V$) is called a **subspace** of V if it is a vector space over F with respect to the addition and scalar multiplication operations of V.
- As the definition states, to prove that something is a subspace you need to show that it is also a vector space, or in other words that it satisfies properties 1) - 8) of the definition of a vector space (Definition 9.1). That may seem discouraging since it sounds like a lot of work. However, fortunately that is not the case since the following theorem guarantees that you only have to check three properties:
- **Theorem 9.10:** Suppose that V is a vector space over a field F. A subset $W \subseteq V$ is a subspace if and only if

  1. $0 \in W$,
  2. (Closed under addition) If $x, y \in W$ then $(x + y) \in W$,
  3. (Closed under scalar multiplication) If $a \in F$ and $x \in W$, then $ax \in W$.

If you prefer, the last two can be written as one condition:

$$\text{“If } a, b \in F \text{ and } x, y \in w, \text{ then } (ax + by) \in W.\text{”}$$

Proof: First suppose that W is a subspace. Then conditions 2) and 3) above obviously need to hold. Condition 1) holds since if you take the zero vector of W and multiply it by (scalar) zero on the left you will get the zero vector of V which must be in W by condition 3) (i.e. the zero of vector of W and V are the same!)

Now suppose that conditions 2) and 3) above hold. Parts 1), 2), 3), 5), 6), 7), 8) of Definition 9.1 (i.e. definition of vector spaces) are automatically satisfied by W because these are properties inherited from V. So the only thing we need to check is that for every $x \in W$ there exists an $\tilde{x} \in W$ such that $x + \tilde{x} = 0$ (i.e. Part 4 of Definition 9.1). Fix any $x \in W$. By condition 3) above, $(-1)x = -x$ is in W and $x + (-x) = 0$. Hence $\tilde{x} = -x$ is what we wanted (i.e. the additive inverse of $x \in W$ in W and V are the same!).

■

- **Example 9.11:** Consider the vector space $\mathbb{R}^3$ (over the field $\mathbb{R}$). The subset $W \subseteq \mathbb{R}^3$ given by

$$W = \{(x, y, z) \in \mathbb{R}^3 : x + 2y - 3z = 0\}$$

is a subspace of $\mathbb{R}^3$. To prove this, we check the three conditions in Theorem 9.10:

1. $0 = (0,0,0)$ is indeed in W since $0 + 2(0) - 3(0) = 0$.
2. If $(x, y, z), (\tilde{x}, \tilde{y}, \tilde{z}) \in W$, then their sum $(x + \tilde{x}, y + \tilde{y}, z + \tilde{z}) \in W$ since

$$(x + \tilde{x}) + 2(y + \tilde{y}) - 3(z + \tilde{z}) = (x + 2y - 3z) + (\tilde{x} + 2\tilde{y} - 3\tilde{z}) = 0 + 0 = 0.$$

3. If $a \in \mathbb{R}$ and $(x, y, z) \in W$, then their scalar multiplication $a(x, y, z) = (ax, ay, az) \in W$ since

$$(ax) + 2(ay) - 3(az) = a(x + 2y - 3z) = a \cdot 0 = 0.$$

You may recall from calculus that in fact $x + 2y - 3z = 0$ is an equation for the plane passing through zero with normal vector $(1,2,-3)$, so this seems to confirm the intuition that planes are subspaces!

- Considering subspaces is very useful since by passing to a subspace you study a vector space that preserves the old properties and introduces additional structure. The next example illustrates this:

- **Example 9.12:** Let $V = \{\text{functions } f : \mathbb{R} \to \mathbb{R}\}$ (with $F = \mathbb{R}$). We can think of $P_2(\mathbb{R})$ of all real polynomials of degree less than or equal to 2:

$$P_2(\mathbb{R}) = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

as a subspace of V. To prove this, we check the three conditions in Theorem 9.10:

1. The constantly zero function 0 is in $P_2(\mathbb{R})$ since it is equal to the polynomial $0x^2 + 0x + 0$.

2. If ax^2+bx+c and $\tilde{a}x^2+\tilde{b}x+\tilde{c}$ are in $P_2(\mathbb{R})$, then so is their sum:

$$(ax^2+bx+c)+\left(\tilde{a}x^2+\tilde{b}x+\tilde{c}\right)=(a+\tilde{a})x^2+\left(b+\tilde{b}\right)x+(c+\tilde{c})\in P_2(x).$$

3. We leave it to the reader to check that $P_2(\mathbb{R})$ is closed under scalar multiplication.

- **Note 9.13:** In this course we will only consider the alternative formulation of $P_n(\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C})$ or $P(\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C})$ as subsets of $\{$functions $f:\mathbb{R}$ or $\mathbb{Q}$ or $\mathbb{C}\to\mathbb{R}$ or $\mathbb{Q}$ or $\mathbb{C}\}$ (when the $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{C}$ all match) and not for general feilds F. The reason for doing so is the following result and observation right after:

- **Theorem 9.14:** Suppose that $p,q\in P_n(\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C})$ are polynomials that are also equal:

$$p(x)=a_nx^n+\cdots+a_1x+a_0=b_nx^n+\cdots+b_1x+b_0.$$

Then each $a_i=b_i$.

Proof: We omit the proof since it lies in analysis. ■

- **Note 9.15:** The reason for Note 9.13 is that the above theorem doesn't work for all fields. For instance, the following polynomials are equal in $\mathbb{Z}_3$ (check this!) but notice that the coefficients are not equal:

$$x^3+0x^2+2x+0=0x^3+0x^2+0x+0 \quad \text{in } \mathbb{Z}_3.$$

- **Theorem 9.16:** Suppose that V is a vector space over a field F and that $W_1,W_2\subseteq V$ are subspaces of V. Then their intersection $W_1\cap W_2$ is also a subspace of V.

Proof: This is a simple exercise in checking that the three conditions in Theorem 9.10 hold for $W_1\cap W_2$.

■

9.3 Linear Combinations

- One way to form or study subspace is using the notion of linear combinations. We start by defining the latter:

- **Definition 9.17:** Suppose that V is a vector space (over a field F) and that $v_1,\dots,v_m\in V$ is a list of vectors in V. A **linear combination** of $v_1,\dots,v_m$ is a finite sum of the form

$$a_1v_1+\cdots+a_mv_m \tag{9.18}$$

where $a_1,\dots,a_m\in F$. The a_i's here are called **coefficients** of the linear combination.

- **Example 9.19:** Consider the vector space $\mathbb{R}^3$ (over the field $\mathbb{R}$). We claim that the vector

$$w=\begin{pmatrix}3\\6\\4\end{pmatrix}$$

is a linear combination of the vectors

$$v_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad v_2 = \begin{pmatrix}3\\2\\0\end{pmatrix}, \quad v_3 = \begin{pmatrix}0\\1\\2\end{pmatrix}.$$

To prove this, we need to show that there exists $a, b, c \in \mathbb{R}$ such that

$$\begin{pmatrix}3\\6\\4\end{pmatrix} = a\begin{pmatrix}1\\0\\0\end{pmatrix} + b\begin{pmatrix}3\\2\\0\end{pmatrix} + c\begin{pmatrix}0\\1\\2\end{pmatrix}.$$

Combining the right-hand side gives

$$\begin{pmatrix}3\\6\\4\end{pmatrix} = \begin{pmatrix}a+3b\\2b+c\\2c\end{pmatrix}.$$

Equating the components gives

$$\begin{aligned} a \quad &+ 3b &&= 3\\ 2b \quad &+ c &&= 6\\ &2c &&= 4 \end{aligned}$$

The third equation gives that $c = 2$. Plugging this into the second equation and solving for b gives $b = 2$. Plugging this into the first equation and solving for a gives $a = -3$. Hence

$$\begin{pmatrix}3\\6\\4\end{pmatrix} = (-3)\begin{pmatrix}1\\0\\0\end{pmatrix} + 2\begin{pmatrix}3\\2\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\\2\end{pmatrix}.$$

Example 9.20: Consider the vector space $P_4(\mathbb{R})$ (over the field $\mathbb{R}$). Is the polynomial $2x^4 + 9x^3 + 5x^2 + 11x$ in the span of $x^4 + 3x^3$, $x^3 + 2x$, and $x^2 + x$? Not obvious, right? Well, let us leverage the power of linear algebra! The answer will be yes if we can find coefficients a, b, c such that

$$2x^4 + 9x^3 + 5x^2 + 11x = a(x^4 + 3x^3) + b(x^3 + 2x) + c(x^2 + x)$$

Distributing on the right-hand side gives

$$2x^4 + 9x^3 + 5x^2 + 11x = ax^4 + (3a + b)x^3 + cx^2 + (2b + c)x$$

This equality will be true if the coefficients of the x^k's on both sides are the same. Equating the coefficients of the x^k's gives the system of equations

$$\begin{aligned} a \quad &\quad &&= 2\\ 3a \quad &+b &&= 9\\ &\quad c &&= 5\\ &2b \quad +c &&= 11 \end{aligned}$$

The first and third equation give $a = 2$ and $c = 5$. Solving for b in the second equation gives $b = 3$. We check that the fourth equation makes sense (important!): indeed $2(3) + 5 = 11$. Hence indeed

$$2x^4 + 9x^3 + 5x^2 + 11x = 2(x^4 + 3x^3) + 3(x^3 + 2x) + 5(x^2 + x).$$

- **Definition 9.21:** Suppose S is a nonempty subset of a vector space V (over a field F): $S \subseteq V$. The **span** of S is the set of all (finite) linear combinations of vectors in S:

$$\operatorname{span} S = \{a_1 v_1 + \cdots + a_m v_m : m \in \mathbb{Z}_+, \text{ each } a_k \in F, \text{ each } v_k \in S\}.$$

  By convention we define $\operatorname{span}\{\emptyset\} = \{0\}$.

- **Example 9.22:** We proved in Example 9.19 that

$$\begin{pmatrix} 3 \\ 6 \\ 4 \end{pmatrix} \in \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}\right\}.$$

- **Example 9.23:** Consider the vector space of functions $V = \{f : \mathbb{R} \to \mathbb{R}\}$ (over $\mathbb{R}$). Then setting $S = \{1, x, x^2, x^3, x^4, x^5\}$

$$P_5(\mathbb{R}) = \operatorname{span} S = \operatorname{span}\{1, x, x^2, x^3, x^4, x^5\}.$$

- **Example 9.24:** At the moment we can only discuss the following intuitively: consider the vector space $\mathbb{R}^3$ and fix any two vectors v_1 and v_2 pointing in two different directions. Let $S = \{v_1, v_2\}$. Recall from calculus that vector addition of the form $v_1 + v_2$ in $\mathbb{R}^3$ can be visualized as putting the base of v_2 at the arrow tip of v_1 and see where the arrow tip of v_2 lands. The same can be done for rescaled versions of v_1 and v_2: $a_1 v_1 + a_2 v_2$, or in other words for linear combinations of v_1 and v_2. With this visualization, you should convince yourself that the set of all points that you can obtain by considering $a_1 v_1 + a_2 v_2$ is the plane passing through $0 = (0,0,0)$, v_1, and v_2. In other words:

$$\operatorname{span} S = \operatorname{span}\{v_1, v_2\} = \text{The plane passing through } 0, v_1, \text{ and } v_2.$$

  This is not a rigorous proof (so you can't use it on the homework yet); we will prove this rigorously later!

- The last two examples seem to indicate that the span of any set is a subspace since recall that $P_5(\mathbb{R})$ is a subspace of $V = \{f : \mathbb{R} \to \mathbb{R}\}$ and planes passing through zero are subspaces of $\mathbb{R}^3$ (we didn't prove the latter rigorously yet). This turns out to be true in general:

- **Theorem 9.25:** Suppose S is a subset of a vector space V (over a field F): $S \subseteq V$.

  a) Then $\operatorname{span} S$ is a subspace of V.

  b) Any subspace $W \subseteq V$ of V that contains S will also contain (the subspace) $\operatorname{span} S$ (i.e. $\operatorname{span} S$ is the smallest subspace of V that contains S).

  Proof: If $S = \emptyset$, then $\operatorname{span} S = \{0\}$ which is a vector space and so both a) and b) are immediately true. So suppose that $S \neq \emptyset$. First we prove a). To show that $\operatorname{span} S$ is a subspace of V, we just need to check the three conditions in Theorem 9.10:

1. First we check that $0 \in \operatorname{span} S$. Take any $v_1 \in S$. Note that $0v_1$ is technically a linear combination of v_1 (i.e. set $a_1 = 0$ in (9.18)) and so $0 = 0v_1 \in \operatorname{span} S$.
2. Take any two $w_1, w_2 \in \operatorname{span} S$. Both of them are linear combinations by definition:

$$w_1 = a_1 v_1 + \cdots + a_m v_m, \quad \text{each } a_k \in F, v_k \in S,$$
$$w_2 = \tilde{a}_1 \tilde{v}_1 + \cdots + \tilde{a}_n \tilde{v}_n, \quad \text{each } \tilde{a}_k \in F, \tilde{v}_k \in S.$$

   Thus their sum $w_1 + w_2$ is also a linear combination:

$$w_1 + w_2 = a_1 v_1 + \cdots + a_m v_m + \tilde{a}_1 \tilde{v}_1 + \cdots + \tilde{a}_n \tilde{v}_n.$$

   So indeed $(w_1 + w_2) \in \operatorname{span} S$.
3. We leave it as an exercise to check that if $a \in F$ and $w \in \operatorname{span} S$, then $aw \in \operatorname{span} S$: this is done very similarly to the previous step (i.e. step 2).

Next let's prove b). Suppose $W \subseteq V$ is a subspace that contains S. We want to show that W contains $\operatorname{span} S$. Since $\operatorname{span} S$ is the set of all linear combinations of vectors in S, we have to show that any linear combination of the form

$$a_1 v_1 + \cdots + a_m v_m, \quad \text{each } a_k \in F, v_k \in S.$$

is in W. By assumption each $v_k \in W$, hence each term $a_k v_k \in W$ by Part 3) of Theorem 9.10, and hence their sum $(a_1 v_1 + \cdots + a_m v_m) \in W$ by Part 2) of Theorem 9.10. So indeed W contains $\operatorname{span} S$.

■

- **Definition 9.26:** A subset S of a vector space V (over a field F) is said to **span** (or **generate**) V if $\operatorname{span} S = V$.
- **Example 9.27:** We have that
  - $S = \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$ spans/generates $\mathbb{R}^2$
  - $S = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$ spans/generates the x, y plane in $\mathbb{R}^3$.
  - $S = \{1, x, x^2\}$ spans/generates $P_2(\mathbb{R})$
  - (Exercise) $\{1 + x, x + 5x^3, x^2 + 3x, x^4 + 2x\}$ spans/generates $P_4(\mathbb{R})$.

9.4 Linear Dependence/Independence

- Vector spaces can be very abstract and hence typically don't lend themselves nicely to computation in their originally given form. So we need a way to represent them in a convenient fashion. In this section we build towards this idea, starting with the following fundamental concept:

- **Definition 9.28:** Suppose that V is a vector space (over a field F). A set of vectors $S \subseteq V$ is called **linearly dependent** if one can find a finite number of distinct vectors $v_1, \dots, v_m \in S$ and scalars $a_1, \dots, a_m \in F$ such that at least one a_i is nonzero and

$$a_1 v_1 + \cdots + a_m v_m = 0. \tag{9.29}$$

- The idea behind the above definition is the following: suppose that (9.29) holds and suppose for example that $a_m \neq 0$. Then one could solve for v_m in terms of the other vectors to get

$$v_m = a_m^{-1} a_1 v_1 + \cdots + a_m^{-1} a_{m-1} v_{m-1}. \tag{9.30}$$

In other words, v_m depends on the other vectors $v_1, \dots, v_{m-1}$ via the linear combination (9.30).

- **Example 9.31:** Consider the vector space $V = \mathbb{R}^3$ (over the field $\mathbb{R}$) and consider the set of vectors

$$S = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix} \right\}.$$

Is this set S linearly dependent? Since S is finite, we can consider the question of if (9.29) is possible to arrange using all vectors of S with at least one a_i not zero. In other words, S will be linearly dependent if and only if there exist scalars $a, b, c \in \mathbb{R}$ at least one of which is nonzero such that

$$a \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + b \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix} + c \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Combing the left-hand side into one vector and then equating the components of both sides of this equation gives

$$\begin{array}{rrrl} a & +2b & +4c & = 0 \\ 2a & & +4c & = 0 \\ & 3b & +3c & = 0 \end{array}$$

Notice that the second and third equations are equivalent to $a = -2c$ and $b = -c$. Plugging these into the first equation gives $(-2c) + 2(-c) + 4c = 0$, which is true regardless of the value of c! So we can choose any value of c, say $c = 1$, and get that $a = -2$ (i.e. the second equation is satisfied) and $b = -1$ (i.e. the third equation is satisfied) and have that the first equation is satisfied automatically. Hence this is a solution:

$$-2 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix} + 1 \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and so S is indeed linearly dependent. Notice, for instance, that this means that the third vector can be written as a linear combination of the first two:

$$\begin{pmatrix}4\\4\\3\end{pmatrix} = 2\begin{pmatrix}1\\2\\0\end{pmatrix} + \begin{pmatrix}2\\0\\3\end{pmatrix}.$$

- **Definition 9.32:** Suppose that V is a vector space (over a field F). A set of vectors $S \subseteq V$ is called **linearly independent** if it is not linearly dependent. A little bit of thought should convince you that this is equivalent to saying that if you take a linear combination

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

of distinct vectors $v_1, \ldots, v_m \in S$, the coefficients $a_1, \ldots, a_m$ must all be zero. By convention we declare the empty set $\emptyset$ to be linearly independent.

- **Example 9.33:** A set of one nonzero vector: $S = \{v\}$, is always linearly independent. Indeed if $a_1 v_1 = 0$ then $a_1 = 0$.

- **Example 9.34:** Consider the vector space $V = P_5(\mathbb{Q})$ (over the field $\mathbb{Q}$) and consider the following set of polynomials

$$S = \{3x^2 + 2,\ \ x^2 + x,\ \ 10x^2 + 2x + 6\}.$$

Is this set linearly independent? Suppose that

$$a(3x^2 + 2) + b(x^2 + x) + c(10x^2 + 2x + 6) = 0.$$

If we can show that this implies that a, b, c are all zero, then S is indeed linearly independent. Combining terms on the left-hand side and writing $0 = 0x^2 + 0x + 0$ gives

$$(3a + b + 10c)x^2 + (b + 2c)x + (2a + 6c) = 0x^2 + 0x + 0.$$

By Theorem 9.14 this is true of and only if the coefficients of the powers of x's are the same on both sides, and so equating coefficients this way gives

$$\begin{array}{rrrl} 3a & +b & +10c & = 0 \\ & b & +2c & = 0 \\ 2a & & +6c & = 0 \end{array}$$

This second and third equations are equivalent to $b = -2c$ and $a = -3c$. Plugging this into the first equation gives $3(-3c) + (-2c) + 10c = 0$ which is equivalent to $-c = 0$, or in other words $c = 0$. Hence a and b must also be zero. Thus all a, b, c are zero and so S is indeed linearly independent.

- We illustrate the relation between span and linear (in)dependence with a few results.

- **Theorem 9.35:** Suppose that S is a linearly dependent subset of a vector space V (over a field F). Then we can always remove a vector v from S to get a new set $\tilde{S}$ such that $\operatorname{span}\tilde{S} = \operatorname{span} S$.

Remark: In other words, linear dependent lists always have vectors that don't contribute any information to the span.

Proof: Since S is linearly dependent, $S \neq \emptyset$ (or else it would be linearly independent). The fact that S is linearly dependent means that there exist distinct vectors $v_1, \dots, v_m \in S$ and $a_1, \dots, a_m \in F$ such that at least one a_i is not zero and

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

Rearranging the v_i's if necessary, we can assume without loss of generality that $a_m \neq 0$. Let's call $v = v_m$ and solve for v:

$$v = -a_m^{-1} a_1 v_1 - \cdots - a_m^{-1} a_{m-1} v_{m-1}. \tag{9.36}$$

Let $\tilde{S}$ be the subset obtained by taking S and removing v from it (set theoretically this is written as $\tilde{S} = S \setminus \{v\}$). We want to show that $\operatorname{span} \tilde{S} = \operatorname{span} S$. The inclusion $\operatorname{span} \tilde{S} \subseteq \operatorname{span} S$ holds because anything that can be written as a linear combination of vectors in $\tilde{S}$ is automatically a linear combination of vectors in S simply because $\tilde{S} \subseteq S$. So let us show that $\operatorname{span} \tilde{S} \supseteq \operatorname{span} S$. Take any vector $w \in \operatorname{span} S$, which means that

$$w = b_1 w_1 + \cdots + b_n w_n \quad \text{for some } w_1, \dots, w_n \in S, \;\; a_1, \dots, a_n \in F \tag{9.37}$$

By combining like terms, we can assume that the w_i's are distinct. If none of the w_i's are equal to v, then they are all in $\tilde{S}$ and so this is a linear combination of vectors only in $\tilde{S}$ and so $w \in \operatorname{span} \tilde{S}$. Now suppose that one of the w_i's is equal to v. Without loss of generality, suppose that $w_n = v$. Then plugging (9.36) into (9.37) gives

$$w = b_1 w_1 + \cdots b_{m-1} w_{m-1} + b_n(-a_m^{-1} a_1 v_1 - \cdots - a_m^{-1} a_{m-1} v_{m-1})$$

which if you distribute the last term shows that this is a linear combination of vectors only in $\tilde{S}$ and so $w \in \operatorname{span} \tilde{S}$ in this case as well. Thus $\operatorname{span} \tilde{S} \supseteq \operatorname{span} S$ and so indeed $\operatorname{span} \tilde{S} = \operatorname{span} S$.

■

- **Corollary 9.38:** Suppose that V is a vector space (over a field F) and that $S = \{v_1, \dots, v_n\}$ is a finite linearly dependent subset. Then there exists a linearly independent subset $\tilde{S} \subseteq S$ with the same span: $\operatorname{span} \tilde{S} = \operatorname{span} S$.

  Proof: Since S is linearly dependent, $S \neq \emptyset$. Now use the previous theorem to remove vectors one by one until you arrive at a linearly independent subset $\tilde{S}$. You know that this process will end eventually because you can't do this process more than n times since if you did it n times you'll arrive at $\tilde{S} = \emptyset$, which we declared to be linearly independent. ■

- **Theorem 9.39:** Suppose that V is a vector space (over a field F) and that $S \subseteq V$ is a linearly independent set. Let $v \in V$ be a nonzero vector not in S. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \operatorname{span} S$.

  Proof: First suppose that $S \cup \{v\}$ is linearly dependent. Then one can find distinct vectors $v_1, \dots, v_m \in S \cup \{v\}$ and scalars $a_1, \dots, a_m$ such that at least one a_i is nonzero so that

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

We note that one of the v_i's here is equal to v and furthermore that the corresponding coefficient $a_i \neq 0$ or else the above would be a linear combination of vectors only in S with at least one coefficient not zero, which would imply that S is linearly dependent while we said that S is linearly independent. By rearranging if the v_i's if necessary, we can assume that $v_m = v$ and so

$$a_1 v_1 + \cdots + a_{m-1} v_{m-1} + a_m v_m = 0 \quad \text{with} \quad a_m \neq 0$$
$$\implies \quad v_m = a_m^{-1} a_1 v_1 + \cdots + a_m^{-1} a_{m-1} v_{m-1}.$$

In other words, v is a linear combination of vectors in S and so $v \in \operatorname{span} S$.

Now suppose that $v \in \operatorname{span} S$. That means we can write

$$v = a_1 v_1 + \cdots + a_m v_m$$

for some vectors $v_1, \ldots, v_m \in S$ and scalars $a_1, \ldots, a_m \in F$. Thus

$$a_1 v_1 + \cdots + a_m v_m + (-1)v = 0.$$

This shows that $S \cup \{v\}$ is linearly dependent.

■

9.5 Bases and Dimension

- **Definition 9.40:** Suppose that V is a vector space (over a field F). A subset $\beta \subseteq V$ is called a **basis** if it is linearly independent and generates V (i.e. $V = \operatorname{span} \beta$).

  Remark: Bases don't have to be unique: a vector space V typically has many bases.

- **Example 9.41:** Since $\emptyset$ is linearly independent and $\operatorname{span} \emptyset = \{0\}$, $\beta = \emptyset$ is a basis for $\{0\}$.

- **Example 9.42:** The set

$$\beta = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

  is a basis for $\mathbb{R}^3$. We'll let you check that it is linearly independent. It spans all of $\mathbb{R}^3$ because any vector in $\mathbb{R}^3$ can be written as a linearly combination of the above three vectors as so:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

  More generally,

$$\beta = \left\{ \underbrace{\begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\\vdots\\0\end{pmatrix}, \dots, \begin{pmatrix}0\\0\\0\\\vdots\\1\end{pmatrix}}_{n \text{ vectors}} \right\}$$

is a basis for $\mathbb{R}^n$ (and even F^n where F is a field). These bases are so famous that they are referred to as the **standard basis** for $\mathbb{R}^n$ (and even F^n).

- **Example 9.43:** The set $\beta = \{1, x, x^2, \dots, x^n\}$ is a basis for $P_n(\mathbb{R})$. We'll let you check the details. This particular basis is in fact called the **standard basis** of $P_n(\mathbb{R})$.
- **Example 9.44:** The set $\beta = \{x^k : k \geq \text{is an integer}\} = \{1, x, x^2, x^3, \dots\}$ is a basis for $P(\mathbb{R})$ (the set of all polynomials: no restriction on the degree).
- Finite bases are nice because they allow us to nicely write down *every* vector in a vector space in a *unique* fashion. This will be critical for when we start doing computations over vector spaces later on.
- **Theorem 9.45:** Suppose that V is a vector space (over a field F). Then a finite subset $\beta = \{u_1, \dots, u_m\}$ of V is a basis of V if and only if every vector $v \in V$ can be written uniquely as a linear combination of $u_1, \dots, u_m$:

  $$v = a_1 u_1 + \cdots + a_m u_m$$

  (i.e. for every v only one set of $a_1, \dots, a_m \in F$ will work here).

  Proof: First suppose that β is a basis. Take any $v \in V$. Since by assumption $\operatorname{span} \beta = V$, we trivially have that $v \in \operatorname{span} \beta$. Thus β can be written as a linear combination of the u_i's:

  $$v = a_1 u_1 + \cdots + a_m u_m. \tag{9.46}$$

  We need to prove that this set of coefficients $a_1, \dots, a_m \in F$ is the only set of coefficients that makes this equation hold. We do this by supposing that $\tilde{a}_1, \dots, \tilde{a}_m \in F$ is another set of coefficients such that

  $$v = \tilde{a}_1 u_1 + \cdots + \tilde{a}_m u_m \tag{9.47}$$

  and showing that $a_i = \tilde{a}_i$. If we subtract (9.47) from (9.46) we get that

  $$0 = v - v = (a_1 - \tilde{a}_1)u_1 + \cdots + (a_m - \tilde{a}_m)u_m. \tag{9.48}$$

  Since $\beta = \{u_1, \dots, u_m\}$ is linearly independent, we have that each coefficient in (9.48) must be zero, or in other words each $a_i - \tilde{a}_i = 0$ and so indeed $a_i = \tilde{a}_i$.

  We leave the other direction of the theorem as an exercise: it should be very quick, approximately a two-line argument.

  ■

- **Theorem 9.49:** Suppose a vector space V (over a field F) is generated by a finite set S (i.e. $\operatorname{span} S = V$). Then V has a finite basis.

  Proof: If S is linearly independent, then we're done since in this case S is a basis for V. Now suppose that S is linearly dependent. Then by Corollary 9.38 we can find a (finite) subset $\beta \subseteq S$ that is linearly independent and $\operatorname{span}\beta = \operatorname{span} S = V$. In this case, β is the finite basis of V we're looking for.

  ■

- You may have heard that $\mathbb{R}^n$ is an n-dimensional space and in Example 9.42 we say that $\mathbb{R}^n$ has a basis with n vectors. Coincidence? No! Let's build towards this idea:

- **Definition 9.50:** We say that a vector space V is **finite dimensional** if it has a finite basis. If V does not have a finite basis, we call it **infinite dimensional**.

- **Definition 9.51:** Suppose that V is a finite dimensional vector space and let $\beta = \{u_1, \dots, u_m\}$ be any finite basis for V. We define the **dimension** of V as the number of vectors in β:

  $$\dim V = m.$$

  Warning: we have to prove that this is a well-defined definition. In other words, how do we know that there doesn't exist another finite basis $\beta = \{\tilde{u}_1, \dots, \tilde{u}_{\tilde{m}}\}$ for some different $\tilde{m}$, in which case we won't know if to set $\dim V = m$ or $\dim V = \tilde{m}$. We will prove soon that this cannot happen and so dimension of V is indeed well-defined.

- **Example 9.52:** Since $\emptyset$ is a basis for $\{0\}$, $\dim\{0\} = 0$.

- **Example 9.53:** In Example 9.42 we saw that $\mathbb{R}^n$ has a basis with n vectors and so $\dim \mathbb{R}^n = n$ (or more generally $\dim F^n = n$). Similarly, from Example 9.43 we see that $\dim P_n(\mathbb{R}) = n + 1$. The latter statement is also true for $P_n(\mathbb{C})$ and $P_n(\mathbb{Q})$.

- **Example 9.54:** We may ask you to prove in the homework that $P(\mathbb{R})$ and $V = \{f : \mathbb{R} \to \mathbb{R}\}$ are both infinite dimensional.

- To prove that our definition of dimension in Definition 9.51 is well defined, we make use of the following powerful theorem:

- **Theorem 9.55: (Replacement Theorem)** Suppose that V is a vector space (over a field F). Suppose that $G = \{w_1, \dots, w_n\}$ is a subset of V with n vectors that generates V (i.e. $V = \operatorname{span} G$). Suppose that $H = \{v_1, \dots, v_m\}$ is a linearly independent subset of V with m vectors. Then $m \leq n$ and there exists a subset L of G with $n - m$ vectors so that $H \cup L$ also generates V (i.e. $V = \operatorname{span}(H \cup L)$).

  Proof: We fix V, F, and G and prove the theorem by induction on the size of m (i.e. the number of vectors in H). The base case is $m = 0$. In this case there are no vectors in H and so $H = \emptyset$. Hence indeed $m \leq n$ and we can set $L = G$ since then L has $n - m$ vectors and $H \cup L = \emptyset \cup G = G$ generates V.

Now suppose that the theorem is true for m. We will show that it holds for $m+1$. Suppose that $H = \{v_1, \dots, v_{m+1}\}$ is linearly independent. We need to prove that $m+1 \leq n$ and that there exists a subset L of G with $n-(m+1) = n-m-1$ vectors so that $H \cup L$ generates V. To get started, let's remove one vector from H, say the last vector and call it $\widetilde{H}$:

$$\widetilde{H} = \{v_1, \dots, v_m\}.$$

Note that $\widetilde{H}$ is a subset of the linearly independent set H and hence $\widetilde{H}$ is also linearly independent. Thus by the inductive hypothesis there exists $\tilde{L} = \{u_1, \dots, u_{n-m}\} \subseteq G$ so that

$$\widetilde{H} \cup \tilde{L} = \{v_1, \dots, v_m, u_1, \dots, u_{n-m}\} \text{ generates } V. \tag{9.56}$$

Now we aim to put v_{m+1} back into this list. Since this list generates all of V,

$$v_{m+1} = a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_{n-m} u_{n-m} \tag{9.57}$$

for some $a_1, \dots, a_m, b_1, \dots, b_{n-m} \in F$. We cannot have all of the b_i's here be equal to zero or else this would imply that v_{m+1} could be written as a linear combination of the v_i's and hence imply that $H = \{v_1, \dots, v_{m+1}\}$ is linearly dependent while we assumed that it was linearly independent. Hence one of the b_i's is nonzero. Since $\widetilde{H} \cup \tilde{L}$ has n vectors (c.f. (9.56)), this already shows us that $m+1 \leq n$.

By rearranging the u_i's if necessary, we can assume without loss of generality that $b_1 \neq 0$ and hence we can solve for u_1 in (9.57) to get

$$u_1 = b_1^{-1} v_{m+1} - b_1^{-1} a_1 v_1 - \cdots - b_1^{-1} a_m v_m - b_1^{-1} b_2 u_2 - \cdots - b_1^{-1} b_{n-m} u_{n-m}. \tag{9.58}$$

This means that any linear combination of elements in $\widetilde{H} \cup \tilde{L}$:

$$c_1 v_1 + \cdots + c_m v_m + d_1 u_1 + \cdots + d_{n-m} u_{n-m} \tag{9.59}$$

can be rewritten as a linear combination of $v_1, \dots, v_{m+1}, u_2, \dots, u_{n-m}$ by plugging (9.58) into u_1 in (9.59). Thus

$$V = \operatorname{span}\left(\widetilde{H} \cup \tilde{L}\right) \subseteq \operatorname{span}\{v_1, \dots, v_{m+1}, \underbrace{u_2, \dots, u_{n-m}}_{\text{Call this } L}\} = V$$

(the last equality follows since you can't span anything bigger than V). Hence the above indicated L is the L we wanted.

■

- **Corollary 9.60:** Suppose that V is a vector space (over a field F) and that $\beta = \{u_1, \dots, u_n\}$ is a basis for V that has n vectors. Then all bases for V have exactly n vectors. Hence dimension in Definition 9.51 (for finite dimensional vector spaces) is indeed well defined.

  Proof: Take another basis $\tilde{\beta}$. First suppose that $\tilde{\beta} = \{v_1, \dots, v_m\}$ has a finite number of vectors. We want to show that $m = n$. Since $\tilde{\beta}$ is linearly independent and β generates V, the previous

theorem (with $H = \tilde{\beta}$ and $G = \beta$) says that $m \leq n$. Similarly, since β is linearly independent and $\tilde{\beta}$ generates V we have that $n \leq m$. So indeed $m = n$.

Now suppose that $\tilde{\beta}$ has an infinite number of vectors, we will show that this actually can't happen. Pick out any $n+1$ vectors $\{v_1, \dots, v_{n+1}\} \subseteq \tilde{\beta}$. Since $\tilde{\beta}$ is linearly independent, its subset $\{v_1, \dots, v_{n+1}\}$ is also linearly independent. Since β generates V, the previous theorem implies that $n+1 \leq n$, a contradiction!

■

- **Corollary 9.61:** Suppose that V is a vector space of dimension n (which implies that it's finite dimensional)

    a. Suppose that $S \subseteq V$ generates V. Then S contains at least n vectors. If furthermore S has exactly n vectors, then S is a basis for V.

    b. Suppose that $S \subseteq V$ is linearly independent and has n vectors. Then it is a basis.

    c. Suppose that $S \subseteq V$ is linearly independent. Then you can add vectors to S to turn it into a basis for V.

    Proof: We will prove parts a) and c) and leave part b) as an exercise (the proof is also written out on pages 48 - 49 of the book).

    We start with part a). Suppose that $S \subseteq V$ generates V (i.e. $V = \operatorname{span} S$). First we show that S has at least n vectors. Suppose not! Then S has less than n vectors. By Corollary 9.38 there exists a linearly independent subset $\tilde{S} \subseteq S$ such that $\operatorname{span} \tilde{S} = \operatorname{span} S = V$ and thus $\tilde{S}$ is a basis for V. But then $\tilde{S}$ is a basis with number of vectors less than n (i.e. the dimension of V) and hence we have contradiction! Next let's prove that if S has exactly n vectors, then it is a basis. Suppose not! Since S spans V, this means that S is not linearly dependent. Again by Corollary 9.38 there exists a linearly independent subset $\tilde{S} \subseteq S$ such that $\operatorname{span} \tilde{S} = \operatorname{span} S = V$ and thus $\tilde{S}$ is a basis for V with less vectors than n, a contradiction!

    Now let's prove c). Suppose that $S \subseteq V$ is linearly independent. Let $\beta = \{w_1, \dots, w_n\}$ be a basis for V. By the replacement theorem (with $G = \beta$ and $H = S$ in the statement) the number of vectors in S is smaller than β, in particular finite, and so we can write it as

    $$S = \{v_1, \dots, v_m\}$$

    with $m \leq n$. The replacement theorem further tells us that we can add a set of $(n-m)$ vectors $L = \{u_1, \dots, u_{n-m}\} \subseteq \beta$ to S to get a new set:

    $$S \cup L = \{v_1, \dots, v_m, u_1, \dots, u_{n-m}\}$$

    that generates V. Notice that $S \cup L$ has no more than n vectors (it could potentially have less if it has repeats). On the other hand, since $S \cup L$ generates V part a) says that $S \cup L$ must have at least n vectors. Hence $S \cup L$ has exactly n vectors, which by part a) means that it is a basis.

    ■

- **Proposition 9.62:** Suppose that W is a subspace of a finite-dimensional vector space V. Then W is also finite dimensional and $\dim W \leq \dim V$. Furthermore, if $\dim W = \dim V$, then $V = W$.

  Proof: Let $n = \dim V$. If $W = \emptyset$, then $\dim W = 0 \leq n$ and if $\dim W = \dim V$ then both $W = V = \emptyset$. Now suppose that $W \neq \emptyset$. Take any vector $v_1 \in W$ can consider the linearly independent list $\{v_1\}$. Now keep adding vectors in W to this list while still maintaining its linear independence until you get

  $$S = \{v_1, \dots, v_k\} \subseteq W$$

  after which you can't add more vectors to it while maintaining its linear independence. This has to happen eventually since if you hit $k = n$, then by Corollary 9.61 b) the above list will be a basis of V and hence any vector $v \in W$ (and hence $v \in V$) is in the span of the above list and hence adding it to the above list will make it linearly dependent by Theorem 9.39. So $k \leq n$.

  Next Theorem 9.39 implies that the fact that we can't add any more vectors to the above list while maintaining its linear independence means that every vector in W is in the span of the above list. Thus S is a basis for W and so $\dim W = k \leq n$. If $k = n$, then the reasoning at the end of the previous paragraph says that S is also a basis for V and hence $\operatorname{span} S = V$. So $W = V$.

  ■

- **Proposition 9.63:** Suppose that W is a subspace of a finite-dimensional vector space V. Any basis for W can be extended to a basis for V.

  Proof: By Proposition 9.62, $k = \dim W \leq \dim V = n$. Hence by Corollary 9.61 c) any basis $\beta = \{u_1, \dots, u_k\}$ for W can be extended to a basis $\tilde{\beta} = \{u_1, \dots, u_k, \tilde{u}_{k+1}, \dots, \tilde{u}_n\}$ for V.

  ■

10 Linear Maps

10.1 Definition

- With this chapter, we finally get to linear algebra: the algebra of linear maps!
- **Definition 10.1:** Suppose that V and W are vector spaces over the same field F. A function $T: V \to W$ is called a **linear transformation** (or **linear map**) from V to W if

  1. For any $x, y \in V$, $T(x + y) = T(x) + T(y)$.
  2. For any $x \in V$ and any $c \in V$, $T(cx) = cT(x)$.

  Note that 1) and 2) together can be equivalently formulated as saying that for any $x, y \in V$ and any $a, b \in F$,

  $$T(ax + by) = aT(x) + bT(y).$$

- **Note 10.2:** A few important properties of linear maps that immediately follow from the definition are

- $T(0) = 0$.
  - Proof: if you take any vector $x \in V$, $T(0) = T(0x) = 0T(x) = 0$.
- $T(x - y) = T(x) - T(y)$
  - Proof: $T(x - y) = T(x + (-1)y) = T(x) + (-1)T(y) = T(x) - T(y)$.

- Linear maps have had a profound influence on mathematics, both pure and applied. They are powerful because they satisfy strong structure. For instance, as we will soon prove, knowing their values at just a few points in a finite dimensional vector space can determine them completely. This is why we can derive an enormous amount of phenomenon about them, even though their definition looks very innocent at first. Hence it is extremely lucky that they appear in many applications including computer vision, (partial) differential equations, X-ray transforms, smooth maps on differential scales (Math 302), optimization in machine learning, etc. Often phenomenon in the real world aren't linear, but linearity is so desired that people will often perform approximations to their models to reduce them to linear ones. You see, linear algebra is useful: it's good that we're learning it!

- **Example 10.3:** The map $T : \mathbb{R}^2 \to M_{2\times 2}(\mathbb{R})$ given by

$$T\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 & 2a_2 \\ a_1 + a_2 & 0 \end{pmatrix}.$$

  is linear. To prove this, we check the two properties of linear maps. Checking the first property:

$$T\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right) = T\begin{pmatrix} a_1 + a_1 \\ a_2 + b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 & 2(a_2 + b_2) \\ (a_1 + b_1) + (a_2 + b_2) & 0 \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & 2a_2 \\ a_1 + a_2 & 0 \end{pmatrix} + \begin{pmatrix} b_1 & 2b_2 \\ b_1 + b_2 & 0 \end{pmatrix} = T\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + T\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

  We leave checking the second property as an exercise.

- **Example 10.4:** The map $\frac{d}{dx} : \{\text{differentiable } f : \mathbb{R} \to \mathbb{R}\} \to \{f : \mathbb{R} \to \mathbb{R}\}$ is a linear map since

$$\frac{d}{dx}(h + g) = \frac{d}{dx}(h) + \frac{d}{dx}(g),$$

$$\frac{d}{dx}(ch) = c\frac{d}{dx}(h),$$

  where $h, g \in \{\text{differentiable } f : \mathbb{R} \to \mathbb{R}\}$ and $c \in \mathbb{R}$ is a constant.

- **Example 10.5:** Integration over any interval $\int_a^b : \{\text{continuous } f : \mathbb{R} \to \mathbb{R}\} \to \mathbb{R}^1$ is a linear map since

$$\int_a^b \big(h(x) + g(x)\big)dx = \int_a^b h(x)dx + \int_a^b g(x)dx,$$

$$\int_a^b ch(x)dx = c\int_a^b h(x)dx$$

where $h, g \in \{\text{continuous } f : \mathbb{R} \to \mathbb{R}\}$ and $c \in \mathbb{R}$ is a constant.

- **Example 10.6:** A linear function $f : \mathbb{R}^1 \to \mathbb{R}^1$ given by $f(x) = mx + b$ is only linear if and only if $b = 0$ (Exercise).

- Two famous linear transformations that you should know:

- **Definition 10.7:** Suppose that V and W are vector spaces over the same field F. The **identity transformation** $I_V : V \to V$ is the linear map given by

$$I_V(x) = x.$$

The **zero transformation** $T_0 : V \to W$ is the linear map given by

$$T_0(x) = 0.$$

It's a very quick exercise to check that these two are indeed linear maps.

- **Definition 10.8:** Suppose that $T : V \to W$ is a linear transformation. The **null space** (or **kernel**) of T is the set of all vectors that T sends to zero:

$$N(T) = \{x \in V : T(x) = 0\} \subseteq V.$$

The **range** (or **image**) of T is the set of all vectors that T can map to:

$$R(T) = \{T(x) : x \in V\} = \{y \in W : \exists x \in V \text{ such that } y = T(x)\} \subseteq W.$$

- One of the purposes of the null space and range is to determine if f has an inverse and if so what type (c.f. Theorem 7.8). The role of range for this purpose is clear. The role of null space is not so clear at the moment, but will follow once we prove the amazing fact that a linear map is injective if and only if its kernel consists of just the zero vector.

- **Example 10.9:** Consider the projection map $P : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$P\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}.$$

Visually, this map projects all points in $\mathbb{R}^3$ perpendicularly onto the x, y plane. Here the null space of P is the z axis and the range is the x, y plane:

$$N(P) = \left\{\begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} : a_3 \in \mathbb{R}\right\} \qquad R(P) = \left\{\begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} : a_1, a_2 \in \mathbb{R}\right\}.$$

Notice that both $N(P)$ and $R(P)$ are subspaces. Furthermore, $\dim N(P) + \dim R(P) = \dim \mathbb{R}^3$. We will prove that this holds more generally! We start with the first of these:

- **Theorem 10.10:** Suppose that $T : V \to W$ is a linear transformation. Then $N(T)$ is a subspace of V and $R(T)$ is a subspace of W.

  Proof: We first show that $R(T)$ is a subspace of W. Checking the three properties:

  1. As we observed in Note 10.2, $T(0) = 0$ and so $0 \in R(T)$.
  2. Suppose $x, y \in R(T)$. Then there exist $v, w \in V$ such that $x = T(v)$ and $y = T(w)$. Thus
  $$x + y = T(v) + T(w) = T(v + w)$$
  and so $(x + y) \in R(T)$.
  3. Suppose that $x \in R(T)$ and $c \in F$. Then there exists $v \in V$ such that $x = T(v)$. Thus
  $$cx = cT(v) = T(cv)$$
  and so $cx \in R(T)$.

  The claim regarding $N(T)$ is left as an exercise (it's easier than $R(T)$).

  ■

- **Proposition 10.11:** Suppose that $T : V \to W$ is a linear map and that $\beta = \{v_1, \dots, v_m\}$ is a basis for V. Then
$$R(T) = \text{span}\{T(v_1), \dots, T(v_m)\} = \underbrace{\text{span}\{T(\beta)\}}_{\text{new notation}}.$$

  Proof: First let's show that $R(T) \subseteq \text{span}\{T(v_1), \dots, T(v_m)\}$. Take any $x \in R(T)$. Then there exists $v \in V$ such that $x = T(v)$. Since $\beta = \{v_1, \dots, v_m\}$ is a basis for V, we can write
  $$v = a_1 v_1 + \cdots + a_m v_m$$
  for some $a_1, \dots, a_m \in F$. Applying T to both sides of this equation gives
  $$x = T(v) = T(a_1 v_1 + \cdots + a_m v_m) = a_1 T(v_1) + \cdots + a_m T(v_m)$$
  and so $x \in \text{span}\{T(v_1), \dots, T(v_m)\}$. Thus indeed $R(T) \subseteq \text{span}\{T(v_1), \dots, T(v_m)\}$.

  Now let's show that $\text{span}\{T(v_1), \dots, T(v_m)\} \subseteq R(T)$. Take any $x \in \text{span}\{T(v_1), \dots, T(v_m)\}$. By definition of span,
  $$x = a_1 T(v_1) + \cdots + a_m T(v_m)$$
  for some $a_1, \dots, a_m \in F$. Notice that by the linearity of T, the above equation can be rewritten as
  $$x = T(a_1 v_1 + \cdots + a_m v_m)$$
  and so $x \in R(T)$. Thus indeed $\text{span}\{T(v_1), \dots, T(v_m)\} \subseteq R(T)$. This proves the theorem.

  ■

- **Definition 10.12:** Suppose that $T : V \to W$ is a linear map. If $N(T)$ is finite dimensional, we define the **nullity** of T to be the dimension of $N(T)$:

$$\text{nullity}(T) = \dim N(T).$$

If $R(T)$ is finite dimensional, we define the **rank** of T to be the dimension of $R(T)$:

$$\text{rank}\, T = \dim R(T).$$

- **Theorem 10.13: (Dimension Theorem)** Suppose that $T : V \to W$ is a linear map and that V is finite dimensional. Then both $N(T)$ and $R(T)$ are finite dimensional and

$$\text{nullity}(T) + \text{rank}(T) = \dim V. \tag{10.14}$$

Proof: By Theorem 10.10 we have that $N(T)$ is a subspace of V. Since V is finite dimensional, by Proposition 9.62 we get that $N(T)$ is indeed finite dimensional. Next let's prove that $R(T)$ is finite dimensional. We will do this by constructing a finite basis for it.

Let $\{u_1, \dots, u_m\}$ be a basis for $N(T)$, which observe implies that $\dim N(T) = m$. By Proposition 9.63 we can add vectors $\{v_1, \dots, v_k\}$ to this to get a basis $\{u_1, \dots, u_m, v_1, \dots, v_k\}$ for all of V, which note implies that $\dim V = m + k$. We claim that

$$\{T(v_1), \dots, T(v_k)\} \tag{10.15}$$

is a basis for $R(T)$, which observe implies that $\dim R(T) = k$. Not only will this prove that $R(T)$ is indeed finite dimensional, but this will also automatically prove (10.14) since then

$$\text{nullity}(T) + \text{rank}(T) = \dim N(T) + \dim R(T) = m + k = \dim V.$$

So let us prove that (10.15) is a basis for $R(T)$. First let's show that (10.15) spans $R(T)$. Take any vector $w \in R(T)$, we will show that w is in the span of (10.15). Since $w \in R(T)$, there exists $x \in V$ such that $w = T(x)$. Since $\{u_1, \dots, u_m, v_1, \dots, v_k\}$ is a basis for V,

$$x = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_k v_k$$

for some $a_1, \dots, a_m, b_1, \dots, b_k \in F$. Applying T to both sides gives

$$w = T(x) = b_1 T(v_1) + \cdots + b_k T(v_k)$$

where we've used that each $T(u_i) = 0$ since each $u_i \in N(T)$. This shows that indeed w is in the span of (10.15).

Next let's show that (10.15) is linearly independent and hence a basis for $R(T)$. Suppose

$$c_1 T(v_1) + \cdots + c_k T(v_k) = 0.$$

We need to show that each $c_k = 0$. Using the linearity of T on the left-hand side gives

$$T(c_1 v_1 + \cdots + c_k v_k) = 0.$$

Hence $c_1 v_1 + \cdots + c_k v_k$ is in the null space of T (i.e. in $N(T)$). Since $\{u_1, \dots, u_m\}$ is a basis for $N(T)$,

$$c_1v_1 + \cdots + c_kv_k = a_1u_1 + \cdots + a_mu_m$$

for some $a_1, \ldots, a_m \in F$. Taking the right-hand side to the left-hand side gives

$$c_1v_1 + \cdots + c_kv_k - a_1u_1 - \cdots - a_mu_m = 0.$$

Recall that $\{u_1, \ldots, u_m, v_1, \ldots, v_k\}$ is a basis for V and hence is linearly independent, and the above is a linear combination of these vectors equal to zero. Hence every coefficient above must be zero, and in particular each $c_i = 0$. As discussed above, this proves the theorem.

■

- Now we are ready to discuss the relationship between the null space and range of a linear map and its injectivity and surjectivity, which will play a central role in the study of the existence of inverses later.
- **Theorem 10.16:** Suppose that $T : V \to W$ is a linear map. Then T is injective if and only if $N(T) = \{0\}$.

  Proof: First suppose that T is injective. Since T is linear, we know that $T(0) = 0$. Since T is injective 0 can be the only vector that gets sent to 0 by T. Hence $N(T) = \{0\}$.

  Now suppose that $N(T) = \{0\}$. Then

  $$T(x) = T(y) \quad \Leftrightarrow \quad T(x) - T(y) = 0 \quad \Leftrightarrow \quad T(x - y) = 0 \quad \Leftrightarrow$$

  $$(x - y) \in N(T) = \{0\} \quad \Leftrightarrow \quad x - y = 0 \quad \Leftrightarrow \quad x = y.$$

  and hence T is injective. Don't write proofs like this on the homework, write them out using words!

  ■

- As the following theorem illustrates, the magic power of the dimension theorem is that injectivity of a linear map can give you information about its range and vice versa. This is due to the strong structure requirements that linear maps satisfy, and is not at all true for general maps.
- **Theorem 10.17:** Suppose that $T : V \to W$ is a linear map where V and W are finite dimensional vector spaces with the same dimension:

  $$\dim V = \dim W.$$

  Then the following are equivalent:

  1. T is injective.
  2. T is surjective.
  3. $\operatorname{rank} T = \dim V$.

  Proof:

$$T \text{ is injective} \xLeftrightarrow{\textbf{Theorem 10.16}} N(T) = \{0\} \quad \Leftrightarrow \quad \text{nullity}(T) = 0$$

$$\xLeftrightarrow{\text{Dimension theorem}} \dim V - \operatorname{rank} T = 0 \quad \Leftrightarrow \quad \operatorname{rank} T = \dim V = \dim W$$

$$\Leftrightarrow \quad \dim R(T) = \dim W \xLeftrightarrow{\textbf{Proposition 9.62}} R(T) = W \quad \Leftrightarrow \quad T \text{ is surjective.}$$

Again, don't write proofs like this on the homework: use words!

■

- **Example 10.18:** Consider the linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$T\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a+b+c \\ 2b+4c \\ b+2c \end{pmatrix}.$$

What is $\operatorname{rank} T$ (i.e. $\dim R(T)$)? Computing range is typically not easy (we'll learn how to do it), so what we can instead do is compute $N(T)$ and apply the dimension theorem. Let's solve for all vectors (a, b, c) that get sent to 0 by T (i.e. compute all vectors in $N(T)$). Setting the above equation to zero to gives the three following systems of three equations:

$$a + b + c = 0,$$

$$2b + 4c = 0,$$

$$b + 2c = 0.$$

The last equation tells us that $b = -2c$. Plugging this into the second equation gives $2(-2c) + 4c = 0c = 0$, an equation that is satisfied automatically! Plugging $b = -2c$ into the first equation gives $a + (-2c) + c = 0$ and so $a = c$. In other words, for any choice of c, the above system will be satisfied if $a = c$ and $b = -2c$ and on the other hand every solution to the above system will be of this form. Hence

$$N(T) = \left\{ \begin{pmatrix} c \\ -2c \\ c \end{pmatrix} : c \in \mathbb{R} \right\} = \left\{ c \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} : c \in \mathbb{R} \right\}.$$

In other words, $(1, -2, 1)$ is a basis for $N(T)$ and so $\dim N(T) = \text{nullity}(T) = 1$. Thus by the dimension theorem, $\operatorname{rank} T = \dim \mathbb{R}^3 - \text{nullity}(T) = 3 - 1 = 2$.

- **Theorem 10.19:** Suppose that V and W are vector spaces and that $\{v_1, \dots, v_m\}$ is a basis for V (in particular V is finite dimensional). For any vectors $w_1, \dots, w_n \in W$ there exists a unique linear map $T : V \to W$ such that each

$$w_i = T(v_i). \tag{10.20}$$

Proof: First let's show that such a map T exists. Take any vector $x \in V$, write it as

$$x = a_1 v_1 + \cdots + a_m v_m$$

for some unique $a_1, \dots, a_m \in F$ (possible since $\{v_1, \dots, v_m\}$ is a basis for V, c.f. Theorem 9.45), and define

$$T(x) = a_1 w_1 + \cdots + a_m w_m. \tag{10.21}$$

This is a good guess for how to define T considering that we want it to be linear and we know that (10.20) holds by assumption. First let's check that T is linear. Take any $a, b \in F$ and any $x = a_1 v_1 + \cdots + a_m v_m$ and $y = b_1 v_1 + \cdots + b_m v_m$ written in the above form and observe that

$$T(ax + by) = T\big(a(a_1 v_1 + \cdots + a_m v_m) + b(b_1 v_1 + \cdots + b_m v_m)\big)$$

$$= T\big((aa_1 + bb_1)v_1 + \cdots + (aa_m + bb_m)v_m\big) = (aa_1 + bb_1)w_1 + \cdots + (aa_m + bb_m)w_m$$

$$= a(a_1 w_1 + \cdots + a_m w_m) + b(b_1 w_1 + \cdots + b_m w_m) = aT(x) + bT(y).$$

Next, observe that if we set $x = v_i$ then by (10.21) we get the desired property $T(v_i) = w_i$. So this T satisfies the conclusions of the theorem. We just have to show that it is unique!

Suppose that $\hat{T} : V \to W$ was another linear map that also satisfies (10.20). We need to show that $T = \hat{T}$. Take any $x = a_1 v_1 + \cdots + a_m v_m$ written in the above form and observe that since both T and $\tilde{T}$ satisfy (10.20),

$$T(x) = T(a_1 v_1 + \cdots + a_m v_m) = a_1 T(v_1) + \cdots + a_m T(v_m) = a_1 w_1 + \cdots + a_m w_m,$$

$$\hat{T}(x) = \hat{T}(a_1 v_1 + \cdots + a_m v_m) = a_1 \hat{T}(v_1) + \cdots + a_m \hat{T}(v_m) = a_1 w_1 + \cdots + a_m w_m.$$

So $T(x) = \hat{T}(x)$ for any $x \in V$. Hence indeed $T = \hat{T}$.

■

- The above theorem conveys the important principle that knowing the values of a linear map over a finite dimensional vector space at just a few points, in particular on a basis, is enough to determine its value everywhere. This is due to the stringent structure that linear maps satisfy. The following corollary is another way to state this precisely:

- **Corollary 10.22:** Suppose that V and W are vector space and that $\beta = \{v_1, \dots, v_m\}$ is a basis for V (in particular V is finite dimensional). If two linear maps $T, L : V \to W$ agree on β (i.e. each $T(v_i) = L(v_i)$) then T and L agree everywhere (i.e. $T(x) = L(x)$ for all $x \in V$).

  Proof: By the previous theorem, there is a unique linear map $\hat{T}$ that has values $T(v_i) = L(v_i)$ at each v_i, in particular that linear map is T and L (i.e. $\hat{T} = T = L$).

  ■

10.2 Matrix Representations

- We've developed a substantial theory of linear maps. However the abstract form that we have at the moment doesn't lend itself naturally to computation. Matrices were invented for this exact purpose, which we now study. First we need a preliminary definition.

- **Definition 10.23:** Suppose that V is a finite dimensional vector space. An **ordered basis** $\beta = \{v_1, v_2, \dots, v_m\}$ for V is a basis where we declare that v_1 is the first vector, v_2 is the second vector, v_3 is the third vector, and so on. In other words, β isn't just a set of vectors, but we also assign an order to the list of vectors. There's a subtle difference!

- **Example 10.24:** The list

$$\left\{ \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{e_1}, \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{e_2}, \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}}_{e_3}, \dots, \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}}_{e_m} \right\}$$

  is called the **standard ordered basis** for $\mathbb{R}^m$. Note that if we consider the ordered bases $\beta = \{e_1, e_2, e_3\}$ and $\gamma = \{e_2, e_1, e_3\}$ of $\mathbb{R}^3$, $\beta \neq \gamma$.

- **Example 10.25:** The list $\{1, x, x^2, \dots, x^n\}$ is the **standard ordered basis** for $P_n(F)$.

- **Note 10.26:** Suppose that we have a linear map $T : V \to W$ between two finite dimensional vector spaces. Let $\beta = \{v_1, \dots, v_m\}$ and $\gamma = \{w_1, \dots, w_n\}$ be ordered basis for V and W respectively. Take any vector $x \in V$ which we can write uniquely as

$$x = a_1 v_1 + \cdots + a_m v_m. \tag{10.27}$$

  Instead of doing algebra on the clunky notation on the right-hand side, we can neatly and uniquely represent the right-hand side in matrix notation:

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}.$$

  For this reason $m \times 1$ matrices are also often referred to as a "column vectors." The brackets "$[\ ]_\beta$" around x remind us that this is only a representation of x and that it highly depends on which ordered basis β we choose (i.e. if you change the β, the representation on the right-hand side changes).

  Now let's look into representing T. Applying T to both sides of (10.27) gives

$$T(x) = a_1 T(v_1) + \cdots + a_m T(v_m). \tag{10.28}$$

  We can write each $T(v_j)$ uniquely as

$$T(v_j) = b_{1,j} w_1 + \cdots + b_{n,j} w_n.$$

  Plugging this into (10.28) gives

$$T(x) = a_1(b_{1,1} w_1 + \cdots + b_{n,1} w_n) + \cdots + a_m(b_{1,m} w_1 + \cdots + b_{n,m} w_n)$$

$$= (b_{1,1}a_1 + \cdots + b_{1,m}a_m)w_1$$
$$+(b_{2,1}a_1 + \cdots + b_{2,m}a_m)w_2$$
$$\vdots$$
$$+(b_{n,1}a_1 + \cdots + b_{n,m}a_m)w_n$$

and so

$$[T(x)]_\gamma = \begin{pmatrix} b_{1,1}a_1 + \cdots + b_{1,m}a_m \\ b_{2,1}a_1 + \cdots + b_{2,m}a_m \\ \vdots \\ b_{n,1}a_1 + \cdots + b_{n,m}a_m \end{pmatrix} \tag{10.29}$$

From here we see that knowledge of the $b_{i,j}$'s completely determine $T(x)$ for any $x \in V$, and the other way around as well. This is a special case of Theorem 10.19. So we can neatly and uniquely represent T in the ordered bases β and γ as the $n \times m$ matrix

$$[T]_\beta^\gamma = \begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,m} \end{pmatrix}.$$

If $V = W$ and $\beta = \gamma$, we simply write $[T]_\beta$.

In fact, by (10.29) we see that we can easily compute $[T(x)]_\gamma$ from $[T]_\beta^\gamma$ and $[x]_\beta$ by multiplying entries of the rows of $[T]_\beta^\gamma$ by the entries of $[x]_\beta$, adding the results and placing this into each row. This defines matrix-vector multiplication:

$$[T(x)]_\gamma = [T]_\beta^\gamma [x]_\beta = \begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,m} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} b_{1,1}a_1 + \cdots + b_{1,m}a_m \\ b_{2,1}a_1 + \cdots + b_{2,m}a_m \\ \vdots \\ b_{n,1}a_1 + \cdots + b_{n,m}a_m \end{pmatrix}$$

This may look like a neat notational trick, but this is the power of matrices: they allow us to efficiently compute operations with linear maps! They will play an even more essential role when we consider compositions.

We also make the remark that if you fix a matrix A, then multiplying it by a column vector as above is a linear map on column vectors (i.e. taking v and outputting Av is a linear map). You will prove this on the homework.

- **Example 10.30:** Consider the linear map $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ given by

$$T(ax^2 + bx + c) = \int_0^x (as^2 + bs + c)ds = \frac{a}{3}s^3 + \frac{b}{2}s^2 + cs + 0.$$

Using the standard ordered basis $\beta_1 = \{1, x, x^2\}$ and $\beta_2 = \{1, x, x^2, x^3\}$ for $P_2(\mathbb{R})$ and $P_3(\mathbb{R})$ respectively, we have the representations

$$[ax^2 + bx + c]_{\beta_1} = \begin{pmatrix} c \\ b \\ a \end{pmatrix},$$

$$\left[\tilde{a}x^3 + \tilde{b}x^2 + \tilde{c}x + \tilde{d}\right]_{\beta_2} = \begin{pmatrix} \tilde{d} \\ \tilde{c} \\ \tilde{b} \\ \tilde{a} \end{pmatrix},$$

$$[T]_{\beta_1}^{\beta_2} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix},$$

since observe that

$$[T(ax^2 + bx + c)]_{\beta_2} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 0c + 0b + 0a \\ 1c + 0b + 0a \\ 0c + (1/2)b + 0a \\ 0c + 0b + (1/3)a \end{pmatrix} = \begin{pmatrix} 0 \\ c \\ b/2 \\ a/3 \end{pmatrix}.$$

- **Note 10.31:** There are two special matrix representations that you should be aware of. Suppose that V and W are vector spaces over the same field and that β and γ are ordered bases for them respectively. Then the matrix representation of the zero map $T_0 : V \to W$ and the identity map $I_V : V \to V$ are given by

$$[T_0]_{\beta}^{\gamma} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad [I_V]_{\beta} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

  The second matrix is called the $n \times n$ **identity matrix**.

10.3 Vector Space of Linear Maps

- Linear maps themselves in fact form vector spaces, a fact that finds many important applications. Let's build towards this concept.

- **Definition 10.32:** Suppose that $T, L : V \to W$ are linear maps and that $c \in F$. Then we define the sum $T + L : V \to W$ and the scalar multiplication $cT : V \to W$ as the linear maps given by

$$(T + L)(x) = T(x) + L(x) \quad \text{and} \quad (cT)(x) = c[T(x)], \qquad \forall x \in V.$$

- The above definition shouldn't seem strange: it's just addition and scaling of functions. You most likely saw similar things in calculus, though you may not have written it down rigorously as above.

- **Theorem 10.33:** Suppose that $T, L : V \to W$ are linear maps and that $c \in F$. Then both

$$T + L : V \to W \quad \text{and} \quad cT : V \to V$$

are linear.

Proof: Let's show that $T + L$ is linear, aT is left as an exercise (it's easier). For any $x, y \in V$ and $a, b \in F$ we have that

$$(T + L)(ax + by) = T(ax + by) + L(ax + by) = aT(x) + bT(y) + aL(x) + bL(y)$$
$$= a[T(x) + L(x)] + b[T(y) + L(y)] = a(T + L)(x) + b(T + L)(y).$$

Hence indeed $T + L$ is linear.

■

- **Definition 10.34:** Suppose that V and W are vector spaces over the same field. We denote the set of all linear maps from V to W by $\mathcal{L}(V, W)$. The set $\mathcal{L}(V, W)$ is a vector space with the addition and scalar multiplication as defined in Definition 10.32 where the zero "vector" is the zero map $T_0 : V \to W$. If $V = W$, then we simply write $\mathcal{L}(V)$.
- Fortunately, matrix representations behave very naturally under addition and scalar multiplication of both vectors and linear maps. This is the content of the following theorem:
- **Theorem 10.35:** Suppose that V and W are vector spaces over the same field F and that β and γ are ordered bases for V and W respectively. Then

  a) If $x, y \in V$ and $a \in F$, then

  $$[x + y]_\beta = [x]_\beta + [y]_\beta \quad \text{and} \quad [ax]_\beta = a[x]_\beta.$$

  b) If $T, L : V \to W$ are linear maps and $a \in F$, then

  $$[T + L]_\beta^\gamma = [T]_\beta^\gamma + [L]_\beta^\gamma \quad \text{and} \quad [aT]_\beta^\gamma = a[T]_\beta^\gamma.$$

Proof: Let $\beta = \{v_1, \dots, v_m\}$ and $\gamma = \{w_1, \dots, w_n\}$. First let's prove a). Take any $x, y \in V$. As in Note 10.26, we can write

$$x = a_1 v_1 + \cdots + a_m v_m \quad \text{and} \quad y = \hat{a}_1 v_1 + \cdots + \hat{a}_m v_m$$
$$\implies \quad x + y = (a_1 + \hat{a}_1)v_1 + \cdots + (a_m + \hat{a}_m)v_m$$
$$\implies \quad [x + y]_\beta = \begin{pmatrix} a_1 + \hat{a}_1 \\ \vdots \\ a_m + \hat{a}_m \end{pmatrix} = [x]_\beta + [y]_\beta.$$

The second equation in a) is left as an exercise. Now let's prove b). Take any linear maps $T, L : V \to W$. We have that

$$T(x) = \begin{matrix} (b_{1,1}a_1 + \cdots + b_{1,m}a_m)w_1 \\ \vdots \\ +(b_{n,1}a_1 + \cdots + b_{n,m}a_m)w_n \end{matrix} \quad \text{and} \quad L(x) = \begin{matrix} (\hat{b}_{1,1}a_1 + \cdots + \hat{b}_{1,m}a_m)w_1 \\ \vdots \\ +(\hat{b}_{n,1}a_1 + \cdots + \hat{b}_{n,m}a_m)w_n \end{matrix}$$

$$\Rightarrow \quad \begin{aligned} T(x)+L(x) \quad &= \left[\left(b_{1,1}+\hat{b}_{1,1}\right)a_1+\cdots+\left(b_{1,m}+\hat{b}_{1,m}\right)a_m\right]w_1 \\ &\quad\ \vdots \\ &+\left[\left(b_{n,1}+\hat{b}_{n,1}\right)a_1+\cdots+\left(b_{n,m}+\hat{b}_{n,m}\right)a_m\right]w_n \end{aligned}$$

$$\Rightarrow \quad [T+L]_\beta^\gamma = \begin{pmatrix} b_{1,1}+\hat{b}_{1,1} & \cdots & b_{1,m}+\hat{b}_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1}+\hat{b}_{n,1} & \cdots & b_{n,m}+\hat{b}_{n,m} \end{pmatrix} = [T]_\beta^\gamma + [L]_\beta^\gamma.$$

The second equation in b) is left as an exercise.

10.4 Composition of Linear Maps

- An important operation of linear maps it to compose them. This for instance appears when you want to perform one rotation, and then another!
- **Note 10.36:** For linear maps T and L, we will typically write LT for their composition instead of the clunkier $L \circ T$. As we'll soon see, this is more than just another shorthand. This notation is introduced because this notation aligns nicely with the fact that in the matrix representation world, composition of linear maps will be associated to matrix multiplication.
- **Theorem 10.37:** Suppose that $T : V \to W$ and $L : W \to Z$ are linear maps. The composition $LT : V \to Z$ is also linear.

  Proof: Take any vectors $x, y \in V$ and any $a, b \in F$. Then

$$\begin{aligned} LT(ax+by) = \ L\big(T(ax+by)\big) &= L\big(aT(x)+aT(y)\big) = aL\big(T(x)\big)+bL\big(T(y)\big) \\ &= a(LT)(x)+b(LT)(y). \end{aligned}$$

  Hence LT is indeed linear.

■

- Some trivial but important properties to be aware of:
- **Theorem 10.38:** The following are true:

  a) If $L, U \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(W, Z)$ then

$$T(L+U) = TL + TU.$$

  If $T \in \mathcal{L}(V, W)$ and $L, U \in \mathcal{L}(W, Z)$ then

$$(L+U)T = LT + UT.$$

  b) If $U \in \mathcal{L}(V, W)$, $L \in \mathcal{L}(W, Z)$, and $T \in \mathcal{L}(Z, Y)$, then

$$(TL)U = T(LU).$$

  c) If $T \in \mathcal{L}(V, W)$, then

$$TI_V = I_W T = T,$$

d) If $L \in \mathcal{L}(V,W)$ and $T \in \mathcal{L}(W,Z)$, then

$$a(TL) = (aT)L = T(aL).$$

Proof: Left as an exercise. Parts a) and d) are a quick calculation, and the rest are just arguments about where points get mapped to (draw diagrams for b) and c) to help visualize them). ■

- **Note 10.39:** Let's now compute the matrix representation of compositions of linear maps. Suppose that $T : V \to W$ and $L : W \to Z$ are linear maps and that

$$\beta = \{v_1, \dots, v_m\} \quad \text{and} \quad \gamma = \{w_1, \dots, w_n\} \quad \text{and} \quad \delta = \{z_1, \dots, z_k\}$$

are ordered bases of V, W, Z respectively. Take any vector $x = a_1v_1 + \cdots + a_mv_m$. Recall from Note 10.26 that representing T as

$$[T]_\beta^\gamma = \begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,m} \end{pmatrix}$$

$$\implies \quad T(x) = \big(b_{1,1}a_1 + \cdots + b_{1,m}a_m\big)w_1 + \cdots + \big(b_{n,1}a_1 + \cdots + b_{n,m}a_m\big)w_n.$$

Similarly, taking any vector $y = c_1w_1 + \cdots + c_nw_n$ and representing L as

$$[L]_\gamma^\delta = \begin{pmatrix} \hat{b}_{1,1} & \cdots & \hat{b}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{b}_{k,1} & \cdots & b_{k,n} \end{pmatrix}$$

$$\implies \quad [L(y)]_\delta = \begin{pmatrix} \hat{b}_{1,1}c_1 + \cdots + \hat{b}_{1,n}c_n \\ \vdots \\ \hat{b}_{k,1}c_1 + \cdots + \hat{b}_{k,n}c_n \end{pmatrix}.$$

Plugging in $y = T(x)$ here gives that

$$\big[L\big(T(x)\big)\big]_\delta = \begin{pmatrix} \hat{b}_{1,1}\big(b_{1,1}a_1 + \cdots + b_{1,m}a_m\big) + \cdots + \hat{b}_{1,n}\big(b_{n,1}a_1 + \cdots + b_{n,m}a_m\big) \\ \vdots \\ \hat{b}_{k,1}\big(b_{1,1}a_1 + \cdots + b_{1,m}a_m\big) + \cdots + \hat{b}_{k,n}\big(b_{n,1}a_1 + \cdots + b_{n,m}a_m\big) \end{pmatrix}.$$

Distributing and rearranging gives that this is equal to

$$\begin{pmatrix} \big(\hat{b}_{1,1}b_{1,1} + \cdots + \hat{b}_{1,n}b_{n,1}\big)a_1 + \cdots + \big(\hat{b}_{1,1}b_{1,m} + \cdots + \hat{b}_{1,n}b_{n,m}\big)a_m \\ \vdots \\ \big(\hat{b}_{k,1}b_{1,1} + \cdots + \hat{b}_{k,n}b_{n,1}\big)a_1 + \cdots + \big(\hat{b}_{k,1}b_{1,m} + \cdots + \hat{b}_{k,n}b_{n,m}\big)a_m \end{pmatrix}.$$

So the matrix representation of the composition LT is given by

$$[LT]_\beta^\delta = \begin{pmatrix} \hat{b}_{1,1}b_{1,1} + \cdots + \hat{b}_{1,n}b_{n,1} & \cdots & \hat{b}_{1,1}b_{1,m} + \cdots + \hat{b}_{1,n}b_{n,m} \\ \vdots & \ddots & \vdots \\ \hat{b}_{k,1}b_{1,1} + \cdots + \hat{b}_{k,n}b_{n,1} & \cdots & \hat{b}_{k,1}b_{1,m} + \cdots + \hat{b}_{k,n}b_{n,m} \end{pmatrix}.$$

This may look messy, however the sums and products in the entries of this matrix have a very good pattern. In particular, we can *define* the **product of the matrices** $[T]_\beta^\gamma$ and $[L]_\gamma^\delta$ as such:

$$\begin{pmatrix} \hat{b}_{1,1}b_{1,1}+\cdots+\hat{b}_{1,n}b_{n,1} & \cdots & \hat{b}_{1,1}b_{1,m}+\cdots+\hat{b}_{1,n}b_{n,m} \\ \vdots & \ddots & \vdots \\ \hat{b}_{k,1}b_{1,1}+\cdots+\hat{b}_{k,n}b_{n,1} & \cdots & \hat{b}_{k,1}b_{1,m}+\cdots+\hat{b}_{k,n}b_{n,m} \end{pmatrix} = \begin{pmatrix} \hat{b}_{1,1} & \cdots & \hat{b}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{b}_{k,1} & \cdots & \hat{b}_{k,n} \end{pmatrix} \begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,m} \end{pmatrix}$$

to get the matrix of the composition LT. In other words,

$$[LT]_\beta^\delta = [L]_\gamma^\delta [T]_\beta^\gamma.$$

Hence matrices give a quick and convenient way to compute compositions of linear maps!

- **Notation 10.40:** If $a_k, \dots, a_n$ are objects that we can sum, a notation for their sum is:

$$\sum_{i=k}^{n} a_i = a_k + \cdots + a_n.$$

- **Example 10.41:** Examples of usage of the above sum notation include

$$\sum_{i=1}^{m} a_i v_i = a_1 v_1 + \cdots a_m v_m \quad \text{and} \quad \sum_{k=-1}^{3} (2k)^2 = (-2)^2 + 0^2 + 2^2 + 4^2 + 6^2 = 60.$$

If we consider matrix multiplication defined in Note 10.39:

$$\begin{pmatrix} c_{1,1} & \cdots & c_{1,m} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & c_{n,m} \end{pmatrix} = \underbrace{\begin{pmatrix} a_{1,1} & \cdots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,k} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{k,1} & \cdots & b_{n,m} \end{pmatrix}}_{B},$$

then we have that each

$$c_{ij} = a_{i1}b_{1j} + \cdots + a_{ik}b_{kj} = \sum_{r=1}^{k} a_{ir}b_{rj}.$$

Note that for the matrix multiplication to make sense we need the width of A be equal to the height of B (both equal to k in this example).

- **Example 10.42:** Consider the linear projection maps $P : \mathbb{R}^3 \to \mathbb{R}^3$ and $\hat{P} : \mathbb{R}^3 \to \mathbb{R}^3$ onto the xy-plane and xz-plane respectively:

$$P(a,b,c) = (a,b,0) \quad \text{and} \quad \hat{P}(a,b,c) = (a,0,c).$$

With respect to the standard ordered basis $\beta = \{e_1, e_2, e_3\}$ of $\mathbb{R}^3$, their matrices are given by

$$[P]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \left[\hat{P}\right]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(check this!). Thus the matrix of the composition $\hat{P} \circ P$ is given by

$$\left[\hat{P} \circ P\right]_\beta = \left[\hat{P}\right]_\beta [P]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It may not be clear what this composition does at first, so let us apply it to a vector $(a, b, c) \in \mathbb{R}^3$. We have that

$$\left[\left(\hat{P} \circ P\right)(a, b, c)\right]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} = [(a, 0, 0)]_\beta.$$

In other words, this shows that the composition $\hat{P} \circ P$ projects vectors onto the x-axis. If you think about it geometrically, that makes sense since if you first project onto the xy-plane and then onto the xz-plane, then that's equivalent to projecting onto the x-axis from the start!

- **Note 10.43:** A warning about matrix multiplication: it is not commutative! In other words, if A and B are square matrices then AB is not necessarily equal to BA (i.e. it may or may not be). In the above example we do have that $[P]_\beta \left[\hat{P}\right]_\beta = \left[\hat{P}\right]_\beta [P]_\beta$. However, an example of where commutativity breaks is

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

since the left-hand side is equal to $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ while the right-hand side is equal to $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$.

If A and B are not both square, then it can be the case that AB makes sense but BA does not make sense (i.e. the sizes of the matrices are not correct for the product to be defined).

- **Note 10.44:** In the homework you will learn about the **transpose** of a matrix A, which is denoted by $A^\top$ and is obtained by flipping the matrix across its diagonal running from the upper left corner. Transposes are important in the theory of inner products and dual spaces. A fact you should be aware of is that

$$(AB)^\top = B^\top A^\top.$$

This is easy to prove and is left as an exercise: simply see what happens when you perform a matrix multiplication and then flip it across a diagonal. Alternatively, the proof is written out on page 89 of the book.

- **Theorem 10.45:** Fix a field F. In the following capital letters denote matrices with entries in F and lower-case letters denote scalars in F. All matrix multiplications are assumed to make sense (i.e. the sizes of the matrices are correct)

a) $A(B+C) = AB + AC$ and $(D+E)A = DA + EA$,
b) $(AB)C = A(BC)$
c) $I_m A = A = AI_n$
d) $a(AB) = (aA)B = A(aB)$
e) $A(b_1B_1 + \cdots + b_kB_k) = b_1AB_1 + \cdots + b_kAB_k$ and $(c_1C_1 + \cdots + c_kC_k)A = c_1C_1A + \cdots + c_kC_kA$.

Proof: We note that e) follows immediately from a) and d). For parts a) – d), you could prove them by direct calculation or alternatively you can note that they follow immediately by observing that they are simply representations of the analogous parts in Theorem 10.38 applied to linear maps over Euclidean spaces. To illustrate this last remark, let's prove b). Let $U : F^m \to F^n$, $L : F^n \to F^k$, and $T : F^k \to F^j$ be linear maps whose matrices with respect to the standard bases are C, B, and A respectively. Then by Theorem 10.38 (here we omit writing the bases)

$$T(LU) = (TL)U \quad \Longrightarrow \quad A(BC) = [T(LU)] = [(TL)U] = (AB)C.$$

■

10.5 Invertibility and Isomorphism

- We will now study the following question: when do linear maps have inverses. One application of this, as we'll see later, is obtaining solutions to systems of equations.
- **Definition 10.46:** Suppose that $T : V \to W$ is a linear map. A function $U : W \to V$ is called an **inverse** of T if both $TU = I_V$ and $UT = I_W$. If such a map U exists, then it is unique, we call T **invertible**, and we write $T^{-1} = U$.
- We remark that the above definition is nothing new: it exactly matches our definition of inverse in Definition 7.7 but just stated for the special case of when the map T (which was f in Definition 7.7) is linear. Recall from Theorem 7.8 that a map is invertible if and only if it is injective and surjective (i.e. is bijective).
- **Note 10.47:** If $U : V \to W$ aand $T : W \to Z$ are invertible linear map, then

$$(TU)^{-1} = U^{-1}T^{-1},$$

$$(T^{-1})^{-1} = T.$$

This is true simply for set theoretical reasons and hence holds not only for linear maps. To see why, draw a diagram about where points get mapped to on both sides of the above equations.

- **Example 10.48:** As we'll learn later, a rotation in space by an angle θ is a linear transformation. The inverse of a rotation will be a rotation in the reverse direction by the same angle (or equivalently by the angle $-\theta$). Note that the inverse is linear, coincidence? No:
- **Theorem 10.49:** Suppose that $T : V \to W$ is an invertible linear map. Then the inverse $T^{-1} : W \to V$ is also linear.

Proof: Take any $x, y \in W$ and any $a, b \in F$. We have that $T^{-1}(ax + by)$ is the unique element of V that gets mapped to $ax + by$ by T. We claim that that unique element is $aT^{-1}(x) + bT^{-1}(y)$. To prove this, observe that

$$T\big(aT^{-1}(x) + bT^{-1}(y)\big) = aTT^{-1}(x) + bTT^{-1}(y) = ax + by.$$

Thus indeed

$$T^{-1}(ax + by) = aT^{-1}(x) + bT^{-1}(y)$$

and hence T^{-1} is linear.

- The existence of an invertible linear map between two vector spaces in fact says that the two vector spaces have very similar structure. We will explore this in the next few theorems, starting with the following:

- **Theorem 10.50:** Suppose that $T : V \to W$ is an invertible linear map. Then V is finite dimensional if and only if W is finite dimensional. If they are finite dimensional, then $\dim V = \dim W$.

- **Proof:** First suppose that V is finite dimensional. Then by the dimension theorem (Theorem 10.13)

$$\dim N(T) + \dim R(T) = \dim V.$$

Since T is injective, by Theorem 10.16, $\dim N(T) = 0$. Since T is surjective, $R(T) = W$. Thus the above equation gives that $\dim W = \dim V$. So both V and W are finite dimensional and their dimensions are equal. The reverse direction (i.e. assuming that W is finite dimensional) is proved similarly but instead using $T^{-1} : W \to V$.

■

- **Definition 10.51:** We say that two vector spaces V and W are **isomorphic** if there exists an invertible linear map $T : V \to W$ (which implies that they are defined over the same field). Any such invertible linear map $T : V \to W$ is called an **isomorphism**.

- **Theorem 10.52:** Suppose that V and W are finite-dimensional vector spaces over the same field. Then V is isomorphic to W if and only if $\dim V = \dim W$.

Proof: If V and W are isomorphic, then Theorem 10.50 says that $\dim V = \dim W$. Now suppose that $\dim V = \dim W$. Let $\beta = \{v_1, \dots, v_m\}$ and $\gamma = \{w_1, \dots, w_n\}$ be bases for V and W respectively. By Theorem 10.19 there exists a unique linear map T such that each

$$T(v_i) = w_i.$$

By Proposition 10.11,

$$R(T) = \text{span}\{T(v_1), \dots, T(v_m)\} = \text{span}\{w_1, \dots, w_n\} = W$$

and hence T is surjective. Since $\dim V = \dim W$, by Theorem 10.17 T is injective, and hence bijective, and thus an isomorphism. So V and W are indeed isomorphic.

■

- **Corollary 10.53:** Suppose that V is a vector space over a field F. Then V is isomorphic to F^n if and only if $\dim V = n$ (here $n \geq 0$ is an integer).

  Proof: This immediately follows from the previous theorem since $\dim F^n = n$. ■

- The above corollary is profound: it says that structure wise finite dimensional vector spaces aren't so varied after all, they all look like F^n. We've actually seen this many times already when we represented vectors in finite dimensional vector spaces as column vectors.

- We've discussed invertibility of linear maps, but what does this look like on the matrix representation side? We endeavor to find this out, starting with the following definition:

- **Definition 10.54:** Suppose that A is an $n \times n$ matrix. We say that A is **invertible** if there exists an $n \times n$ matrix B such that $AB = BA = I_n$. We will prove below that such a matrix B is unique and hence we call B the **inverse** of A and write $A^{-1} = B$.

- **Lemma 10.55:** Suppose that A is an invertible $n \times n$ matrix. Then there exists only one matrix B such that $AB = BA = I_n$.

  Proof: Just like in the proof of Theorem 10.45, we use the power of representation! Let $\hat{B}$ be another such matrix. Let $T, U, \hat{U} : F^n \to F^n$ be linear maps whose matrices with respect to the standard ordered basis β of F^n are $A, B, \hat{B}$ respectively. Then

$$AB = BA = I_n = A\hat{B} = \hat{B}A \quad \Leftrightarrow \quad [TU]_\beta = [UT]_\beta = I_n = \left[T\hat{U}\right]_\beta = \left[\hat{U}T\right]_\beta$$

$$\Leftrightarrow \quad TU = UT = I_{F^n} = T\hat{U} = \hat{U}T$$

  Because of the uniqueness of inverses of maps, this implies that

$$U = \hat{U} \quad \Leftrightarrow \quad B = [U]_\beta = \left[\hat{U}\right]_\beta = \hat{B}.$$

■

- **Example 10.56:** We will later learn how to efficiently compute the inverses of matrices. For now, you can verify by direct computation that the inverse of $\begin{pmatrix} 7 & 4 \\ 3 & 2 \end{pmatrix}$ is $\begin{pmatrix} 1 & -2 \\ -1.5 & 3.5 \end{pmatrix}$ since

$$\begin{pmatrix} 7 & 4 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -1.5 & 3.5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -2 \\ -1.5 & 3.5 \end{pmatrix}\begin{pmatrix} 7 & 4 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

- **Theorem 10.57:** Suppose that V and W are finite dimensional vector spaces over the same field of the same dimension. Let β and γ be ordered bases for V and W respectively. Then a linear map $T : V \to W$ is invertible if and only if $[T]_\beta^\gamma$ is invertible. If T is invertible then $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Proof: First suppose that T is invertible. Let $A = [T]_\beta^\gamma$ and let $B = [T^{-1}]_\gamma^\beta$. Then

$$AB = [T]_\beta^\gamma [T^{-1}]_\gamma^\beta = [TT^{-1}]_\gamma^\gamma = [I_W]_\gamma^\gamma = I_n.$$

Similarly one shows that $BA = I_n$. Hence $A = [T]_\beta^\gamma$ is invertible and its inverse is $B = [T^{-1}]_\gamma^\beta$ (i.e. $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$).

Now suppose that $[T]_\beta^\gamma$ is invertible. Let $U : W \to V$ be the (unique) linear map whose matrix with respect to the above ordered bases is $([T]_\beta^\gamma)^{-1}$ (i.e. $[U]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$). Then

$$[TU]_\gamma^\gamma = [T]_\beta^\gamma [U]_\gamma^\beta = [T]_\beta^\gamma ([T]_\beta^\gamma)^{-1} = I_n = [I_W]_\gamma^\gamma \quad \Longrightarrow \quad TU = I_W.$$

Similarly one shows that $UT = I_V$. Hence T is invertible (i.e. $T^{-1} = U$).

■

- The following theorem summarizes many of our ideas on representations, and says a little more:
- **Theorem 10.58:** Suppose that V and W are finite dimensional vector spaces over the same field F. Let $n = \dim V$ and $m = \dim W$. Let β and γ be ordered bases for V and W respectively. Then the maps

$$\Phi_\beta^\gamma : \mathcal{L}(V, W) \to M_{m\times n}(F) \quad \text{given by} \quad \Phi_\beta^\gamma(T) = [T]_\beta^\gamma$$

$$\phi_\beta : V \to F^n \quad \text{given by} \quad \phi_\beta(x) = [x]_\beta$$

are isomorphisms.

Proof: We already discussed why they are bijective in Note 10.26, and you guys proved their linearity in the homework. Hence Φ_β^γ and ϕ_β are isomorphisms. ■

- **Corollary 10.59:** Suppose that V and W are finite dimensional vector spaces over the same field F. Let $n = \dim V$ and $m = \dim W$. Then

$$\dim \mathcal{L}(V, W) = mn.$$

Proof: By the previous theorem $\mathcal{L}(V, W)$ is isomorphic to $M_{m\times n}(F)$ and $\dim M_{m\times n}(F) = mn$. Hence by Theorem 10.52 $\dim \mathcal{L}(V, W) = mn$.

■

10.6 Change of Basis

- As observed already, ordered bases are powerful tools for computing and performing operations on linear maps and vectors by passing to their representations. However, sometimes the standard bases that come with vector spaces are not the most convenient ones for performing certain calculations. The solution is straightforward: just use another ordered basis. This is what we endeavor to study now, and name of the technique is unsurprisingly called “change of bases.”

We want to emphasize that change of bases only applies to representations and not linear maps and vectors themselves since bases are irrelevant to the definition of the latter two.

- **Theorem 10.60:** Suppose V is a finite dimensional vector space and that β and β' are ordered bases for V. Letting $Q = [I_V]_{\beta}^{\beta'}$, we have that

$$[v]_{\beta'} = Q[v]_{\beta} \qquad \forall v \in V. \tag{10.61}$$

Furthermore, Q is invertible and $Q^{-1} = [I_V]_{\beta'}^{\beta}$.

Remark: Because of (10.61) we say that Q changes **$\boldsymbol{\beta}$-coordinates to $\boldsymbol{\beta'}$-coordinates** and we say that Q is a **change of basis/coordinates matrix**. Observe that because $Q^{-1} = [I_V]_{\beta'}^{\beta}$, Q^{-1} is also a change of basis matrix that does the reverse: it changes β'-coordinates to β-coordinates.

Proof: By definition of matrix representations (see Note 10.26)

$$Q[v]_{\beta} = [I_V]_{\beta}^{\beta'}[v]_{\beta} = [I_V(v)]_{\beta'} = [v]_{\beta'}.$$

Next, the matrix Q is invertible by Theorem 10.57 because it is a representation of an invertible map (i.e. the identity). Lastly, since (c.f. Note 10.39)

$$\overbrace{[I_V]_{\beta}^{\beta'}}^{Q}[I_V]_{\beta'}^{\beta} = [I_V I_V]_{\beta'}^{\beta'} = [I_V]_{\beta'}^{\beta'} = I_n,$$

$$[I_V]_{\beta'}^{\beta}\underbrace{[I_V]_{\beta}^{\beta'}}_{Q} = [I_V I_V]_{\beta}^{\beta} = [I_V]_{\beta}^{\beta} = I_n,$$

we have by the uniqueness of matrix inverses that $Q^{-1} = [I_V]_{\beta'}^{\beta}$.

■

- **Example 10.62:** Consider the ordered bases

$$\beta = \{v_1 = (1,0), v_2 = (0,1)\} \quad \text{and} \quad \beta' = \{v_1' = (1,1), v_2' = (-1,1)\}$$

of $\mathbb{R}^2$. Let us compute the change of bases matrices $Q = [I_V]_{\beta}^{\beta'}$ and $R = Q^{-1} = [I_V]_{\beta'}^{\beta}$. Let's start with R since that's easier. Notice that the matrix R

$$\text{takes } [v_1']_{\beta'} = \begin{pmatrix}1\\0\end{pmatrix} \text{ to } [v_1']_{\beta} = \begin{pmatrix}1\\1\end{pmatrix} \quad \text{and} \quad \text{takes } [v_2']_{\beta'} = \begin{pmatrix}0\\1\end{pmatrix} \text{ to } [v_2']_{\beta} = \begin{pmatrix}-1\\1\end{pmatrix}.$$

In other words, writing $R = \begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}$ we must have that

$$\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}b_{11}\\ b_{21}\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}b_{12}\\ b_{22}\end{pmatrix} = \begin{pmatrix}-1\\1\end{pmatrix}.$$

From this we can straight away read off the values of the entries of R (i.e. the b_{ij}'s) to get that

$$R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Computing Q is also easy: simply set $Q = R^{-1}$ and compute the inverse of R. However, we don't have advanced tools to compute inverses just yet, so at the moment we can do this via systems of equations as follows.

First we ask: what are the representations of the bases vectors of β with respect to β'. In other words, if we write

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \end{pmatrix} + b\begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} = c\begin{pmatrix} 1 \\ 1 \end{pmatrix} + d\begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

what are the values of a, b, c, d? The answer is $a = 0.5$, $b = -0.5$, $c = 0.5$, and $d = 0.5$, we let you verify the details. Hence Q

$$\text{takes } [v_1]_\beta = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ to } [v_1]_{\beta'} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}$$

$$\text{and takes } [v_2]_\beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ to } [v_2]_{\beta'} = \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}.$$

Hence, writing $Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ we must have that

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

from which we can read off the value of Q:

$$Q = \underbrace{\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1}}_{R^{-1}} = \begin{pmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{pmatrix}.$$

- **Note 10.63:** In general, if you want to change bases from β to $\beta' = \{v_1, \dots, v_n\}$ and you can write the column vector representation $[v_i]_\beta$ of each v_i in terms of β, then the R in the above example is obtained by simply making $[v_i]_\beta$ its columns and setting $Q = R^{-1}$. As the previous note illustrated this is often easy to do when $\beta = F^n$, and so when changing bases from β'' to β' one option is to pass through $\beta = F^n$ to make life easier (i.e. change coordinates $\beta'' \to \beta = F^n \to \beta'$).

- **Theorem 10.64:** Suppose that $T : V \to V$ is a linear map and suppose that β and β' are ordered bases for V. Let $Q = [I_V]_\beta^{\beta'}$. Then

$$[T]_{\beta'}^{\beta'} = Q[T]_\beta^\beta Q^{-1}. \tag{10.65}$$

Proof: We have that

$$[T]_{\beta'}^{\beta'} = [I_V T I_V]_{\beta'}^{\beta'} = [I_V]_\beta^{\beta'}[T]_\beta^\beta [I_V]_{\beta'}^\beta = Q[T]_\beta^\beta Q^{-1}.$$

■

- **Note 10.66:** In (10.65) we say that we changed bases for the matrix representation of T from β to β'. In general an equation of the form (10.65) can be interpreted as that the matrix on the left-hand side and the matrix stuck in between Q and Q^{-1} on the right-hand side represent the same linear map but in different coordinates/bases (this is just an interpretation). Hence they are "similar" in a sense. Thus mathematicians came up with the following definition.

- **Definition 10.67:** We say that two $n \times n$ matrices A and B are **similar** if there exists an $n \times n$ invertible matrix Q such that

$$A = QBQ^{-1}.$$

Notice that by multiplying through this equation on the left and right by Q^{-1} and Q respectively gives that $B = Q^{-1}AQ$ and thus is a symmetric property (the interested reader may also verify that it is transitive).

- **Example 10.68:** Let us write down the matrix for the "reflection map" T across the line $y = -x$ in $\mathbb{R}^2$ with respect to the standard basis. Consider the bases β and β' in Example 10.62. Off the bat, it may not be clear how to write down the matrix for T in the standard basis β, but it is easy to do so in β' because T simply takes v_1' to $-v_1'$ and v_2' to v_2'. On the representation side, $[T]_{\beta'}^{\beta}$

$$\text{takes } [v_1']_{\beta'} = \begin{pmatrix}1\\0\end{pmatrix} \text{ to } [-v_1']_{\beta'} = \begin{pmatrix}-1\\0\end{pmatrix} \quad \text{and} \quad \text{takes } [v_2']_{\beta'} = \begin{pmatrix}0\\1\end{pmatrix} \text{ to } [v_2']_{\beta'} = \begin{pmatrix}0\\1\end{pmatrix}.$$

Hence if we write $[T]_{\beta'}^{\beta'} = \begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}$, we have that

$$\begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}c_{11}\\ c_{21}\end{pmatrix} = \begin{pmatrix}-1\\0\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}c_{12}\\ c_{22}\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}.$$

From this we can read off the entries of $[T]_{\beta'}^{\beta'}$:

$$[T]_{\beta'}^{\beta'} = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}.$$

We already computed $Q = [I_V]_{\beta}^{\beta'}$ and its inverse $Q = R^{-1}$ in Example 10.62, and so by (10.65)

$$[T]_{\beta}^{\beta} = \underbrace{\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}}_{Q^{-1}}\underbrace{\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}}_{[T]_{\beta'}^{\beta'}}\underbrace{\begin{pmatrix}0.5 & 0.5\\ -0.5 & 0.5\end{pmatrix}}_{Q} = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}0.5 & 0.5\\ -0.5 & 0.5\end{pmatrix} = \begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix}.$$

Now that we see the answer, we actually see that this clearly has to be the matrix since this reflection takes $(1,0)$ to $(0,-1)$ and $(0,1)$ to $(-1,0)$ in the standard basis β. So the theory is right! An example of where the final answer will be less obvious is if you instead try to redo the above to reflect across a more general line $y = mx$.

11 Row and Column Operations

11.1 Row/Column Operations and Elementary Matrices

- **Note 11.1:** Suppose we want to solve the linear system of equations

$$2x + 2y = 2,$$

$$x + 3y = 3.$$

From the perspective of linear algebra, this can be reformulated and solved as follows:

$$\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \Longrightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}^{-1}\begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \Longrightarrow \quad \text{Voila!} \tag{11.2}$$

However, this is not helpful since we don't have good ways of computing inverses of matrices just yet. So instead, we can do this as follows:

$$\begin{matrix} 2x & +2y & = 2 \\ x & +3y & = 3 \end{matrix} \Longrightarrow \begin{matrix} x & +y & = 1 \\ x & +3y & = 3 \end{matrix} \Longrightarrow \begin{matrix} x & +y & = 1 \\ & 2y & = 2 \end{matrix} \Longrightarrow \begin{matrix} x & +y & = 1 \\ & y & = 1 \end{matrix}$$

$$\Longrightarrow \begin{matrix} x & & = 0 \\ & y & = 1 \end{matrix} \Longrightarrow \text{Voila!}$$

Notice the operations that we did: we either multiplied a row through by a constant or subtracted one row from another. Observe that we could have ignored the x's and y's floating around in the above calculation and consider everything as entries of a matrix for writing efficiency. We remark that a slight modification of this algorithm can be used to compute the inverse matrix in (11.2). Thus, as we'll see, such operations lie at the heart of solving systems of linear equations and computing inverses of matrices.

- **Definition 11.3:** Suppose that A is an m by n matrix. The following are called **elementary row operations of type 1, 2, and 3** respectively:

    1. Interchanging two rows.
    2. Multiplying a row by a scalar.
    3. Adding a scalar multiple of a row to another row.

    Elementary column operations of type 1, 2, and 3 are defined exactly the same way but instead by doing the above operations on columns.

- Surprisingly, for any matrix A it's possible to perform elementary row/column operations by multiplying it by suitable "elementary matrices:"

- **Definition 11.4:** An $n \times n$ matrix E is called an **elementary matrix of type 1, 2, or 3** if it is obtained by performing one operation of type 1, 2, or 3 to the identity matrix I_n.

- **Note 11.5:** If E is an elementary matrix obtained by performing a row operation of type $k = 1,2,3$ to the identity matrix, then multiplying it by A from the left performs that same row operation on A. The same goes for column operations, but then you need to multiply it by A from the right. For instance, suppose that

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}$$

If we take the 4×4 matrix E_1 obtained from the identity matrix by switching 1st and 3rd rows, then multiplying this by A from the left switches the 1st and 3rd rows of A:

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{E_1} \underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}}_{A} = \begin{pmatrix} 11 & 12 & 13 & 14 & 15 \\ 6 & 7 & 8 & 9 & 10 \\ 1 & 2 & 3 & 4 & 5 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}.$$

Similarly, if we take the 5×5 matrix E_2 obtained from the identity matrix by adding -2 times the 4th column to the 2nd column, then multiplying this by A from the right adds -2 times the 4th column of A to the 2nd column of A:

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{E_2} = \begin{pmatrix} 1 & 2-2\cdot 4 & 3 & 4 & 5 \\ 6 & 7-2\cdot 9 & 8 & 9 & 10 \\ 11 & 12-2\cdot 14 & 13 & 14 & 15 \\ 16 & 17-2\cdot 19 & 18 & 19 & 20 \end{pmatrix}.$$

- **Proposition 11.6:** Every elementary matrix is invertible, and its inverse is an elementary matrix of the same type.

  Proof: This is trivial: if E is an elementary matrix obtained from the identity matrix by performing an elementary operation of type k, then undoing that elementary operation will be an elementary operation of the same type, which is equivalent to multiplying by an elementary matrix of the same type to get back to the identity matrix.

  ■

11.2 Matrix rank

- Considering that we've defined rank for linear maps, it's not surprising that we can define rank for matrices as well since the latter represent the former. To define the rank of the latter, we fix particular linear maps that they represent:

- **Definition 11.7:** Suppose that $A \in M_{m\times n}(F)$ is an m by n matrix. Let $L_A : \mathbb{R}^n \to \mathbb{R}^m$ denote the linear map $L_A v = Av$ where Av denotes matrix vector multiplication. We define the **rank** of A as

$$\operatorname{rank} A = \operatorname{rank} L_A.$$

- In the above definition, it turns out that you don't necessarily need to use the linear map L_A that the matrix A represents: you can use any other linear map that A represents. This is the content of the following theorem.

- **Theorem 11.8:** Suppose that $T : V \to W$ is a linear map between finite-dimensional vector spaces and that β and γ are ordered bases for V and W respectively. Then

$$\operatorname{rank} T = \operatorname{rank}[T]_\beta^\gamma. \tag{11.9}$$

Proof: The book leaves this as an exercise: we'll sketch the argument. Let $m = \dim V$ and $n = \dim W$ and let $A = [T]_\beta^\gamma$. If you let $\{w_1, \dots, w_k\}$ be a basis of $R(T)$, then $\{[w_1]_\gamma, \dots, [w_k]_\gamma\}$ will be a linearly independent subset of $R(L_A)$ and hence $\dim R(T) \leq \dim R(L_A)$. Similarly you can argue that $\dim R(T) \geq \dim R(L_A)$. Hence $\dim R(T) = \dim R(L_A)$, which is equivalent to (11.9).

■

- **Lemma 11.10:** A $n \times n$ matrix A is invertible if and only if it is of rank n (i.e. $\operatorname{rank} A = n$).

Proof: This follows immediately from the fact that A represents $L_A : \mathbb{R}^n \to \mathbb{R}^n$ and L_A is invertible if and only if it has rank n (or equivalently is surjective) – see Theorem 10.57 and Theorem 10.17.

■

- **Theorem 11.11:** Suppose that $A \in M_{m\times n}(F)$. Suppose also that $P \in M_{m\times m}(F)$ and $Q \in M_{n\times n}(F)$ are invertible matrices. Then

1. $\operatorname{rank}(PA) = \operatorname{rank} A$,
2. $\operatorname{rank}(AQ) = \operatorname{rank} A$,
3. $\operatorname{rank}(PAQ) = \operatorname{rank} A$.

Proof: This will be a HW problem. ■

- Using our theory above, we have a neat way of proving that elementary row and column operations preserve rank:

- **Corollary 11.12:** If we apply an elementary row/column operation to a matrix A to get a matrix $\tilde{A}$, then $\operatorname{rank} A = \operatorname{rank} \tilde{A}$.

Proof: By our discussion in Note 11.5, there exists an elementary matrix E such that either $\tilde{A} = EA$ or $\tilde{A} = AE$ (depending on if we're doing a row or column operation on A). Then by Theorem 11.11,

$$\operatorname{rank} \tilde{A} = \operatorname{rank} EA = \operatorname{rank} A \quad \text{if} \quad \tilde{A} = EA$$

$$\text{or} \quad \operatorname{rank} \tilde{A} = \operatorname{rank} AE = \operatorname{rank} A \quad \text{if} \quad \tilde{A} = AE.$$

■

- **Theorem 11.13:** The rank of a matrix equals the dimension of the span of its columns. This is equivalent to the maximum number of linearly independent columns that you can choose from A.

Proof: Let $A \in M_{m\times n}(F)$. By Proposition 10.11,

$$\operatorname{rank} A = \dim R(L_A) = \dim \operatorname{span}\{L_A(e_1), \dots, L_A(e_n)\}.$$

Letting c_j denote the j^{th} column of A, by multiplying out Ae_j it's not hard to see that $c_j = Ae_j = L_A(e_j)$. Hence the above equation gives that

$$\operatorname{rank} A = \dim \operatorname{span}\{c_j, \dots, c_j\},$$

which proves the first statement of the theorem.

For the second statement, by Corollary 9.38 we can remove vectors from $\{c_j, \dots, c_j\}$ to get a basis $\beta = \{\tilde{c}_j, \dots, \tilde{c}_k\}$ of $R(L_A)$ with $k = \dim R(L_A) = \operatorname{rank} A$. Any linear independent list of columns has to have less than k vectors since β is a basis. On the other hand, $\{\tilde{c}_j, \dots, \tilde{c}_k\}$ is a linear independent list of columns. Hence the maximum number of linearly independent columns is k, or in other words the rank of A.

■

- **Theorem 11.14:** Suppose that $A \in M_{m\times n}(F)$ is a matrix with rank r. Then $r \leq m$ and $r \leq n$ and you can apply a finite number of row and column operations to turn A into the form

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

where the O_1, O_2, O_3 are zero matrices (draw D out!). Hence there exist elementary matrices $E_1, \dots, E_k$ and $G_1, \dots, G_j$ such that

$$D = \underbrace{E_1 \dots E_k}_{\text{Call this } B} A \underbrace{G_1 \dots G_j}_{\text{Call this } C}$$

$$D = BAC,$$

where observe that both B and C are invertible since they are products of elementary matrices. Observe that this implies that

$$\operatorname{rank} D = \operatorname{rank} A.$$

Proof: The fact that A can be turned into D by applying a finite number of row and column operations is proved by an algorithm. The rest of the conclusions are immediate (the last one follows from Corollary 11.12). Instead of writing the algorithm out explicitly, we will do an example below from which it will be straightforward to understand how it works in the general case.

■

- **Example 11.15:** Here we illustrate the algorithm used to prove Theorem 11.14:

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 4 \\ 0 & 3 & 5 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 4 \\ 3 & 0 & 5 & 1 \end{pmatrix} \to \begin{pmatrix} 2 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 5 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 5 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -5 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2.5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = D.$$

- **Note 11.16:** We remark that as the algorithm progresses, it eventually arrives at a matrix of the form

$$D_{\text{Unfinished}} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right).$$

  Then you continue the algorithm inductively on A'.

- We now present some applications of Theorem 11.14, starting with invertible matrices:

- **Corollary 11.17:** Every invertible matrix $A \in M_{m\times m}(F)$ can be written as a product of elementary matrices:

  $$A = E_1 \dots E_l. \tag{11.18}$$

  Remark: You can think of (11.18) as decomposing A into "simpler" matrices, though keep in mind that this decomposition is not unique.

  Proof: By Theorem 11.14 we can write $D = BAC$ where D is as stated there and $B = E_1 \dots E_k$ and $C = G_1 \dots G_j$ are products of elementary matrices (and hence B and C are invertible). Since A is invertible, by Lemma 11.10 we have that $\operatorname{rank} A = n$. By looking at the form of D, we see that this implies that $D = I_n$. So

  $$A = B^{-1}DC^{-1} = (E_1 \dots E_k)^{-1} I_n \left(G_1 \dots G_j\right)^{-1} = E_k^{-1} \dots E_1^{-1} G_1^{-1} \dots G_1^{-1}.$$

  By Proposition 11.6, inverses of elementary matrices are also elementary matrices. Hence this prove the corollary.

  ■

- Theorem 11.14 is also useful to conclude properties of general matrices too:

- **Corollary 11.19:** Suppose that $A \in M_{m\times n}(F)$ is a matrix. Then

  1. $\operatorname{rank} A^{\mathsf{T}} = \operatorname{rank} A$.

2. $\operatorname{rank} A$ equals the dimension of the span of its rows. This is equivalent to the maximum number of linearly independent rows that you can choose from A.
3. Dimensions of the span of the rows and the span of the columns of A are equal and both equal to $\operatorname{rank} A$.

Proof: First let's prove 1). By Theorem 11.14 we can write $D = BAC$ where D is as stated there and B and C are both finite products of elementary matrices. Then

$$A = B^{-1}DC^{-1}$$

$$\implies \quad A^{\top} = (B^{-1}DC^{-1})^{\top} = (C^{-1})^{\top}D(B^{-1})^{\top}.$$

Since B and C are products of elementary matrices, the inverse of elementary matrices are elementary matrices, and transposes of elementary matrices (check this last claim!), we have that both $(C^{-1})^{\top}$ and $(B^{-1})^{\top}$ are also products of elementary matrices. Hence by Corollary 11.12,

$$\operatorname{rank} A^{\top} = \operatorname{rank}[(C^{-1})^{\top}D^{\top}(B^{-1})^{\top}] = \operatorname{rank} D^{\top}.$$

By looking at the form of D, we see that the maximum number of linearly independent columns in D and $D^{\top}$ are the same and hence $\operatorname{rank} D = \operatorname{rank} D^{\top}$. Since $\operatorname{rank} D = \operatorname{rank} A$, the above equation indeed gives us that $\operatorname{rank} D = \operatorname{rank} A$.

Part 2) follows from observing that rows of A are columns of $A^{\top}$ and then applying Theorem 11.13. Part 3) follows from part 2) and Theorem 11.13.

■

Corollary 11.20: Suppose that $T : V \to W$ and $U : W \to Z$ are linear maps over finite dimensional vector spaces. Suppose also that A and B are matrices such that the product AB makes sense. Then

1. $\operatorname{rank} UT \leq \operatorname{rank} U$
2. $\operatorname{rank} UT \leq \operatorname{rank} T$
3. $\operatorname{rank} AB \leq \operatorname{rank} A$
4. $\operatorname{rank} AB \leq \operatorname{rank} B$.

Proof: We'll prove 1), then 3), then 4), and then 2). To prove 1), observe that

$$R(UT) = \{UT(v) : v \in V\} = \big\{U\big(T(v)\big) : v \in V\big\} \subseteq \{U(w) : w \in V\} = R(U).$$

Hence $\dim R(UT) \leq \dim R(U)$, which is equivalent to 1). Part 3) follows from 1) simply by representing 1) in matrix form when $U = L_A$ and $T = L_B$.

Let's prove 4). From 3) and Corollary 11.19 we have that

$$\operatorname{rank} AB = \operatorname{rank}(B^{\top}A^{\top})^{\top} = \operatorname{rank}(B^{\top}A^{\top}) \leq \operatorname{rank} B^{\top} = \operatorname{rank} B.$$

Part 2) follows from 3) simply by representing 2) in matrix form.

■

- **Note 11.21:** Understanding the rank of a matrix is important in linear algebra and its applications, such as in differential geometry. Our theory above allows us to compute this efficiently. In particular, to compute the rank of a matrix A, you can apply row and column operations to turn it into the form D described in Theorem 11.14, from where you can read off the rank of A as r. You don't actually have to go as far as D. For instance, if you perform elementary operation as follows:

$$A = \begin{pmatrix} 2 & 1 & 3 & 1 \\ 1 & 2 & 3 & 4 \\ 5 & 4 & 9 & 6 \end{pmatrix} \xrightarrow{\text{some row/column operations}} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -6 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

then from here you can see that the biggest list of linearly independent columns that you can pull out of the matrix on the right is 2. Hence the rank of the matrix on the right is 2. Since you only performed elementary operations to A to get the matrix on the right, this shows that the rank of A is 2 as well.

11.3 Computing Matrix Inverses

- We now study an efficient algorithm for computing inverses of matrices. We start with the following notation:

- **Notation 11.22:** Suppose that $A \in M_{m\times n}(F)$ and $B \in M_{m\times k}(F)$ are matrices. The augmented matrix $(A|B)$ is the m by $(n+k)$ matrix given by

$$(A|B) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_{m1} & \cdots & b_{mk} \end{array}\right).$$

The dividing line " | " in the middle doesn't mean anything: it's just a psychological placeholder to remind us that the A and B typically represent different quantities in application.

- **Note 11.23:** Suppose that A is an invertible $n \times n$ matrix. By Corollary 11.17 we can write A as a product of elementary matrices:

$$A = E_1 \dots E_k.$$

Multiplying both sides by $E_k^{-1} \dots E_1^{-1}$ on the left gives:

$$E_k^{-1} \dots E_1^{-1} A = I_n,$$

where recall that each E_i^{-1} is also an elementary matrix (observe that this implies that $A^{-1} = E_k^{-1} \dots E_1^{-1}$). An interpretation of this equation is that it's possible to apply row operations to A to get I_n.

Now comes the marvelous observation. Consider the augmented matrix $(A|I_n)$ and let us perform the elementary row operations of $E_k^{-1} \dots E_1^{-1}$ to $(A|I_n)$. As a quick aside: it's an elementary exercise to check that for any matrices R, M, N,

$$R(M|N) = (RM|RN) \tag{11.24}$$

whenever all operations involved make sense. Hence we get that

$$E_k^{-1} \dots E_1^{-1}(A|I_n) = (E_k^{-1} \dots E_1^{-1}A|E_k^{-1} \dots E_1^{-1}I_n) = (I_n|A^{-1}).$$

So here is the amazing algorithm: to compute the inverse of an invertible matrix A, take the augmented matrix $(A|I_n)$, apply elementary row operations to it to turn the A into I_n, in which case the augmented matrix will become$(I_n|B)$, and then read off the "B" to get the inverse matrix (i.e. $B = A^{-1}$)!

The above reasoning is reversible. Suppose A is $n \times n$ and you were able to perform elementary row operations $G_1 \dots G_j$ to turn the augmented matrix $(A|I_n)$ into $(I_n|B)$ for some B:

$$G_1 \dots G_j(A|I_n) = (I_n|B),$$

Applying (11.24) to the left-hand side gives that

$$\Big(\big(G_1 \dots G_j\big)A\Big|\big(G_1 \dots G_j\big)I_n\Big) = (I_n|B).$$

Comparing the portions "(here|)" implies that $\big(G_1 \dots G_j\big)A = I_n$ and hence $\big(G_1 \dots G_j\big) = A^{-1}$. Comparing the portions "(|here)" then gives that $A^{-1} = \big(G_1 \dots G_j\big)I_n = B$. In other words, if we were able to do this, we can conclude that A is invertible and similarly as before we can read of its inverse by looking at B.

The algorithm will fail if you try to perform it on a noninvertible matrix A (as we know it should!). In particular, it fails because no matter how hard you try it will be impossible to turn the A into I_n via row operations in the augmented matrices $(A|I_n)$ and $(I_n|B)$.

- **Example 11.25:** Suppose we want to find the inverse of

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}. \tag{11.26}$$

To do this, we do

$$(A|I_2) = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 3 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 0 & 3 & -1 \\ 0 & 1 & -2 & 1 \end{array}\right).$$

So the inverse is

$$A^{-1} = \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix}. \tag{11.27}$$

You can check this by multiplying (11.26) and (11.27) out and seeing that you indeed get I_2.

11.4 Systems of Equations

- As you have seen in the homework, matrix equations and systems of linear equations are intimately connected. Precisely, one can be used to rewrite the other. Moreover, inverting a matrix can be seen as solving an enormous system of linear equations and hence it's not surprising that a slight modification of our algorithm for inverting matrices can be adapted to solving general systems of equations. We will demonstrate this with an example and then discuss what happens in the general case.

- **Example 11.28:** Suppose we want to solve the system of linear equations:

$$\begin{aligned} x_1 - 4x_2 - x_3 + x_4 &= 3, \\ 2x_1 - 8x_2 + x_3 - 4x_4 &= 9, \\ -x_1 + 4x_2 - 2x_3 + 5x_4 &= -6. \end{aligned} \tag{11.29}$$

Here we have three equations for four unknowns. This is equivalent to solving a matrix equation of the form $Ax = b$, in particular

$$\underbrace{\begin{pmatrix} 1 & -4 & -1 & 1 \\ 2 & -8 & 1 & -4 \\ -1 & 4 & -2 & 5 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 3 \\ 9 \\ -6 \end{pmatrix}}_{b}.$$

We cannot solve this simply by inverting the matrix since A isn't square. The matrix A is called the **coefficient matrix**. If this system has a solution, we say that it is **consistent**, otherwise we call it **inconsistent**. Observe that whether this system is consistent or not is equivalent to asking if the vector on the right-hand side is in the range of L_A. To solve the system (11.29), let's subtract two times the first equation from the second equation and add the first equation to the third equation to get the system:

$$\begin{aligned} x_1 - 4x_2 - x_3 + x_4 &= 3, \\ 3x_3 - 6x_4 &= 3, \\ -3x_3 + 6x_4 &= -3. \end{aligned}$$

Notice that x_1, x_2, x_3, x_4 are a solution of (11.29) if and only if they are a solution to this system. In that sense, these two systems are **equivalent**: they have the same **solution set** (i.e. set of solutions)! Notice that the operation we did with the equations in (11.29) is equivalent to performing row operations on the augmented matrix:

$$(A|b) = \left(\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{array}\right) \xrightarrow{\text{two row operations}} \left(\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & -3 & 6 & -3 \end{array}\right).$$

Hence let's continue solving this system by instead performing row operations on this augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & -3 & 6 & -3 \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & -1 & 2 & -1 \end{array}\right) \to \underbrace{\left(\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\text{(Row) echelon form}}$$

$$\to \underbrace{\left(\begin{array}{cccc|c} 1 & -4 & 0 & -1 & 4 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\text{(Row) reduced echelon form}}$$

(either echelon form requires putting a 1 on the first row to the let of the " | " if it is not the zero matrix). This last augmented matrix represents the system of equations

$$\begin{aligned} x_1 - 4x_2 \qquad - x_4 &= 4, \\ x_3 - 2x_4 &= 1. \\ 0 &= 0. \end{aligned} \tag{11.30}$$

Hence we can set x_2 and x_4 to be anything (they're called **free variables**) and set $x_1 = 4 + 4x_2 + x_4$ and $x_3 = 1 + 2x_4$. If we let $x_2 = a$ and $x_4 = b$ denote x_2's and x_4's arbitrary values, in vector notation we can write the solution set as:

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_3 \end{pmatrix} \text{ is a solution to } Ax = b \right\} = \left\{ \begin{pmatrix} 4 + 4a + b \\ a \\ 1 + 2b \\ b \end{pmatrix} = \underbrace{\begin{pmatrix} 4 \\ 0 \\ 1 \\ 0 \end{pmatrix}}_{x_0} + a \underbrace{\begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{v_1} + b \underbrace{\begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}}_{v_2} : a, b \in \mathbb{R} \right\}.$$

The algorithm that we just did to solve this system is called **Gaussian elimination**. Let's make a few remarks about the process. If the system was inconsistent to start with, you would see this in the algorithm by encountering an impossible statement, such the third equation in (11.30) instead being something like $0 = 12$. Next, we solved the system by going from the reduced echelon form (i.e. wrote (11.30)), but you can also do this from the echelon form but you'll have to do some extra substitutions. Thirdly, although we won't prove this, you should be aware that the reduced echelon form of a matrix A is unique.

The 1's that have only zeros to their left in the echelon form (and also zeros above them in the reduced echelon form) are called **pivots**, which represent variables that are not free variables. We note that pivots are not allowed to be to the right of the " | " in the augmented matrix.

Since $b \neq 0$ our system $Ax = b$ is called **nonhomogeneous**, in which case observe that its solution set takes the form

$$\{x \in \mathbb{R}^4 \text{ is a solution to } Ax = b\} = \{x_0 + av_1 + bv_2 : a, b \in \mathbb{R}\} \tag{11.31}$$

where observe that x_0 is a solution to $Ax = b$ (i.e. $Ax_0 = b$). If $b = 0$, then the system $Ax = 0$ is called **homogeneous**, and it's not hard to see by looking back at our algorithm that we would get that

$$\{x \in \mathbb{R}^4 \text{ is a solution to } Ax = 0\} = \{av_1 + bv_2 : a, b \in \mathbb{R}\}.$$

Notice that this is equivalent to $N(L_A)$ and that v_1 and v_2 are a basis for $\ker L_A$. So we can rewrite this as

$$\{x \in \mathbb{R}^4 \text{ is a solution to } Ax = b\} = x_0 + \{x \in \mathbb{R}^4 \text{ is a solution to } Ax = 0\} = x_0 + N(L_A).$$

This is an important principle!

The above discussion generalizes directly to bigger systems of equations, which you'll have practice doing computations on in the homework. In particular, notice that in general the algorithm tells us that

$$\dim \operatorname{dom} L_A = \text{Width of } A = \#\text{ of pivots} + \underbrace{\#\text{ of free variables}}_{\dim N(L_A)}$$

and so by the dimension theorem we get the important principle that

$$\#\text{ of pivots} = \dim \operatorname{dom} L_A - \dim N(L_A) = \dim R(L_A) = \operatorname{rank} A.$$

$$\implies \quad \#\text{ of pivots} = \operatorname{rank} A.$$

12 Determinants

- We transition to the study of determinants. I'm going to warn you: this is a controversial subject since it is ***extremely*** useful while on the other hand it is not intuitive at first. Their uses come up in the definition of eigenvalues, in the proof that matrix groups are Lie groups, Wronskians in differential equations, change of variables for multiple integrals, implicit function theorem, etc.

  So what are determinants? At the basic level, they "determine" whether a square matrix is invertible or equivalently if a system with a square coefficient matrix is solvable for any right-hand side (c.f. Note 11.1). We will study the determinant as follows, we will define it algebraically, and then study its properties and geometric meaning. First consider the 1 by 1 matrix $[a]$. Clearly it is invertible if and only if $a \neq 0$. So we set

  $$\det[a] = a.$$

  Next, in the homework you proved that a square 2 by 2 matrix

  $$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

  is invertible if and only if $a(d) - b(c) \neq 0$. Hence we set

  $$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a \cdot \det[d] - b \cdot \det[c].$$

  If you did a similar exercise for 3 by 3 matrices, you would get that

  $$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

  is invertible if and only if $a(ei - fh) - b(di - fg) + c(dh - eg) \neq 0$. Hence we set

  $$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \cdot \det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \cdot \det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \cdot \det\begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

  See a pattern!?

- **Definition 12.1:** Define the determinant of 1 by 1 matrices by $\det[a] = a$. We define the determinant of bigger square matrices inductively as follows. Suppose we defined the algebraic expression for the determinant of $m \times m$ matrices. We define the determinant of a $(m+1) \times (m+1)$ matrix A by

  $$\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m+1} \\ \vdots & \ddots & \vdots \\ a_{m+1,1} & \cdots & a_{m+1,m+1} \end{pmatrix} = a_{1,1} \det\overbrace{[A \text{ but remove first row and first column}]}^{\text{an } m\times m \text{ matrix}}$$

  $$-a_{1,2} \det[A \text{ but remove first row and second column}] + \cdots$$

  $$+(-1)^m a_{1,m+1} \det\big[A \text{ but remove first row and } (m+1)^{\text{th}} \text{ column}\big] \tag{12.2}$$

$$= a_{1,1}\det\tilde{A}_{1,1} - a_{1,2}\det\tilde{A}_{1,2} + \cdots + (-1)^m a_{1,m+1}\det\tilde{A}_{1,m+1} = \sum_{j=1}^{k+1}(-1)^{j-1}a_{1,j}\det\tilde{A}_{1,j}.$$

- **Definition 12.3:** Suppose that $T : V \to V$ is a linear map. Let β be a basis for V. We define the determinant of T as the determinant of its representation with respect to β:

$$\det T = \det[T]_\beta^\beta.$$

In the homework you will prove that this definition is well-defined. In other words, if you choose a different basis $\tilde{\beta}$ of V, then you will get the same answer for the determinant of T (tip: first read this chapter before trying to prove this).

12.1 Determinants and Elementary Row Operations

- We will now study how determinants behave under elementary row operations.
- **Theorem 12.4:** Suppose that $A \in M_{m\times m}(F)$ is a square matrix and that B is obtained by switching two rows in A (i.e. an elementary row operation of type 1). Then

$$\det B = -\det A. \tag{12.5}$$

Proof: We will prove this by induction. It is obviously true for 1×1 matrices and you can also check it directly for 2×2. Now suppose that we proved it for $k \times k$ matrices. Let A be a $(k+1)\times(k+1)$ matrix. If the two rows that we're switch do not involve the first row, then (12.5) follows immediately from (12.2) and our inductive hypothesis. So suppose that one of the rows that we're switching is the first row. It's actually enough to prove (12.5) when we're switching the first and the second row for the following reason. If we're switching the first and the j^{th} row, this can be seen as first switching the first and the second row, then switching the second row and the j^{th} row (which we already argued causes a sign flip in the determinant), and finally switching the first and second row again. Overall we did an odd number of row switches (three in fact), and hence we get (12.5).

So suppose we're switching the first and second row. By (12.2),

$$\det A = a_{1,1}\det\tilde{A}_{1,1} - \cdots + (-1)^k a_{1,k+1}\det\tilde{A}_{1,k+1} = \sum_{i=1}^{k+1}(-1)^{i-1}a_{1,i}\det\tilde{A}_{1,i}$$

Letting $\widetilde{\left(\tilde{A}_{1,i}\right)}_{2,j}$ denote removing the second row and j^{th} column <u>of A</u> from $\tilde{A}_{1,i}$, applying (12.2) again to each $\tilde{A}_{1,i}$ in the above equation gives

$$\det A = \sum_{i=1}^{k+1}(-1)^{i-1}a_{1,i}\sum_{\substack{j=1\\ j\neq i}}^{k+1}(-1)^{?}a_{2,j}\det\widetilde{\left(\tilde{A}_{1,i}\right)}_{2,j}$$

$$= \sum_{\substack{i,j\in\{1,\dots,k+1\}\\ i\neq j}} (-1)^{?} a_{1,i}a_{2,j}\det\left(\widetilde{\tilde{A}_{1,\iota}}\right)_{2,j}. \tag{12.6}$$

Expanding B similarly gives

$$\det B = \sum_{\substack{,j\in\{1,\dots,k+1\}\\ i\neq j}} (-1)^{?} a_{2,i}a_{1,j}\det\left(\widetilde{\tilde{A}_{1,\iota}}\right)_{2,j}. \tag{12.7}$$

So the question is how do the $(-1)^{?}$ relate to each other in (12.6) and (12.7). We can figure this out by tracking the signs in the expansions involved in the above two equations diagrammatically as follows. Take any two $i \neq j$ and without loss of generality assume that $i < j$. Then (this is a demonstration, you need to be in class to understand it)

Computing $\det A$ using (12.2) involves

$$\left.\begin{array}{ccccccc} & & \overset{(-1)^{i-1}}{\overbrace{\phantom{a_{1,i}}}} & & & & \\ a_{1,1} & \cdots & a_{1,i} & \cdots & a_{1,j} & \cdots & a_{1,k+1} \\ a_{2,1} & \cdots & a_{2,i} & \cdots & \underbrace{a_{2,j}}_{(-1)^{j-2}} & \cdots & a_{2,k+1} \end{array}\right\} \text{ sign of } a_{1,i}a_{2,j}\det\left(\widetilde{\tilde{A}_{1,\iota}}\right)_{2,j} \text{ is } (-1)^{i+j-3},$$

Computing $\det B$ using (12.2) involves

$$\left.\begin{array}{ccccccc} & & & & \overset{(-1)^{j-1}}{\overbrace{\phantom{a_{2,j}}}} & & \\ a_{2,1} & \cdots & a_{2,i} & \cdots & a_{2,j} & \cdots & a_{2,k+1} \\ a_{1,1} & \cdots & \underbrace{a_{1,i}}_{(-1)^{i-1}} & \cdots & a_{1,j} & \cdots & a_{1,k+1} \end{array}\right\} \text{ sign of } a_{1,i}a_{2,j}\det\left(\widetilde{\tilde{A}_{1,\iota}}\right)_{2,j} \text{ is } (-1)^{i+j-2}.$$

Notice that $(-1)^{i+j-2} = -(-1)^{i+j-3}$. Hence the signs in (12.6) and (12.7) are related as $(-1)^{?} = -(-1)^{?}$. Hence indeed $\det B = -\det A$.

■

- **Corollary 12.8:** If two rows of a square matrix are identical, then the determinant of the matrix is zero.

  Proof: Let A be a square matrix with two identical rows. Switch those two identical rows to get a matrix B. Since those two rows are identical, $B = A$. On the other hand, $\det A = \det B = -\det A$. So indeed $\det A = 0$.

  ■

- **Theorem 12.9:** Suppose that $A \in M_{m\times m}(F)$ is a square matrix. Then for any $a, b \in F$ and any row vectors $v = (v_1, \dots, v_m)$ and $w = (w_1, \dots, w_m)$,

$$\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ av_1 + bw_1 & \cdots & av_m + bw_m \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix} \begin{matrix} \\ \\ i \\ \\ \\ \end{matrix} = a\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ v_1 & \cdots & v_m \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix} + b\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ w_1 & \cdots & w_m \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix}.$$

Remark: This can be thought of as linearity of the determinant in each row <u>when</u> all of the other rows are fixed.

Proof: We prove this similarly to Theorem 12.4: by induction. It's clearly true for 1×1 matrices and you can directly check it for 2×2 matrices. Now suppose that we proved it for $k \times k$ matrices. Let A be a $(k+1) \times (k+1)$ matrix. As in the previous proof, you can similarly argue that the inductive hypothesis already implies that the above equation is true if $i \geq 2$. If $i = 1$, then by (12.2), the left-hand side of the above equation is equal to

$$(av_1 + bw_2)\det\tilde{A}_{1,1} - \cdots + (-1)^k(av_{k+1} + bw_{k+1})\det\tilde{A}_{1,k+1}$$
$$= a\left[v_1\det\tilde{A}_{1,1} - \cdots + (-1)^k v_{k+1}\det\tilde{A}_{1,k+1}\right]$$
$$+b\left[w_1\det\tilde{A}_{1,1} - \cdots + (-1)^k w_{k+1}\det\tilde{A}_{1,k+1}\right].$$

By (12.2), this is precisely the right-hand side of the equation in this theorem (with $m = k+1$).

■

- **Corollary 12.10:** Suppose that $A \in M_{m\times m}(F)$ is a square matrix. If you multiply a row by a constant $c \in F$ to get a matrix B (i.e. an elementary row operation of type 2), then

$$\det B = c\det A.$$

Proof: Fix a row index i. Then this result follows immediately from Theorem 12.9 by setting there $a = c$, $v = (a_{i,1}, \ldots, a_{i,m})$, $b = 0$, and $w = 0$.

■

- **Corollary 12.11:** Suppose that $A \in M_{m\times m}(F)$ is a square matrix. If you add a scalar multiple of a row to another row to get a matrix B (i.e. an elementary row operation of type 3), then

$$\det B = \det A.$$

Proof: Suppose that we're adding $c \in F$ times the j^{th} row to the i^{th} row to get B. By Theorem 12.9,

$$\det B = \det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{i,1} + ca_{j,1} & \cdots & a_{i,m} + ca_{j,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix}$$

$$= \underbrace{\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{i,1} & \cdots & a_{i,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix}}_{\det A} + c\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{j,1} & \cdots & a_{j,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix}.$$

Notice that the last matrix has two identical rows since the i^{th} row is equal to the j^{th} row. Hence by Corollary 12.8 the last matrix is zero. This proves this corollary.

■

- **Corollary 12.12:** If a square matrix $A \in M_{m\times m}(F)$ is not invertible, then $\det A = 0$.

  Proof: If $A \in M_{m\times m}$ is not invertible, then by Lemma 11.10 it does not have (max) rank m. Hence by Corollary 11.19, its row vectors are not spanning. Since there are m rows, this means that they are linearly dependent. This means that one of the rows can be written as a linear combination of the other rows. Letting $r_1, \ldots, r_m$ denote the rows of A, suppose that the first row can be written as a linear combination of the other rows:

  $$r_1 = a_2 r_2 + \cdots + a_m r_m.$$

  since the proof in the other cases is similar. Then by Theorem 12.9 and Corollary 12.8,

  $$\det A = \det\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \det\begin{pmatrix} a_2 r_2 + \cdots + a_m r_m \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = a_2 \underbrace{\det\begin{pmatrix} r_2 \\ r_2 \\ \vdots \\ r_m \end{pmatrix}}_{0} + \cdots + a_m \underbrace{\det\begin{pmatrix} r_m \\ r_2 \\ \vdots \\ r_m \end{pmatrix}}_{0} = 0.$$

  ■

- It is also true that if $A \in M_{m\times m}(F)$ is invertible, then $\det A \neq 0$ (i.e. the converse of the above corollary). We will prove that in Corollary 12.18 below. Recall that this is partly what we're aiming for: the determinant "determines" whether a matrix is invertible.

12.2 Properties of Determinants

- Just as we studied rank of square matrices by decomposing them into elementary matrices, we will use a similar technique to study the properties of determinants.

- **Lemma 12.13:** The determinant of the identity matrix I_m is one.

  Proof: We prove this by induction on m. It's obviously true for $m = 1$: $\det[1] = 1$. Now suppose that we proved the lemma for I_k, let's prove it for I_{k+1}. By (12.2) and the inductive hypothesis,

  $$\det I_{k+1} = 1 \cdot \det I_k - 0 \cdot (\text{something}) + \cdots + (-1)^k 0 \cdot (\text{something}) = 1.$$

  ■

- **Lemma 12.14:** Suppose that $E_1, E_2, E_3 \in M_{m\times m}(F)$ are elementary matrices of types 1, 2, and 3 respectively. Suppose E_2 is associated with the elementary operation of multiplying a row by $c \in F$. Then

$$\det E_1 = -1, \quad \det E_2 = c, \quad \det E_3 = 1.$$

Proof: We will prove that $\det E_2 = c$, the others are proved similarly. By definition E_2 is obtained by multiplying a row of I_m by c, and hence by Corollary 12.10 and Lemma 12.13 $\det E_2 = c \cdot 1 = c$.

■

- **Lemma 12.15:** Suppose that $E, B \in M_{m\times m}(F)$ where E is an elementary matrix. Then

$$\det(EB) = \det E \det B.$$

Proof: First suppose that E is an elementary matrix of type 1. Hence EB switches two rows of B and thus by Theorem 12.4, $\det(EB) = -\det B$. On the other hand, by Lemma 12.14, $\det E = -1$ and so $\det E \det B = -\det B$. So indeed $\det(EB) = \det E \det B$. The proofs in the cases when E is an elementary matrix of types 2 or 3 are handled similarly.

■

- **Theorem 12.16:** Suppose that $A, B \in M_{m\times m}(F)$ are square matrices. Then

$$\det(AB) = \det A \det B. \tag{12.17}$$

Proof: First suppose that both A and B are invertible. Then by Corollary 11.17 we can decompose A and B into elementary matrices: $A = E_1 \dots E_k$ and $B = G_1 \dots G_j$. Applying Lemma 12.15 many times gives that

$$\det(AB) = \det\big(E_1 \dots E_k G_1 \dots G_j\big) = \det E_1 \dots \det E_k \det G_1 \dots \det G_j$$

$$= \det(E_1 \dots E_k) \det\big(G_1 \dots G_j\big) = \det A \det B.$$

Now suppose that one of A or B is not invertible and hence not of (max) rank m. By Corollary 11.20 we get that AB is not of (max) rank m and hence not invertible. Thus by Corollary 12.12 both sides of (12.17) are simply equal to zero.

■

- **Corollary 12.18:** If $A \in M_{m\times m}(F)$ is invertible, then $\det A \neq 0$ and

$$\det(A^{-1}) = \frac{1}{\det A}. \tag{12.19}$$

Proof: From

$$\det A \det(A^{-1}) = \det(AA^{-1}) = \det I_m = 1,$$

we get that $\det A$ cannot be zero. By solving for $\det(A^{-1})$, we get (12.19).

■

- Now we carry out a similar program to study transposes.
- **Lemma 12.20:** If E is an elementary matrix, then $\det E = \det(E^\top)$.

  Proof: This is proved very similarly to Lemma 12.14, we leave the details to the reader. ■

- **Theorem 12.21:** For any square matrix $A \in M_{m\times m}(F)$, $\det(A^\top) = \det A$.

  Proof: First suppose that A is invertible and hence we can decompose it into elementary matrices: $A = E_1 \dots E_k$. Then by Lemma 12.20,

$$\det A^\top = \det[(E_1 \dots E_k)^\top] = \det(E_k^\top \dots E_1^\top) = \det(E_k^\top) \dots \det(E_1^\top) = \det(E_k) \dots \det(E_1)$$
$$= \det(E_1) \dots \det(E_k) = \det(E_1 \dots E_k) = \det A.$$

  Now suppose that A is not invertible and hence not of (max) rank m. By Corollary 11.19 we have that $\operatorname{rank} A = \operatorname{rank}(A^\top)$ and so $A^\top$ also does not have rank m and thus is not invertible. So $\det(A^\top) = \det A$ simply because both $\det(A^\top)$ and $\det A$ are zero.

■

- **Corollary 12.22:** Theorem 12.4, Corollary 12.10, and Corollary 12.11 all hold for column operations as well.

  Proof: To see that this is true for any square matrix $A \in M_{m\times m}(F)$, note that column operations on A are given by row operations on $A^\top$ and then use the fact that taking the transpose of a matrix doesn't change the determinant (i.e. use Theorem 12.21).

■

- **Note 12.23:** The formula (12.2) computes the determinant by expanding along the first row. You can in fact expand along a different row or even column! Expanding along an i^{th} row looks as follows:

$$\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{i,1} & \cdots & a_{i,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix} = (-1)^{i-1}[a_{i,1} \det \overbrace{[A \text{ but remove } i^{\text{th}} \text{ row and first column}]}^{\text{an } (m-1)\times(m-1) \text{ matrix}}$$
$$-a_{i,2} \det[A \text{ but remove } i^{\text{th}} \text{ row and second column}] + \cdots$$
$$+(-1)^{m-1} a_{i,m} \det[A \text{ but remove } i^{\text{th}} \text{ row and } m^{\text{th}} \text{ column}]]$$
$$= (-1)^{i-1} \sum_{j=1}^{m} (-1)^{j-1} a_{i,j} \det A_{ij} = \sum_{j=1}^{m} (-1)^{i+j} a_{i,j} \det A_{ij}.$$

(here we used that $(-1)^{-2} = 1$). To prove this, first interchange the 1st and i^{th} row, expand along the first row using (12.2), and then in the submatrices interchange the rows $(i-2)$ times to get the matrices A_{ij} above. You will do $(i-1)$ interchanges in total, which is the origin of the $(-1)^{i-1}$ above.

Expanding along the j^{th} column looks as follows

$$\det\begin{pmatrix} \overbrace{a_{1,1}}^{j} & \cdots & a_{1,j} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,j} & \cdots & a_{m,m} \end{pmatrix} = (-1)^{j-1}[a_{1,j} \det \overbrace{[A \text{ but remove first row and } j^{\text{th}} \text{ column}]}^{\text{an } (m-1)\times(m-1) \text{ matrix}}$$

$$-a_{2,j} \det\big[A \text{ but remove second row and } j^{\text{th}} \text{ column}\big] + \cdots$$

$$+(-1)^{m-1} a_{m,j} \det\big[A \text{ but remove } m^{\text{th}} \text{ row and } j^{\text{th}} \text{ column}\big]]$$

$$= \sum_{i=1}^{m} (-1)^{i+j} a_{i,j} \det A_{ij}.$$

Using the fact that $\det A = \det A^{\top}$, this is proved by simply taking the transpose of the matrix A, computing the determinant of $A^{\top}$ by expanding along the j^{th} row, which will be equivalent to the above expansion along the j^{th} column.

- The determinant has a geometric interpretation as well:
- **Theorem 12.24:** If you have vectors $v_1, \ldots, v_m \in \mathbb{R}^m$, the volume (or area if $m = 2$) of the parallelepiped that they span is equal to the absolute value of the determinant of the matrix obtained by making $v_1, \ldots, v_m$ its columns or rows:

$$\text{Vol}\{t_1 v_1 + \cdots t_m v_m : \text{each } 0 \le t_i \le 1\}$$

$$= \left|\det\begin{pmatrix} (v_1)_1 & \cdots & (v_m)_1 \\ \vdots & \ddots & \vdots \\ (v_1)_m & \cdots & (v_m)_m \end{pmatrix}\right| = \left|\det\begin{pmatrix} (v_1)_1 & \cdots & (v_1)_m \\ \vdots & \ddots & \vdots \\ (v_m)_1 & \cdots & (v_m)_m \end{pmatrix}\right|.$$

Proof: You will most likely explore this in the homework. The proof is technically beyond the scope of the class because we need to define volume, which you do in a measure theory class. We remark that the absolute values above are needed since, unlike volume, determinants can be negative. The two determinants are equal because of the law $\det A = \det A^{\top}$.

■

- In particular, the above geometric interpretation explains immediately why a matrix is invertible if and only if its determinant is nonzero: can you figure out why?

- The following is a famous and useful rule of thumb that you should be aware of called **Cramer's rule**. It has a higher dimensional analog (see Theorem 4.9 in the book), though we won't cover it in this class.

- **Theorem 12.25:**

$$\text{If } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ is invertible, then } A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof: We know that if A is invertible, then $\det A \neq 0$ and so we're not dividing by zero above. The fact that the above formula gives A^{-1} can be checked directly by multiplying it by A to get I_2.

■

13 Eigenspaces and Diagonalization

13.1 Eigenvalues and Eigenvectors

- We now get to the most exciting part of linear algebra: the theory of eigenvalues and eigenvectors and their application to diagonalization – a powerful technique in linear algebra. This theory also goes by the name of "spectral theory" whose extensions to infinite dimensional vector spaces (something you'd study in a class on partial differential equations for instance) has had profound influence on analysis, differential geometry, and inverse problems.

  This theory is stunning: it starts with a simple idea which surprisingly develops into an entire collection of nontrivial and powerful facts that fit together perfectly and intertwine almost all of the ideas we have covered so far. The simple idea is the following: if we have a linear map $T: V \to V$, can we find directions that are preserved under T:

- **Definition 13.1:** Suppose that $T: V \to V$ is a linear map. A nonzero vector $v \in V$ is called an **eigenvector** of T if there exists a scalar $\lambda \in F$ such that $T(v) = \lambda v$ (note that λ can be zero). The scalar λ is called the **eigenvalue** corresponding to the eigenvector v.

- The definition for matrices is analogous.

- **Definition 13.2:** Suppose that $A \in M_{m\times m}(F)$ is a square matrix. A nonzero vector $v \in F^m$ is called an **eigenvector** of A if there exists a scalar $\lambda \in F$ such that $Av = \lambda v$. The scalar λ is called the **eigenvalue** corresponding to the eigenvector v. This is equivalent to v being an eigenvector of $L_A: F^m \to F^m$ with eigenvalue λ since $Av = \lambda v$ represents $L_A(v) = \lambda v$.

- **Note 13.3:** So how does one find eigenvectors and eigenvalues? A neat way to do this is to first find eigenvalues and then find eigenvectors as follows. Suppose you have a linear map $T: V \to V$ where V is finite dimensional whose representation is given by a square matrix $A \in M_{m\times m}(F)$. The following also work if you simply started with a square matrix A with no reference to a

linear map. A scalar $\lambda \in F$ is an eigenvalue if and only if (the v below is an eigenvector associated to λ)

$$\exists \text{ nonzero } v \in F^m : Av = \lambda v \quad \iff \quad \exists \text{ nonzero } v \in F^m : Av - \lambda v = 0$$

$$\iff \quad \exists \text{ nonzero } v \in F^m : (A - \lambda I_m)v = 0 \quad \iff \quad \exists \text{ nonzero } v \in F^m : v \in \text{null}(L_A - \lambda I_{F^m})$$

$$\overset{\text{dimension theorem}}{\iff} \quad \det(A - \lambda I_m) = 0.$$

The matrix $A - \lambda I_m$ looks as follows:

$$\begin{pmatrix} a_{1,1} - \lambda & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} - \lambda & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,m} - \lambda \end{pmatrix}. \tag{13.4}$$

If you take the determinant of this and set it zero you will get an equation of the form

$$b_m\lambda^m + b_{m-1}\lambda^{m-1} + \cdots + b_1\lambda + b_0 = 0.$$

The formal polynomial (c.f. Definition 12.3 for the second "=" here)

$$f(t) = \det(T - \lambda I_V) = \det(A - tI_m) = b_m t^m + \cdots + b_0$$

is called the **characteristic polynomial** of T and A and so we get the important *principle* that $\lambda \in F$ is an eigenvalue of A if and only if it is a root of the characteristic polynomial $f(t) = b_m t^m + \cdots + b_0$ when the latter is thought of as a function. This is how you solve for eigenvalues!

Before we move on, let's study the polynomial f itself. By computing the determinant of (13.4) using (12.2) iteratively, you will show in the homework that

$$b_m = (-1)^m \tag{13.5}$$

$$\text{and} \quad b_{m-1} = (-1)^{m-1}a_{1,1} + \cdots + (-1)^{m-1}a_{m,m} = (-1)^{m-1}\,\text{tr}\,A = \text{tr}\,T.$$

Moreover, we have that

$$b_0 = f(0) = \det(A - 0I_m) = \det A = \det T. \tag{13.6}$$

Computing the other coefficients b_i of f is harder. Before we continue, we interrupt this note with an important definition and result from algebra:

- **Note 13.7:** Suppose that $h = a_n x^n + \cdots + a_0 \in P(F)$ is a (formal) polynomial of **degree** n over F (i.e. $a_n \neq 0$). We say that h **splits** (over F) if it can be written as the formal product

$$h(x) = a_n(x - b_1) \ldots (x - b_n). \tag{13.8}$$

for some $b_1, \ldots, b_n \in F$ (the b_i's can repeat). Clearly each b_i is a **root** of h when h is thought of as a function $h : F \to F$ (i.e. each $h(b_i) = 0$). Conversely, it's a theorem from algebra that if $b_1, \ldots, b_n$ are roots of h, then (13.8) holds and furthermore it is the unique way to split h up to

rearranging the $(x - b_i)$ terms. This is not hard to prove: it's done by induction. Note that this implies that h cannot have more than n distinct roots.

Not all polynomials split. For instance, good luck splitting $x^2 + 1$ over $\mathbb{R}$. However there is one very special field over which polynomials always split:

- **Theorem 13.9: (Fundamental Theorem of Algebra)** Any polynomial over $F = \mathbb{C}$ splits.

  Remark: Because of this theorem, we say that $\mathbb{C}$ is an **algebraically closed field**.

  Proof: Proving this theorem is beyond the scope of this class. The simplest proofs use complex analysis or homotopy of continuous maps over the circle. ■

- **Note 13.3 continued:** Suppose that we can split our characteristic polynomial of T and A:

$$f(t) = b_m(t - \lambda_1) \dots (t - \lambda_m) \tag{13.10}$$

You will show in the homework that distributing (13.10) and using (13.5) and (13.6) gives that

$$b_{m-1} = b_m[-\lambda_1 - \cdots - \lambda_m] = (-1)^{m-1} \operatorname{tr} A \quad \Longrightarrow \quad \underbrace{\lambda_1 + \cdots + \lambda_m = \operatorname{tr} A = \operatorname{tr} T}_{\text{famous fact}},$$

$$b_m = b_m(-1)^m \lambda_1 \dots \lambda_m = \det A \quad \Longrightarrow \quad \underbrace{\lambda_1 \dots \lambda_m = \det A = \det T}_{\text{famous fact}}.$$

Now, suppose we fix an eigenvalue $\lambda \in F$. Solving for an eigenvector associated to it is easy. From our reasoning before we see that v is an eigenvector if and only if

$$(A - \lambda I_m)v = 0 \quad \Leftrightarrow \quad \begin{pmatrix} a_{1,1} - \lambda & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} - \lambda & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,m} - \lambda \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \Leftrightarrow$$

$$\begin{gathered} (a_{1,1} - \lambda)v_1 + a_{1,2}v_2 + \cdots + a_{1,m}v_m = 0, \\ a_{2,1}v_1 + (a_{1,2} - \lambda)v_2 + \cdots + a_{2,m}v_m = 0, \\ \vdots \\ a_{m,1}v_1 + a_{m,2}v_2 + \cdots + (a_{m,m} - \lambda)v_m = 0. \end{gathered}$$

Hence to find all eigenvectors associated to λ (there is no reason for there to only be one eigenvector!), just solve the above system of equations. There will be at least one (nonzero) eigenvector since $\det(A - \lambda I_m) = 0$ and so the above system has at least one free variable. In fact, the above shows that the set of all eigenvectors associated to the eigenvalue λ is a vector space given by:

$$E_\lambda = \operatorname{null}(T - \lambda I_V) \quad \text{which is represented by} \quad \operatorname{null}(A - \lambda I_m).$$

This is called the **eigenspace of** λ.

One more thing. Suppose that λ is an eigenvalue of A. The **algebraic multiplicity** of λ is defined as the largest value of k such that it's possible to write

$$f(t) = (t-\lambda)^k \cdot (\text{some polynomial}). \tag{13.11}$$

(here we say that $f(t)$ is "divisible" by $(t-\lambda)^k$). If you split $f(t)$ as

$$f(t) = (t-\lambda_1)^{k_1} \dots \left(t-\lambda_j\right)^{k_j}$$

where all of the λ_i are distinct, then each k_i is the algebraic multiplicity of λ_i (this is proved similarly to the theorem from algebra mentioned in Note 13.7). Note the important fact that in this case

$$k_1 + \cdots + k_j = m. \tag{13.12}$$

The **geometric multiplicity** of λ is defined as the dimension of its eigenspace E_λ:

$$\text{geometric multiplicity of } \lambda = \dim(E_\lambda).$$

So how do the two multiplicities relate?

- **Theorem 13.13:** Suppose that you have a linear map $T: V \to V$ where V is finite dimensional and suppose that $\lambda \in F$ is an eigenvalue. Then

$$1 \leq \underbrace{\text{geometric multiplicity of } \lambda}_{\dim E_\lambda} \leq \text{algebraic multiplicity of } \lambda.$$

Remark: When we study diagonalization later, we'll see that geometric multiplicity of λ being in fact equal to algebraic multiplicity of λ is a highly sought after condition.

Proof: Let $m = \dim V$. Let "$g = \dim E_\lambda$" and "a" denote the geometric and algebraic multiplicities of λ respectively. We need to show that $1 \leq g \leq a$. The fact that $1 \leq g$ follows from the fact that, as noted above, there is at least one (nonzero) eigenvector for every eigenvalue. Now, let $\{v_1, \dots, v_g\}$ be an ordered basis for E_λ and extend it to an ordered basis $\beta = \{v_1, \dots, v_g, w_1, \dots, w_{m-g}\}$ of V. It's not hard to see that with respect to this ordered basis, the representation of T is of the form

$$[T]_\beta^\beta = A = \begin{pmatrix} \lambda I_g & B \\ 0 & C \end{pmatrix}$$

(the "0" here represents a zero matrix). Hence the characteristic polynomial of T is given by (in the third equality below we use the result of Exercise 21 on page 229 in the book)

$$f(t) = \det(A - tI_m) = \det\begin{pmatrix} \lambda I_g - tI_g & B \\ 0 & C - tI_{m-g} \end{pmatrix} = \det\left(\lambda I_g - tI_g\right)\det\left(C - tI_{m-g}\right)$$

$$= (\lambda - t)^g \det\left(I_g\right)\det\left(C - tI_{m-g}\right) = (t-\lambda)^g \cdot (\text{some polynomial}).$$

Since the algebraic multiplicity "a" of λ is the largest power of $(t-\lambda)$ so that one can write $f(t)$ as in (13.11), we conclude that $g \leq a$.

■

- **Example 13.14:** Consider the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates the plane by 90 degree counterclockwise. With respect to the standard basis, its representation is given by:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Its characteristic polynomial is

$$f(t) = \det(A - tI_2) = \det\begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} = t^2 + 1.$$

Notice that we cannot split this polynomial (over $\mathbb{R}$). This in fact makes sense because geometrically we know that rotation by 90 degrees does not preserve any directions!

If, however, you view the above matrix as a complex matrix (i.e. $A \in M_{2\times 2}(\mathbb{C})$), then you will get the splitting $f(t) = (t - i)(t + i)$. Geometrically this means that if you view A as representing a rotation in $\mathbb{C}^2$, then there are fixed directions (they're hard to visualize!).

13.2 Diagonalization

- We now study one of the most important applications of spectral theory: diagonalization. When possible, this technique allows one to find an ordered basis with respect to which the representation of a linear map is a diagonal matrix. This not only helps understand the structure of the linear map, but also allows one to take powers of it very quickly. This, for instance, opens the gateway to composing functions with linear maps by running the latter through the former's Taylor series.

- **Definition 13.15:** Suppose that $T : V \to V$ is a linear map where V is finite dimensional. We say that T is **diagonalizable** if there exists an ordered basis β of V such that $[T]_\beta^\beta$ is a diagonal matrix:

$$[T]_\beta^\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix}. \tag{13.16}$$

It's not an accident that we use the letters λ_i here for the diagonal matrix, see Note 13.18 below!

- **Definition 13.17:** A square matrix $A \in M_{m\times m}(F)$ is called **diagonalizable** if $L_A : F^m \to F^m$ is diagonalizable.

- **Note 13.18:** Let's discuss Definition 13.15 a little bit further. Suppose that we diagonalized T as in (13.16). Notice then that $e_1, \dots, e_m$ are eigenvectors of $[T]_\beta^\beta$ with associated eigenvalues $\lambda_1, \dots, \lambda_m$. Thus $e_1, \dots, e_m$ are representations of an *ordered basis of eigenvectors* $\{v_1, \dots, v_m\}$. Conversely, if we had an ordered basis of eigenvectors $\beta = \{v_1, \dots, v_m\}$, then it's not hard to see that the representation $[T]_\beta^\beta$ of T is going to be of the form (13.16).

Let's see what diagonalization of a square matrix $A \in M_{m\times m}(F)$ looks like (c.f. Definition 13.17). If L_A is diagonalizable, then by the previous paragraph there exists an ordered basis of eigenvectors $\beta = \{v_1, \dots, v_m\}$ such that $[L_A]_\beta^\beta$ is equal to a diagonal matrix D. By changing bases from $\{e_1, \dots, e_m\} \to \{v_1, \dots, v_m\}$, we get that

$$A = QDQ^{-1} \tag{13.19}$$

for some matrix $Q \in M_{m\times m}(F)$. In other words, A is *similar* to a diagonal matrix. In fact, some thought should convince you that the columns of Q will be $v_1, \dots, v_m$ represented with respect to the standard ordered basis $\{e_1, \dots, e_m\}$. Similar to the previous paragraph, you can get the converse: if A is similar to a diagonal matrix (i.e. (13.19) holds), then A is diagonalizable. This proves the following two corollaries:

- **Corollary 13.20:** Suppose that $T : V \to V$ is a linear map where V is finite dimensional. Then T is diagonalizable if and only if there exists a basis of eigenvectors $\{v_1, \dots, v_m\}$ for V.
- **Corollary 13.21:** A square matrix $A \in M_{m\times m}(F)$ is diagonalizable if and only if it is similar to a diagonal matrix.
- We detract for a moment to talk about the relation between diagonalizability and the ability to split characteristic polynomials:
- **Theorem 13.22:** The characteristic polynomial of any diagonalizable linear map $T : V \to V$, where V is finite dimensional, splits. The same is true for square matrices $A \in M_{m\times m}(F)$.

  Proof: Suppose that $T : V \to V$ is diagonalizable. Then there exists an ordered basis β of V such that $[T]_\beta^\beta$ is diagonal as in (13.16). The characteristic polynomial of T is given by

$$f(t) = \det\left([T]_\beta^\beta - tI_m\right) = \det\begin{pmatrix} \lambda_1 - t & 0 & \cdots & 0 \\ 0 & \lambda_2 - t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m - t \end{pmatrix} \tag{13.23}$$

$$= (\lambda_1 - t) \dots (\lambda_m - t) = (-1)^m (t - \lambda_1) \dots (t - \lambda_m),$$

  and hence indeed splits.

  To prove the theorem for a diagonalizable matrix A, simply apply the above to $L_A : F^m \to F^m$.

  ■

- **Note 13.24:** The converse of Theorem 13.25 is not true: if the characteristic polynomial split, that does not necessarily imply that the linear map or matrix is diagonalizable. For instance, the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

only has the eigenvalue $\lambda = 1$ but as you can check $E_1 = \text{span}\{(0,1)\}$ is only one-dimensional. Hence by Corollary 13.20 (applied to $L_A : \mathbb{R}^2 \to \mathbb{R}^2$) A is not diagonalizable.

- Although Corollary 13.21 is conceptually important, it doesn't provide an algorithm to determine whether a matrix is diagonalizable and to diagonalize it if that's the case. Corollary 13.20 and (13.19) give us such an algorithm, but they require us to both find a list of eigenvectors and furthermore ensure that they're a basis if possible. It turns out that eigenvectors possess certain nice properties that allow us to circumvent some of that work, beginning with the following:

- **Theorem 13.25:** Suppose that $T : V \to V$ is a linear map. Suppose that $\lambda_1, \dots, \lambda_l$ are distinct eigenvalues of T. For each $i = 1, \dots, l$, let $S_i = \{v_{i,1}, \dots, v_{i,n_i}\} \subseteq E_{\lambda_i}$ be a linearly independent set of eigenvectors associated to λ_i. Then the list

$$S_1 \cup \dots \cup S_l = \{\underbrace{v_{1,1}, \dots, v_{1,n_1}}_{\text{Eigenvectors of } \lambda_1}, \dots, \underbrace{v_{l,1}, \dots, v_{l,n_l}}_{\text{Eigenvectors of } \lambda_l}\}$$

  is also linearly independent.

  Proof: First let's do the case $l = 2$ where the notation is easier. Let $S_1 = \{v_1, \dots, v_k\}$ and $S_2 = \{w_1, \dots, w_n\}$. We prove that $S_1 \cup S_2$ is linearly independent by contradiction: suppose not! Then there exist constants $a_1, \dots, a_k, b_1, \dots, b_n \in F$ not all zero so that

$$a_1 v_1 + \dots + a_k v_k + b_1 w_1 + \dots + b_n w_n = 0. \tag{13.26}$$

  Observe that $a_1, \dots, a_k$ cannot all be zero because if they were, then this would imply that not all of the $b_1, \dots, b_n$ are zero and hence the above would imply that $w_1, \dots, w_n$ are linearly dependent – a contradiction! Now, applying T to both sides of the above equation gives that

$$a_1 \lambda_1 v_1 + \dots + a_k \lambda_1 v_k + b_1 \lambda_2 w_1 + \dots + b_n \lambda_2 w_n = 0. \tag{13.27}$$

  Subtracting λ_2 times (13.26) from (13.27) gives:

$$a_1(\lambda_1 - \lambda_2)v_1 + \dots + a_k(\lambda_1 - \lambda_2)v_k + b_1(\lambda_2 - \lambda_2)w_1 + \dots + b_n(\lambda_2 - \lambda_2)w_n = 0$$
$$\implies \quad a_1(\lambda_1 - \lambda_2)v_1 + \dots + a_k(\lambda_1 - \lambda_2)v_k = 0.$$

  In other words, since $\lambda_1 \neq \lambda_2$ and not all of the $a_1, \dots, a_k$ are zero, we obtained a linear combination of the form

$$b_1 v_1 + \dots + b_k v_k = 0, \tag{13.28}$$

  where not all of the $b_1, \dots, b_k \in F$ are zero. But this implies that the $v_1, \dots, v_k$ are linearly dependent – a contradiction! To summarize: by utilizing the eigenvalue properties of T, we were able to "eliminate" the $w_1, \dots, w_n$ from the linear combination (13.26) to get the linear combination (13.28) and obtain a contradiction.

  In the general case, this is done similarly. By taking a linear combination

$$\underbrace{a_{1,1}v_{1,1} + \dots + a_{1,n_1}v_{1,n_1}}_{\text{Linear comb. of eigenvectors of } \lambda_1} + \dots + \underbrace{a_{l,1}v_{l,1} + \dots + a_{l,n_l}v_{l,n_l}}_{\text{Linear comb. of eigenvectors of } \lambda_l} = 0,$$

where not all of the $a_1, \dots, a_{l,n_l} \in F$ are zero, just as above one eliminates $v_{l,1}, \dots, v_{l,n_l}$ to get a linear combination of the form

$$\underbrace{b_{1,1}v_{1,1} + \cdots + b_{1,n_1}v_{1,n_1}}_{\text{Linear comb. of eigenvectors of } \lambda_1} + \cdots + \underbrace{b_{l-1,1}v_{l-1,1} + \cdots + b_{l-1,n_{l-1}}v_{l-1,n_{l-1}}}_{\text{Linear comb. of eigenvectors of } \lambda_{l-1}} = 0$$

where not all of the $b_1, \dots, b_{l-1,n_{l-1}} \in F$ are zero. Then one repeats such an "elimination" procedure until one arrives at a contradiction similar to the one obtained from (13.28) (i.e. in this case that the $v_{1,1}, \dots, v_{1,n_1}$ are linearly dependent).

■

- **Theorem 13.29:** Suppose that $T : V \to V$ is a linear map where V is finite dimensional. Suppose also that the characteristic polynomial of T splits:

$$f(t) = (-1)^m(t - \lambda_1)^{k_1} \dots \left(t - \lambda_j\right)^{k_j}$$

where $\lambda_1, \dots, \lambda_j$ are all of T's distinct eigenvalues (and hence $k_1, \dots, k_j$ are their multiplicities). Then

  a. T is diagonalizable if and only if the geometric multiplicity of each λ_i is equal to its algebraic multiplicity:

$$\dim E_{\lambda_i} = k_i.$$

  b. If T is diagonalizable and $\beta_1, \dots, \beta_j$ are bases for $E_{\lambda_1}, \dots, E_{\lambda_j}$ respectively, then $\beta = \beta_1 \cup \dots \cup \beta_j$ is a basis for V.

The same theorem holds with T replaced by a square matrix $A \in M_{m\times m}(F)$.

Proof: First we'll prove a). Suppose that T is diagonalizable, we want to show that $\dim E_{\lambda_i} = k_i$. Since T is diagonalizable, there exists an ordered basis β for V so that $[T]_\beta^\beta$ is diagonal:

$$[T]_\beta^\beta = \begin{pmatrix} \lambda_1 I_{m_1} & 0 & \cdots & 0 \\ 0 & \lambda_2 I_{m_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_2 I_{m_2} \end{pmatrix}$$

for some sizes $m_1, \dots, m_j$ to be determined. On the one hand, a computation like in (13.23) gives that

$$f(t) = (-1)^m(t - \lambda_1)^{m_1} \dots \left(t - \lambda_j\right)^{m_j},$$

and so each $m_i = k_i$ by the uniqueness of splitting. On the other hand, the equation for $[T]_\beta^\beta$ above shows that $e_1, \dots, e_m$ are representations of linearly independent eigenvectors of T and that for each λ_i there are m_i eigenvectors associated to it. Hence $k_i = m_i \leq \dim E_{\lambda_i}$. Since $\dim E_{\lambda_i} \leq k_i$ by Theorem 13.13, we get that $\dim E_{\lambda_i} = k_i$.

Now suppose that each $\dim E_{\lambda_i} = k_i$, we want to show that T is diagonalizable. Let β_i be a basis for each E_{λ_i}, which of course has $\dim E_{\lambda_i}$ number of vectors. Consider $\beta = \beta_1 \cup ... \cup \beta_j$. On the one hand, by Theorem 13.25 we have that $\beta = \beta_1 \cup ... \cup \beta_j$ is linearly independent and in particular has no repeats. On the other hand, by (13.12) we have that the number of vectors in β is given by

$$\dim E_{\lambda_1} + \cdots + \dim E_{\lambda_j} = k_1 + \cdots + k_m = m.$$

Hence β is a basis for V consisting of eigenvectors. Thus by Corollary 13.20 T is diagonalizable. This also proves b).

As before, we prove the theorem for a square matrix A simply by applying the above to $L_A : F^m \to F^m$.

■

- **Corollary 13.30:** Suppose that $T : V \to V$ is a linear map where V is m-dimensional. If T has m distinct eigenvalues $\lambda_1, \dots, \lambda_m$, then T is diagonalizable. The same holds for square matrices $A \in M_{m\times m}(F)$.

  Proof: We have that the characteristic polynomial splits as:

  $$f(t) = (-1)^m(t - \lambda_1) ... (t - \lambda_m).$$

  Hence every λ_i has an algebraic multiplicity of $k_i = 1$. Thus by Theorem 13.13

  $$1 \leq \dim E_{\lambda_i} \leq k_i = 1$$

  and so each $\dim E_{\lambda_i} = k_i$. Thus by Theorem 13.29 a) our linear map T or matrix A is diagonalizable.

■

- Amazingly there is one class of matrices that is always guaranteed to be diagonalizable which appears in many places of mathematics:

- **Theorem 13.31: (Spectral Theorem)** Suppose that $A \in M_{m\times m}(\mathbb{R})$ is a **symmetric** real matrix, which means that $A^\top = A$ (i.e. it's symmetric across the diagonal). Then all eigenvalues of A are real, the characteristic polynomial of A splits, and A is diagonalizable. Moreover, there exists an **orthonormal** basis $u_1, \dots, u_m$ of eigenvector with respect to the dot product of A, which means that the length of each u_i is one and $u_i \perp u_j$ if $i \neq j$. In this case

  $$A = UDU^{-1}, \tag{13.32}$$

  where the columns of U are $u_1, \dots, u_j$. Furthermore, because the columns of U are orthonormal,

  $$U^{-1} = U^\top \tag{13.33}$$

  and so (13.32) can be rewritten as

$$A = UDU^{\top}.$$

Proof: This is proved by induction, and uses the following equation relating dot products and transposes:

$$(Bx) \cdot y = x \cdot (B^{\top} y) \quad \text{and} \quad x \cdot (Cy) = (C^{\top} x) \cdot y,$$

which is an elementary exercise. We note that proving (13.33) is also on the level of an elementary exercise. We don't have time to prove this theorem in this course.

■

Have a great summer!