Logic and Set Theory

A fundamental object of study in this course is a *vector space*. On one hand, this is a simple object to define. A vector space is simply a module over a field. But that definition is meaningless unless you know what a module and a field are. So, we will begin this course from the beginning, and build linear algebra "from the ground up". That means starting with a rigorous introduction to set theory and logic.

1. Set Theory

Definition: Set,elements

A *set* X is a well-defined collection of objects, called *elements*. One should be able to determine membership in a set. We write $a \in X$ to say an element is in the set.

Example. Important sets to know are the following:

(1) $\emptyset$, the empty set
(2) $\mathbb{N}$, the natural numbers $1, 2, 3, \ldots$
(3) $\mathbb{Z}$, the integers (positive and negative whole numbers, and zero)
(4) $\mathbb{Q} = \left\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\right\}$, the rational numbers
(5) $\mathbb{R}$, the real numbers
(6) $\mathbb{R}_+$, the positive real numbers
(7) $\mathbb{C}$, the complex numbers

Definition: Subset, equal sets, proper subset

A *subset* of a set X is a set Y such that for all $y \in Y$, $y \in X$. We write $Y \subseteq X$. We say sets X and Y are *equal* and write $X = Y$ if $X \subseteq Y$ and $Y \subseteq X$. We say Y is a *proper subset* of X if $Y \subseteq X$ and $Y \neq X$. In this case we typically write $Y \subsetneq X$.

It is commonly understood that $\subset$ means the same thing as $\subseteq$. (It is actually my personal preference to use $\subset$ but I will try to stay consistent with your text.)

Example. $\emptyset \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

The previous example is actually a bit of an abuse of notation. The set $\mathbb{Z}$ is not formally a subset of $\mathbb{Q}$. But there is a subset of $\mathbb{Q}$ consisting of fractions $m/1$ which is *in bijection* with $\mathbb{Z}$. An important theme in this course is that mathematical objects should be defined independently of a particular representation.

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapters 1. Last Updated: September 14, 2021

One should also note that the empty set $\emptyset$ is a subset of *every* set, even itself. If you're in the mood to stretch your brain, consider whether there is a set that contains all sets.

Example. Suppose X is a set with n elements (we write $|X| = n$ to denote this when X has finitely many elements). How many subsets does X have? Can you justify your conjecture?

Say A and B be sets. A *universal set* for A and B is any set containing both of them. In most cases, there will be a natural choice for U. For example, if A and B are both sets of integers, then we may take $U = \mathbb{Z}$.

Operations on sets

Let A and B be subsets of a universal set U.

- Union on A and B: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- Intersection of A and B: $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- Complement of A in U: $A' = \{x : x \in U \text{ and } x \notin A\}$
- Difference: $A \backslash B = \{x : x \in A \text{ and } x \notin B\}$
- Cartesian Product: $A \times B = \{(a, b) : a \in A, b \in B\}$

Set difference is sometimes denoted with a minus sign $(-)$ instead of a backslash $(\backslash)$. The next proposition explains how some of the set operations play together.

Proposition 1: Properties of $\cup$ and $\cap$

Let A, B, C be sets.

(1) $A \cup A = A$, $A \cap A = A$, $A \backslash A = \emptyset$.
(2) $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$.
(3) $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$.
(4) $A \cup B = B \cup A$, $A \cap B = B \cap A$.
(5) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
(6) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

To prove an equality between two sets, it suffices to show that all the elements in one set are in the other, and vice-versa. That is, we want to show that each subset is a subset of the other, or otherwise denoted as a "double inclusion".

(5). Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. (Note that the mathematical "or" is not "exclusive or"). If $x \in A$, then $x \in A \cup B$ and $x \in A \cup C$, so $x \in (A \cup B) \cap (A \cup C)$. If $x \in B \cap C$, then $x \in B$ so $x \in A \cup B$, and $x \in C$ so $x \in A \cup C$. Again we have $x \in (A \cup B) \cap (A \cup C)$. Thus, in either case we have $A \cup (B \cap C) \in (A \cup B) \cap (A \cup C)$.

For the converse, suppose $y \in (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$. Suppose $y \in A$, then $y \in A \cup (B \cap C)$. Now suppose $y \notin A$. Since $y \in A \cup B$, then $y \in B$. Similarly, since $y \in A \cup C$,

then $y \in C$. Thus, $y \in B \cap C$ so again we have $y \in A \cup (B \cap C)$. Thus, in either case we have $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$. □

Theorem 2: DeMorgan's Laws (for sets)

Let A and B be subsets of a (universal) set U.

(1) $(A \cup B)' = A' \cap B'$.

(2) $(A \cap B)' = A' \cup B'$.

Proof. Let $x \in (A \cup B)'$. Then $x \in U$ and $x \notin A \cup B$. Suppose $x \in A$, then $x \in A \cup B$, a contradiction. Hence, $x \notin A$. Similarly, $x \notin B$. It follows that $x \in A'$ and $x \in B'$. Thus, $x \in A' \cap B'$. That is, $(A \cup B)' \subseteq A' \cap B'$.

Conversely, suppose $x \in A' \cap B'$. Then $x \in A'$ and $x \in B'$. Since $x \in A'$, then $x \in U$ and $x \notin B$. Since $x \in B'$, then $x \in U$ and $x \notin B$. In either case, $x \in U$. If $x \in A \cup B$ then $x \in A$ or $x \in B$, a contradiction. Thus, $x \notin A \cup B$, so $x \in (A \cup B)'$. That is, $A' \cap B' \subseteq (A \cup B)'$.

Now by double inclusion, $(A \cup B)' = A' \cap B'$. □

Definition: Statement

A *statement* is a sentence or mathematical expression that is either true or false.

An example of a (mathematical) statement is:

Example. The following are (mathematical) statements:

- A circle with radius r has area πr^2. (True)
- The number $\sqrt{2}$ is rational. (False)
- Every square is a rectangle. (True)
- Every rectangle is a square. (False)
- Some rectangles are squares. (True)

We often refer to statements with letters (typically, P, Q, R). There are many ways to join statements. We will discuss several here. These should be compared to the set properties above.

Let P and Q be statements. The statement "P *and* Q" (denoted $P \wedge Q$) is true if both P and Q are true, otherwise it is false. We can express the different scenarios for $P \wedge Q$ using a *truth table*.

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

The statement "P *or* Q" (denoted $P \vee Q$) is true if at least one of P or Q are true. It is false only when both P and Q are false.

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

Always, *or* in mathematics is defined as above. Do not confuse this with *exclusive or*, which is true when *exactly* one of P or Q is true.

The *negation* of the statement P is the "*not* P" (denoted $\sim P$). The truth value of $\sim P$ is the opposite of P:

P	$\sim P$
T	F
F	T

Statements involving and, or, not can be combined to form more complex statements. It is also worth considering how these statements play with each other. We say two statements P and Q are *equivalent* (and write $P \equiv Q$) if their truth values align.

Theorem 3: DeMorgan's Laws (for logic)

Let P and Q be statements

(1) $\sim (P \wedge Q) \equiv (\sim P) \vee (\sim Q)$

(2) $\sim (P \vee Q) \equiv (\sim P) \wedge (\sim Q)$

Proof. We will prove (1) and leave (2) as an exercise. To prove the equivalence, it suffices to show that the sides have have the same truth table:

P	Q	$P \wedge Q$	$\sim (P \wedge Q)$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T

P	Q	$(\sim P)$	$(\sim Q)$	$(\sim P) \vee (\sim Q)$
T	T	F	F	F
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T

The result follows by comparing the last columns of the two truth tables. □

Another way to combine statements is to use conditionals. A *conditional statement* is of the form "If P, then Q" (denoted $P \Rightarrow Q$). The statement $P \Rightarrow Q$ means that *if P is true, then Q must be true.*

P	Q	$P \Rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

This is a little awkward to many students at first. Note in the last two rows, P is *false*, and so the conditional statement is *vacuously true.*

There are many other ways to phrase a conditional statement. One is P *implies* Q (see Page 44 of Hammack's text for many more examples). Importantly, in this instance we say that P is a *sufficient condition* for Q, and we say that Q is a *necessary condition* for P.

Exercise. The negation of a conditional statement *is not* another conditional statement. Show that $\sim (P \Rightarrow Q)$ is logically equivalent to $P \wedge \sim Q$.

From a conditional statement $P \Rightarrow Q$ one can form several other conditional statements:

Converse: $Q \Rightarrow P$

Inverse: $\sim P \Rightarrow \sim Q$

Contrapositive: $\sim Q \Rightarrow \sim P$

Exercise. Show that a conditional statement $P \Rightarrow Q$ and its contrapositive $\sim Q \Rightarrow \sim P$ are logically equivalent. Show that the converse and inverse are logically equivalent to each other.

The statement $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ is called a *biconditional* and is true only when a statement and its converse have the same truth value. We denote a biconditional simply by $P \Leftrightarrow Q$. This is read "P if and only if Q".

P	Q	$P \Rightarrow Q$	$Q \Rightarrow P$	$P \Leftrightarrow Q$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

The symbol $\forall$ means "for all/every/each" (universal quantifier) and the symbol $\exists$ means "there exists" (existential quantifier). Note, in most situations it is preferable to use the words in place of the symbols.

Example 4. Consider the sentence

$$\text{There is (exists) an integer } n \in \mathbb{Z} \text{ for which } n^2 = 2.$$

This could be written, symbolically, as

$$\exists n \in \mathbb{Z}, n^2 = 2.$$

But the statement in English is preferable. Of course, in either case, the statement is false.

Example 5. Consider the sentence

$$\text{Every integer that is not odd is even.}$$

This could be written, symbolically, as

$$\forall n \in \mathbb{Z}, \sim(n \text{ is odd}) \Rightarrow (n \text{ is even}).$$

(Aside: Here is a general remark/rule-of-thumb. Never (ever!) start a sentence in one of your proofs with a symbol. One reason is that the period "." has mathematical meaning (e.g., it can be used for inner products), so if your sentence ends with a symbol and the next one begins with a symbol, this can cause confusion. This is unlikely in most situations, but ultimately it is just good style.

To make the above more precise, an *open statement* is one of the form $P(x)$, whose truth value cannot be determined until we substitute a value for x (this need not be a number, it can be almost anything). In terms of an open statement $P(x)$ and the set X, the statements above are,

$$\text{universal quantifier: } \forall x \in X, P(x)$$

$$\text{existential quantifier: } \exists x \in X P(x).$$

The negation of a "for all" statement is a "there exists" statement and vice-versa. In particular,

$$\sim (\forall x \in X, P(x)) \equiv \exists x \in X, \sim P(x)$$

$$\sim (\exists x \in X, P(x)) \equiv \forall x \in X, \sim P(x)$$

3. Functions

Here we formally define functions between sets.

Definition: Relation, function, domain, codomain, image

Let S and T be sets. A *relation* is a subset of the cartesian product $S \times T$. A *function* $f \subset S \times T$ is a relation such that if $(s,t),(s,t') \in f$ then $t = t'$. The set $\{s \in S : (s,t) \in f\}$ is the *domain of* f and the set $\{t \in T : (s,t) \in f\}$ is the *image of* f, denoted $\mathrm{Im}(f)$.

A function may also be called a *map* or *mapping*. Our standard notation for a function $f \subseteq S \times T$, however, will be $f : S \to T$. In this context, we will call the S the domain and T the *codomain* of f. To rephrase the above definition, we say f is *well-defined* if for every value $s \in S$ there is one and only one $t \in T$ such that $f(s) = t$. In this notation,

$$\mathrm{Im}(f) = \{t \in T : f(s) = t \text{ for some } s \in S\}.$$

Let $f : S \to T$ be a function and $t \in T$. An element $s \in S$ such that $f(s) = t$ is called a *preimage of* t *(under* f*)*. A given codomain element need not have a unique preimage, or any at all.

Definition: Injective, surjective, bijective

Let $f : S \to T$ be a function. If whenever $f(s) = f(s')$ we have $s = s'$, then f is said to be *injective*. If $\mathrm{Im}(f) = T$, then f is said to be *surjective*. A function that is both injective and surjective is said to be *bijective*.

The term injective is sometimes called *one-to-one* and surjective is sometimes called *onto*.

Example. (1) Let S be a nonempty set. The *identity map* (on S) is the function

$$\begin{aligned} 1_S : S &\to S \\ s &\mapsto s \end{aligned}$$

This function is bijective.

(2) Consider the function

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

Since $f(1) = f(-1)$, then f is not injective. Since the image does not contain any negative real numbers, then it is not surjective. It we restrict the domain to nonegative real numbers ($\mathbb{R}_+$), then the map is injective. If we restrict the codomain to $\mathbb{R}_+$, then the function is surjective. If we restrict both to $\mathbb{R}_+$, then the function is bijective.

(3) We denote by $\mathscr{C}^0(\mathbb{R})$ the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$, and by $\mathscr{C}^1(\mathbb{R})$ the set of differentiable functions $f : \mathbb{R} \to \mathbb{R}$ with continuous derivatives. Consider the function

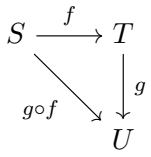
$$D : \mathscr{C}^1(\mathbb{R}) \to \mathscr{C}^0(\mathbb{R})$$
$$f \mapsto \frac{df}{dx}$$

This map is surjective (by the Fundamental Theorem of Calculus) but not injective (e.g., x and $x + 1$ have the same derivative).

Definition: Composition

Let $f : S \to T$ and $g : T \to U$ be functions. The *composition* $g \circ f : S \to U$ is defined by the rule $(g \circ f)(s) = g(f(s))$ for all $s \in S$.

We may represent composition by the *commutative diagram*:



By a commutative diagram, we mean that if you follow an element around the diagram in any way you will arrive at the same result.

Let 1_S and 1_T denote the identity functions of the sets S and T, respectively, and let $f : S \to T$ be a function. Then we have

$$f \circ 1_S = f \quad \text{and} \quad 1_T \circ f = f.$$

Proposition 6: Composition of functions is associative

Let $f : S \to T$, $g : T \to U$, and $h : U \to V$ be functions. Then

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Proof. The domain of both sides is S, so given an element in S, it suffices to show that each side evaluates to the same element in V.

Let $s \in S$. Set $t = f(s)$, $u = g(t)$. That is, $(g \circ f)(s) = g(f(s)) = u$. Set $v = h(u)$. Then

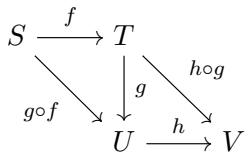
$$(h \circ (g \circ f))(s) = h((g \circ f)(s)) = h(u) = v.$$

Similarly, $(h \circ g)(t) = h(g(t)) = h(u) = v$ so

$$((h \circ g) \circ f)(s) = (h \circ g)(f(s)) = (h \circ g)(t) = v.$$

This completes the proof. □

If we were clever, we might try to prove the above proposition by diagram:



In general, function composition is *noncommutative*. That is, $f \circ g \neq g \circ f$. For one, the reverse composition may not even be defined. But even if they are, commutativity can fail. For example, consider $f(x) = x^2$ and $g(x) = x+1$ (both functions defined $\mathbb{R} \to \mathbb{R}$). Of course, there are examples where two functions do commute (e.g., one is the identity function).

Proposition 7: Properties of composition

Let $f : S \to T$ and $g : T \to U$ be functions.

(1) If f and g are injective, then $g \circ f$ is injective.
(2) If f and g are surjective, then $g \circ f$ is surjective.
(3) If f and g are bijective, then $g \circ f$ is bijective.
(4) If $g \circ f$ is injective, then f is injective.
(5) If $g \circ f$ is surjective, then g is surjective.

Proof. (1) Let $s, s' \in S$ such that $(g \circ f)(s) = (g \circ f)(s')$. We claim $s = s'$.

From our hypothesis, we have $g(f(s)) = g(f(s'))$. Since g is injective, then $f(s) = f(s')$. Now by the injectivity of f, $s = s'$.

(2) (Diagram chase) Let $u \in U$. We claim that u has a preimage in S. Since g is surjective, then there exists $t \in T$ such that $g(t) = s$. Because f is surjective, there exists $s \in S$ such that $f(s) = t$. But then $(g \circ f)(s) = g(f(s)) = g(t) = u$. That is, s is a preimage of u under $g \circ f$.

(3) This follows from (1) and (2).

(4) Let $s, s' \in S$ such that $f(s) = t = f(s')$. Then $(g \circ f)(s) = g(f(s)) = g(t) = g(f(s')) = (g \circ f)(s')$. By injectivity of $g \circ f$, then $s = s'$ as desired.

(5) Let $u \in U$. Since $g \circ f$ is surjective, there exists $s \in S$ such that $(g \circ f)(s) = u$. Then $g(f(s)) = u$, so $f(s)$ is a preimage of u under g. □

Definition: Invertible, inverse function

A function $f : S \to T$ is said to be *invertible* if there exists a function $g : T \to S$ such that

$$g \circ f = \mathrm{id}_S \quad \text{and} \quad f \circ g = \mathrm{id}_T .$$

In this case, the map g is called the *inverse function of* f.

The following result is crucial to our study of functions on vector spaces. To prove uniqueness, our standard strategy will be to assume there are two objects with a given property and show they are the same.

Proposition 8: Inverses are unique

If $f : S \to T$ is an invertible function, then the inverse function $g : T \to S$ of f is unique.

Proof. Let $h : T \to S$ be an inverse function of f. Then

$$h = h \circ 1_T = h \circ (f \circ g) = (h \circ f) \circ g = 1_S \circ g = g. \qquad \square$$

Since the inverse is unique, when f is invertible we will denote its inverse by f^{-1}. By the symmetry in the definition of inverse function, f^{-1} is also invertible with inverse f. By uniqueness of the inverse, this now implies that $(f^{-1})^{-1} = f$.

Proposition 9: Invertibility equivalent to bijectivity

A function $f : S \to T$ is invertible if and only if is bijective.

Proof. Assume f is invertible. By definition, there exists an inverse function f^{-1}. Let $s, s' \in A$ such that $f(s) = f(s')$. Applying f^{-1} to both sides gives $s = s'$, so f is injective. Let $t \in T$ and set $s = f^{-1}(t)$. Applying f to both sides gives $f(s) = t$, so f is surjective and hence bijective.

Now assume f is bijective. Then for every $s \in S$, there is exactly one element $t \in T$ such that $f(a) = b$. (Existence follows from surjectivity and uniqueness follows from injectivity.) Define $g : T \to S$ by the rule $g(t) = s$ if $f(s) = t$. Then g is well-defined (by the above). Moreover, $g(f(s)) = s$ and $f(g(t)) = t$, so $g = f^{-1}$. $\square$

Example. The symbol $\mathbb{R}_+^\times$ denotes the set of positive real numbers. The functions

$$\begin{aligned} \mathbb{R} &\to \mathbb{R}_+^\times & \qquad \mathbb{R}_+^\times &\to \mathbb{R} \\ x &\mapsto e^x & \qquad x &\mapsto \ln(x). \end{aligned}$$

are both bijective and inverses of one another.

Let $f : S \to T$ be a function between sets S and T and let $V \subseteq T$. Then the preimage of V is defined as

$$f^{-1}(V) = \{s \in S : f(s) \in T\}.$$

This notation *does not* imply that f is invertible. If $V = \{t\}$ is a singleton, then by an abuse of notation, we will write $f^{-1}(t)$. Now if f is a bijection, then $f^{-1}(t)$ will be a singleton, say $\{s\} = f^{-1}(t)$. By another abuse of notation, we will more commonly write $s = f^{-1}(t)$.

4. Cardinality

We now discuss how one compare the "sizes" of infinite set.

Definition: Cardinality

Nonempty sets S and T are said to have the same *cardinality* if there exists a bijection $f : S \to T$.

When a set S is finite, say with n elements, then there is a bijection $f : \{1, \ldots, n\} \to S$ and we write $\text{Card}(S) = n$ (or $|S| = n$). So cardinality is really only interesting with regards to infinite sets. But, not all infinities are created equal.

If there exists an injective function $f : S \to T$, then we write $\text{Card}(S) \leq \text{Card}(T)$. If there exists an injection but no bijection, then we write $\text{Card}(S) < \text{Card}(T)$. Similarly, if there exists a surjective function $f : S \to T$, then $\text{Card}(T) \leq \text{Card}(S)$.

Theorem 10: Comparing cardinalities

Let S and T be nonempty sets. If $\text{Card}(S) \leq \text{Card}(T)$ and $\text{Card}(T) \leq \text{Card}(S)$, then $\text{Card}(S) = \text{Card}(T)$.

One version of the *pigeonhole principle* states that if n pigeons fly into m pigeonholes and $n > m$, then at least one hole must contain two or more pigeons. Here is the statement for functions.

Theorem 11: Pigeonhole principle

Let S and T be sets of the same finite cardinality. A function $f : S \to T$ is injective if and only if it is surjective.

Proof. Suppose f is injective. Then $|f(S)| = |S| \leq |T|$. If f is not surjective, then $|f(S)| < |T|$, so $|S| < |T|$, a contradiction. Thus, f is surjective. The converse is similar and left to the reader. □

Example. The sets $\mathbb{N}$ (natural numbers starting at 0) and $\mathbb{Z}$ (integers) have the same cardinality. To prove this, we need to set up a bijection. Define $f : \mathbb{N} \to \mathbb{Z}$ in the following way:

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

We will verify now that f is bijective.

Let $m, n \in \mathbb{N}$ such that $f(m) = f(n)$. By definition of f, this implies m, n are either both even or both odd. Suppose they are both even. Then we have $-m/2 = -n/2$, so $m = n$. The odd case is similar. Hence, f is injective.

Let $x \in \mathbb{Z}$ and suppose $x > 0$. Set $n = 2x - 1$. Then

$$f(n) = \frac{(2x-1)+1}{2} = x.$$

The case $x \leq 0$ is similar. Thus, f is surjective.

We commonly write the cardinality of $\mathbb{N}$ (and also $\mathbb{Z}$ by the previous example), by $\aleph_0$. We say that a set with this cardinality is *countably infinite*.

Example. We will show that $\text{Card}(\mathbb{Z}) = \text{Card}(\mathbb{Q})$.

First, there is an injection $f : \mathbb{Z} \to \mathbb{Q}$ given by $f(n) = n/1$. Hence, $\text{Card}(\mathbb{Z}) \leq \text{Card}(\mathbb{Q})$. One way to show the reverse inequality is in Hammack's book (page 269). One simply needs to define an "ordering" on $\mathbb{Q}$, and there are many ways to do this.

Example. On the other hand, $\mathbb{R}$ is not countably infinite. To do this, we use the decimal expansion of a real number.

Let $f : \mathbb{N} \to \mathbb{R}$ be a function. Choose an element $b = b_0.b_1b_2b_3b_4 \ldots \in \mathbb{R}$ in the following way:

- Choose b_0 so that it is different from the whole part of $f(0)$.
- Choose b_1 so that is is different from the first decimal place of $f(1)$.
- Choose b_2 so that is is different from the second decimal place of $f(2)$.
- And so on...
- Choose b_k so that is is different from the kth decimal place of $f(k)$.

Then for every n, b differs from $f(n)$ in at least one decimal place, so $f(n) \neq b$ for every n. That is, f is not surjective. Since f was chosen arbitrarily, this implies that there is no surjective function (and hence no bijective function) $\mathbb{N} \to \mathbb{R}$. However, there is an obvious injective map $\mathbb{N} \to \mathbb{R}$ (just send each element of $\mathbb{N}$ to itself in $\mathbb{R}$), so $\text{Card}(\mathbb{N}) < \text{Card}(\mathbb{R})$.

5. Permutations

Definition: Permutation, symmetric group

Set $\mathcal{P}_n = \{1, 2, \ldots, n\}$. A *permutation* is a bijective function $\mathcal{P}_n \to \mathcal{P}_n$. The set of all permutations on $\mathcal{P}_n$ is called the *symmetric group on n letters*, denoted $\mathcal{S}_n$.

Example. The function $f : \mathcal{P}_3 \to \mathcal{P}_3$ given by $f(1) = 2$, $f(2) = 1$, and $f(3) = 3$ is a permutation.

The following facts are a consequence of our study of functions.

(1) The composition of two permutations is a permutation. That is, if $f, g \in \mathcal{S}_n$, then $f \circ g \in \mathcal{S}_n$.
(2) Function composition is associative.
(3) There is an element in $\mathcal{S}_n$ that acts as the identity with respect to composition. For all $f \in \mathcal{S}_n$, $1_{P_n} \circ f = f = f \circ 1_{P_n}$.
(4) Every permutation is invertible and the inverse of a permutation is a permutation. That is, if $f \in \mathcal{S}_n$, then there is some $g \in \mathcal{S}_n$ such that $f \circ g = 1_{P_n} = g \circ f$.

The above shows that $\mathcal{S}_n$ is an example of a *group*. We will study groups in more detail soon.

We can list the elements of $\mathcal{S}_n$ is *two-line* notation. Let $f \in \mathcal{S}_n$, then the two-line notation for f is

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$$

Example. The elements of $\mathcal{S}_3$ in two-line notation are

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

We can compose permutations by "stacking". Note that we compose right-to-left (because these are functions).

Example 12. Let $f = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ and $g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Then

$$f \circ g = f \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$

$$g \circ f = g \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Note that $f \circ g \neq g \circ f$.

Proposition 13: Cardinality of $\mathcal{S}_n$

The cardinality of $\mathcal{S}_n$ is $n!$.

Proof. There are n choices for the image of 1, $n-1$ choices for the image of 2, $n-2$ choices for the image of 3, and so on. Ultimately, there is only 1 choice for the image of n. Hence, there are a total of $n \cdot (n-1) \cdot (n-2) \cdots 1 = n!$ choices for the images of elements in $\mathcal{P}_n$. Thus, $|\mathcal{S}_n| = n!$. □

Now we discuss a more compact way of writing elements of $\mathcal{S}_n$.

Definition: k-cycle, transposition

Let $a_1, \ldots, a_k \in \mathcal{P}_n$. The *$k$-cycle* $(a_1\ a_n\ \ldots\ a_n)$ is the permutation defined by the rule

$$a_1 \mapsto a_2, \quad a_2 \mapsto a_3, \quad \ldots \quad a_n \mapsto a_1,$$

and fixes all other elements of $\mathcal{P}_n$. A *transposition* is a 2-cycle.

Example. (1) The permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 2 \end{pmatrix}$$

can be expressed in cycle form as $\pi = (1\ 2\ 4)$.

(2) The permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

can be expressed as the product of two disjoint cycles as $\pi = (1\ 4) \circ (2\ 3)$.

(3) Consider the permutation $\pi = (1\ 2\ 4\ 3) \circ (1\ 4)$. These cycles are not disjoint, and so it can be simplified. We could write this out in two-line format to compute, but it's easier to just track the image of each element through successive cycles, remembering to close cycles when we get back to where we started. So $\pi = (1\ 3)(2\ 4)$.

It turns out that every permutation can be written as the product of (not necessarily disjoint) transpositions. One way to do this, is the following:

$$(a_1\ a_n\ \ldots\ a_n) = (a_1\ a_n)(a_1\ a_{n-1})(a_1\ a_{n-2}) \cdots (a_1\ a_2).$$

If n is odd, then this decomposition involves an even number of transpositions, and vice-versa. While this particular decomposition is not unique, it turns out the parity is determined by the permutation.

Theorem 14: Invariance of Parity

Suppose that a permutation in $\mathcal{S}_n$ may be expressed as the product of an even (respectively, odd) number of transpositions. Then every factorization into transposition likewise involves an even (respectively, odd) number of factors.

We say that a permutation is *even* or *odd* if it can be expressed by an even or odd number of transpositions, respectively. This gives rise to a well-defined map $\sigma : \mathcal{S}_n \to \{\pm 1\}$ defined by

$$\sigma(\pi) = \begin{cases} 1 & \text{if } \pi \text{ is even} \\ -1 & \text{if } \pi \text{ is odd} \end{cases}$$

This map is called the *sign homomorphism*. It has some extra structure because it respects composition.

Groups

We continue in our goal of defining a vector space. We previously defined a set, now we begin to put some structure on that set.

1. Groups and subgroups

We begin with a concept you are very familiar with, but likely have not seen it given a formal name.

Definition: Binary operation

A *binary operation* on a set S is a function $S \times S \to S$.

We will often use $*$ or $\cdot$ for an arbitrary operation. So, formally, a binary operation is an assignment

$$S \times S \to S$$
$$(s, t) \mapsto s * t$$

Another way of stating this is to say that S *is closed under the operation.*

Example. The following are examples of binary operations:

(1) The addition operation $+$ is a binary operation on $\mathbb{N}$. It is also a binary operation on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$.
(2) Multiplication is a binary operation on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$.
(3) Composition on permutations in $\mathcal{S}_n$ is a binary operation.

The subtraction operation is not a binary operation on $\mathbb{N}$ (since $3 - 5 = -2 \notin \mathbb{N}$). It is a binary operation on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$.

Let $*$ be a binary operation on S. The operation $*$ is *associative* if $(s * t) * u = s * (t * u)$ for all $s, t, u \in S$. An element $e \in S$ is an *identity element* for $*$ if $s * e = s = e * s$ for all $s \in S$. The operation $*$ is commutative if $s * t = t * s$ for all $s, t \in S$.

Definition: Group

A *group* $(G, *)$ is a set G along with a binary operation $*$ such that

(1) the operation $*$ is associative
(2) there exists an element $e \in G$ which is an identity for $*$,
(3) and for every $s \in G$ there exists a $t \in G$ such that $s * t = e = t * s$.

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapters 2. Last Updated: October 21, 2021

Another way we could phrase part (3) is to say that every element must have an inverse in G.

We say a group is *commutative* (or *abelian*) if the operation $*$ is commutative. Any group with the addition operation is assumed to be commutative, typically with identity element 0. When working in an arbitrary group, where we do not assume commutativity, we will often use multiplicative notation (and often drop the operation symbol).

Example. The following are examples of binary operations:

(1) $(\mathbb{Z}, +)$ is a group. The identity element is 0 and the additive inverse of $n \in \mathbb{Z}$ is $-n$.
(2) $(\mathbb{Q}, \times)$ is not a group because 0 is not invertible.
(3) $(\mathbb{Q}^\times, \times)$ is a group.
(4) $(\mathcal{S}_n, \circ)$ is a group (we have already verified this).

Note that $(\mathbb{N}, +)$ is not a group because it fails part (3) of the definition

Here are some more exotic examples.

Example. Consider the set $S = \{0, 1, 2, 3, 4\}$. The operation (called *addition mod* 5) is defined by first taking the sum of two numbers and then taking the remainder when the sum is divided by 5. So, $1 + 2 = 3$ but $3 + 4 = 2$. Then S with this operation is a commutative group, which can be easily observed by making a table (called a *Cayley table*):

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

From the Cayley table, we can observe that 0 is an identity element, and that every element has an inverse. Only associativity requires some thought.

There is nothing special about 5 in the previous example. Addition mod n can be defined for any positive integer n. We denote the integers $\{0, 1, \ldots, n-1\}$ with addition mod n by $\mathbb{Z}_n$.

Example. Let $\mathbb{Q}[x]$ denote the set of polynomials (in the variable x) with rational coefficients. That is, elements of $\mathbb{Q}[x]$ have the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i, \qquad a_i \in \mathbb{Q}.$$

The smallest nonnegative integer such that $a_i \neq 0$ is called the *degree* of $p(x)$. For technical reasons that are not relevant right now, the degree of the zero polynomial is $-\infty$.

We add polynomials by grouping like terms. If two polynomials $p(x)$ and $q(x)$ with coefficients a_i and b_i, respectively, have the same degree n, then

$$p(x) + q(x) = \left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} b_i x^i\right) = \sum_{i=0}^{n} (a_i + b_i) x^i.$$

If they do not have the same degree, say $\deg q(x) = m < n$, then we can simply extend $q(x)$ by setting $b_i = 0$ for $i = m+1, m+2, \ldots, n$. It is clear that this is a binary operation, since $a_i + b_i \in \mathbb{Q}$ for all i.

Associativity of polynomial addition will follow from associativity on $\mathbb{Q}$. Let $p(x), q(x) \in \mathbb{Q}[x]$ be as above and let $r(x) = \sum c_i x^i$. As above, there is no loss in assuming each summation goes to n. We have,

$$\begin{aligned}
(p(x) + q(x)) + r(x) &= \left(\sum_{i=0}^{n} (a_i + b_i) x^i\right) + \left(\sum_{i=0}^{n} c_i x^i\right) \\
&= \left(\sum_{i=0}^{n} ((a_i + b_i) + c_i) x^i\right) \\
&= \left(\sum_{i=0}^{n} (a_i + (b_i + c_i)) x^i\right) \\
&= \left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} (b_i + c_i) x^i\right) \\
&= p(x) + (q(x) + r(x)),
\end{aligned}$$

as desired.

It is easy to show that the identity element is the zero polynomial, $z(x) = 0$. For $p(x) \in \mathbb{Q}[x]$ (written as above),

$$p(x) + z(x) = \left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} 0 x^i\right) = \sum_{i=0}^{n} (a_i + 0) x^i = \sum_{i=0}^{n} a_i x^i = p(x).$$

Since addition is commutative (check!), then $z(x) + p(x) = p(x)$ as well.

Finally, we claim that the inverse of a polynomial $p(x)$ (written as above) is the polynomial $-p(x)$ defined as $-p(x) = \sum(-a_i)x^i$. By the addition rule,

$$p(x) + (-p(x)) = \sum_{i=0}^{n} (a_i + (-a_i)) x^i = \sum_{i=0}^{n} 0 x^i = z(x).$$

Thus, $\mathbb{Q}[x]$ is a group under polynomial addition. (We typicaly just identify the zero polynomial with the scalar 0.)

One can easily replace $\mathbb{Q}$ in the previous example with $\mathbb{R}$ or $\mathbb{C}$.

Example. A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous at* $a \in \mathbb{R}$ if for every $\varepsilon > 0$, there is some $\delta > 0$ such that for all x with $0 < |x - a| < \delta$, we have $|f(x) - f(a)| < \varepsilon$. That is, we can guarantee that the y-values of f are sufficiently close together by choosing x values in a small enough interval (there are no big "jumps"). If f is continuous at all $a \in \mathbb{R}$, we say f is *continuous on* $\mathbb{R}$. We denote by $\mathscr{C}^0(\mathbb{R})$ the set of continuous real-valued functions on $\mathbb{R}$. That is, continuous functions $f : \mathbb{R} \to \mathbb{R}$.

Let $f, g \in \mathscr{C}^0(\mathbb{R})$. We define an operation on $\mathscr{C}^0(\mathbb{R})$ called *pointwise-addition*. That is, if $f, g \in \mathscr{C}^0(\mathbb{R})$, then $f + g$ is defined by the rule $(f + g)(x) = f(x) + g(x)$.

We claim that pointwise addition is a binary operation on $\mathscr{C}^0(\mathbb{R})$. Fix some $a \in \mathbb{R}$ and let $\varepsilon > 0$. Because f is continuous there is some $\delta_1 > 0$ such that $|f(x) - f(a)| < \varepsilon/2$ for all x such that $0 < |x - a| < \delta_1$. Similarly, g is continuous so there is some $\delta_2 > 0$ such that $|g(x) - g(a)| < \varepsilon/2$ for all x such that $0 < |x - a| < \delta_2$. Choose $\delta = \min\{\delta_1, \delta_2\}$. Then for all x such that $0 < |x - a| < \delta$, we have

$$\begin{aligned}
|(f + g)(x) - (f + g)(a)| &= |f(x) + g(x) - f(a) - g(a)| \\
&= |(f(x) - f(a)) + (g(x) - g(a))| \\
&\leq |f(x) - f(a)| + |g(x) - g(a)| \quad \text{by the triangle inequality} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

It follows that the sum of two continuous functions is continuous.

It is easy to check that pointwise addition is associative. Let $f, g, h \in \mathscr{C}^0(\mathbb{R})$ and let $x \in \mathbb{R}$. Then

$$\begin{aligned}
(f + (g + h))(x) &= f(x) + (g + h)(x) \\
&= f(x) + (g(x) + h(x)) \\
&= (f(x) + g(x)) + h(x) \quad \text{by associativity of real numbers} \\
&= (f + g)(x) + h(x) \\
&= ((f + g) + h)(x).
\end{aligned}$$

Let $z : \mathbb{R} \to \mathbb{R}$ be the function defined by $z(x) = 0$ for all $x \in \mathbb{R}$. Then z is continuous since for any $a \in \mathbb{R}$ and any $\epsilon > 0$, $|z(x) - z(a)| = 0 < \epsilon$ for all x. Moreover, z is an identity element on $\mathscr{C}^0(\mathbb{R})$ since for any $f \in \mathscr{C}^0(\mathbb{R})$,

$$(f + z)(x) = f(x) + z(x) = f(x) + 0 = f(x).$$

Finally for $f \in \mathscr{C}^0(\mathbb{R})$, define $-f : \mathbb{R} \to \mathbb{R}$ by $(-f)(x) = -f(x)$. We leave it as an exercise to verify that $-f$ is continuous. Note that

$$(f + (-f))(x) = f(x) + (-f)(x) = f(x) - f(x) = 0,$$

so $-f$ is the inverse of f under pointwise addition.

It follows that $\mathscr{C}^0(\mathbb{R})$ is a group under pointwise addition.

We now consider some general properties for groups.

Proposition 1: Cancellation laws

Let G be a group. Then for any three elements $s, t, u \in G$,

$$st = su \Rightarrow t = u$$

$$st = ut \Rightarrow s = u.$$

Proof. We prove the first property. The second is similar. Suppose $st = su$ for some $s, t, u \in G$. Since G is a group, then s has an inverse element, say $x \in G$, such that $xs = e$. Then

$$st = su$$

$$xst = xsu$$

$$et = eu$$

$$t = u,$$

as desired. □

We call the first property *left cancellation* and the second property *right cancellation*. Of course, for commutative groups these are the same thing. In particular, in additive notation we have

$$s + t = s + u \Rightarrow t = u.$$

These cancellation rules are sometimes called *sudoku rules* because they tell us that there cannot be repeated entries in a row or column of the Cayley table.

Exercise. Let $G = \{e, a, b, c\}$ be a group with identity element e. Show that there are exactly two possible Cayley tables for G.

Proposition 2: Properties of groups

Let G be a group.

(1) The identity element $e \in G$ is unique.
(2) The inverse of an element $s \in G$ is unique.
(3) If x is a left inverse of $s \in G$, then x is a right inverse.
(4) For all $s \in G$, $(s^{-1})^{-1} = s$.
(5) For all $s, t \in G$, $(st)^{-1} = t^{-1}s^{-1}$.
(6) If $s \in G$, then $ss = s$ if and only if $s = e$.

Proof. (1) Suppose $e, e' \in G$ are both identities. Then $e = ee' = e'$, so $e = e'$.

(2) Suppose $x, y \in G$ are inverses of $s \in G$. Then

$$x = xe = x(sy) = (xs)y = ey = y,$$

so $x = y$.

(3) Suppose $x \in G$ is a left inverse for $s \in G$, so $xs = e$. But then

$$s^{-1} = (xs)s^{-1} = x(ss^{-1}) = xe = x.$$

That is, a left inverse is *the* two-sided inverse of s.

(4) This follows by symmetry of the definition of inverse, and the uniqueness of inverses.

(5) By uniqueness, it suffices to show that $t^{-1}s^{-1}$ is *an* inverse of st. We check:

$$(st)(t^{-1}s^{-1}) = s(tt^{-1})s^{-1} = ses^{-1} = ss^{-1} = e.$$

(6) If $ss = s$, then $ss = se$ and the result follows by left cancellation. □

Property (5) is often called *socks on-shoes on-shoes off-socks off*.

We can define exponents as one is used to. For positive n:

$$s^n = s \cdot s \cdots s \quad (n \text{ times})$$
$$s^{-n} = s^{-1} \cdot s^{-1} \cdots s^{-1} \quad (n \text{ times})$$

We set $s^0 = e$. However, remember that groups are not assumed to be commutative, so $(st)^2 = stst$ but we cannot reduce this unless we know something about the group (so, in general, $(st)^2 \neq s^2t^2$).

In additive notation, we use coefficients instead of exponents. Remember here we do assume the operation is commutative:

$$-(-s) = s$$
$$-(s+t) = -s - t$$
$$ns = s + s + \cdots s \quad (n \text{ times})$$
$$n(s+t) = ns + nt$$

When studying a new mathematical object, there are often three main ideas that one should explore early on: sub-objects, maps between objects, and quotient objects.

Definition: Subgroup

Let $(G, *)$ be a group. A nonempty subset H of G is called a *subgroup* of G if it is a group with respect to the operation $*$ on G.

Associativity (and commutativity) are properties of the operation $*$, and not the set G, so there is no need to check these properties for H. What remains to be checked are the following properties:

(1) The operation $*$ is a binary operation on H (H is closed under the operation).
(2) The identity element $e \in G$ is an element in H. (Hence, e is the identity element in H.)
(3) If $s \in H$, then $s^{-1}H$ (H is closed under inverses).

Theorem 3: The subgroup test

A nonempty subset H of a group G is a subgroup of G if and only if for all $s, t \in H$, $st^{-1} \in H$.

Proof. ($\Rightarrow$) Assume H is a subgroup. Then $H \neq \emptyset$ because $e \in H$. If $s, t \in H$, then $t^{-1} \in H$ by closure under inverses. Now by closure of multiplication, $st^{-1} \in H$.

($\Leftarrow$) Assume $H \neq \emptyset$ and $s, t \in H$ implies $st^{-1} \in H$. We will verify (1), (2), and (3) above.

(1) Let $s \in H$ (this exists because H is non-empty). Then $e = ss^{-1} \in H$.
(2) Since $e \in H$ by (1), then for any $s \in H$ we have $s^{-1} = es^{-1} \in H$.
(3) Given $s, t \in H$, we have $t^{-1} \in H$ by (2) and so $st = s(t^{-1})^{-1} \in H$.

Thus, H is a subgroup. □

Example. The following are examples of subgroups:

(1) In any group, the *trivial subgroup* consisting of only identity is a subgroup. Also, the group itself is a subgroup.
(2) Let n be an integer. Consider the set

$$n\mathbb{Z} = \{na : a \in \mathbb{Z}\}.$$

Since $0 = n0 \in n\mathbb{Z}$, then $n\mathbb{Z} \neq \emptyset$. Let $na, nb \in n\mathbb{Z}$, then $na - nb = n(a - b) \in n\mathbb{Z}$. Thus, $n\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

(3) Let G be a group and $a \in G$. Consider the set

$$\langle a \rangle = \{a^n : n \in \mathbb{Z}\}.$$

Of course, $1 = a^0 \in \langle a \rangle$ so $\langle a \rangle \neq \emptyset$. For any $n, m \in \mathbb{Z}$, $a^n(a^m)^{-1} = a^{n-m} \in \langle a \rangle$. Hence, $\langle a \rangle$ is a subgroup of G. We call $\langle a \rangle$ the *cyclic subgroup* of G generated by a.

(4) Recall that $\mathscr{C}^0(\mathbb{R})$ is the group of continuous real-valued functions on $\mathbb{R}$. Let

$$I = \{f \in \mathscr{C}^0(\mathbb{R}) : f(0) = 0\}.$$

The zero function is clearly in I so $I \neq \emptyset$. For $f, g \in I$ we have

$$(f - g)(0) = f(0) - g(0) = 0 - 0 = 0,$$

so $f - g \in I$. Hence, I is a subgroup of $\mathscr{C}^0(\mathbb{R})$.

(5) The group $\mathbb{Z}_6$ has four subgroups:

$$\{0\}, \quad \{0,2,4\}, \quad \{0,3\}, \quad \mathbb{Z}_6.$$

2. Group homomorphisms

Now we study maps between groups. In general, maps between mathematical objects should be *structure preserving* in some way.

Definition: Group homomorphism

Let G and H be a groups. A *group homomorphism* is a function $\phi : G \to H$ such that

$$\phi(st) = \phi(s)\phi(t)$$

for all $s, t \in G$.

A group homomorphism is, first and foremost, a (well-defined) function on sets. Note that on the left-hand side of the definition, the operation between s and t occurs in G, while on the right-hand side the operation between $\phi(s)$ and $\phi(t)$ occurs in H.

In additive notation, the homomorphism property becomes

$$\phi(s + t) = \phi(s) + \phi(t).$$

Be careful! At times we might have homomorphisms between additive and non-additive groups.

Example. The following are examples of group homomorphisms:

(1) The map $\phi : \mathbb{Z} \to 2\mathbb{Z}$ defined by $\phi(n) = 2n$ is a group homomorphism. Let $n, m \in \mathbb{Z}$, then

$$\phi(n + m) = 2(n + m) = 2n + 2m = \phi(n) + \phi(m).$$

(2) Recall that $\mathscr{C}^1(\mathbb{R})$ is the set of functions with continuous derivative. The *differentiation map* is $D : \mathscr{C}^1(\mathbb{R}) \to \mathscr{C}^0(\mathbb{R})$ is defined by $D(f) = f'$. Then for $f, g \in \mathscr{C}^1(\mathbb{R})$,

$$D(f + g) = (f + g)' = f' + g' = D(f) + D(g).$$

A similar argument works for integration.

(3) Let $\psi : \mathbb{R} \to \mathbb{R}_+^\times$ be the exponential map, so $\psi(x) = e^x$. Then for $x, y \in \mathbb{R}$,

$$\psi(x + y) = e^{x+y} = e^x e^y = \psi(x)\psi(y).$$

We now consider various properties of group homomorphisms.

Proposition 4: Composition of homomorphisms

Let $\phi : G \to H$ and $\psi : H \to K$ be group homomorphisms. Then $\psi \circ \phi : G \to K$ is a group homomorphism.

Proof. By set theory, $\psi \circ \phi$ is a function. Let $s, t \in G$. Then

$$(\psi \circ \phi)(st) = \psi(\phi(st)) = \psi(\phi(s)\phi(t)) = \psi(\phi(s))\psi\phi(t) = (\psi \circ \phi)(s)(\psi \circ \phi)(t). \qquad \square$$

Given a group G, we will often write the identity as e_G (or 0_G if the group is additive).

Proposition 5: Properties of group homomorphisms

Let $\phi : G \to H$ be a group homomorphism. Then the following properties hold:

(1) $\phi(e_G) = e_H$,
(2) $\phi(s^{-1}) = \phi(s)^{-1}$ for all $s \in G$, and
(3) $\phi(s^m) = \phi(s)^m$ for all $s \in G$ and all integers m

Proof. (1) Let $t = \phi(e_G)$. Then

$$t = \phi(e_G) = \phi(e_G e_G) = \phi(e_G)\phi(e_G) = tt.$$

By properties of groups, $t = e_H$.

(2) By part (1) we have

$$e_H = \phi(e_G) = \phi(ss^{-1}) = \phi(s)\phi(s^{-1}).$$

Uniqueness of inverses gives $\phi(s)^{-1} = \phi(s^{-1})$.

(3) Exercise. □

In additive notation, the properties above become

(1) $\phi(0_G) = 0_H$,
(2) $\phi(-s) = -\phi(s)$ for all $s \in G$, and
(3) $\phi(ms) = m\phi(s)$ for all $s \in G$ and all integers m.

Definition: Isomorphism

A bijective group homomorphism is called an *isomorphism*. If there is an isomorphism between groups G and H, we write $G \cong H$.

Given a group homomorphism $\phi : G \to H$, the *image of* ϕ is the usual function-theoretic image:

$$\mathrm{Im}(\phi) = \{t \in H : \phi(s) = t \text{ for some } s \in G\}.$$

We now come to another definition which will play a very important role throughout the semester.

Definition: Kernel

Let $\phi : G \to H$ be a group homomorphism. The *kernel of* ϕ is defined by

$$\mathrm{Ker}(\phi) = \{s \in G : \phi(s) = e_H\}.$$

Another way to think about the kernel is that it is the preimage of e_H. That is, $\mathrm{Ker}(\phi) = \phi^{-1}(e_H)$.

Example. Consider the differentiation function $D : \mathscr{C}^1(\mathbb{R}) \to \mathscr{C}^0(\mathbb{R})$. Then $\mathrm{Ker}(D)$ is the set of constant functions in $\mathscr{C}^1(\mathbb{R})$. We make two observations that will be explained in subsequent results. The first is that the set of constant functions in $\mathscr{C}^1(\mathbb{R})$ is in fact a subgroup (easy exercise!). The

second is that, if $f \in \mathscr{C}^0(\mathbb{R})$, then it has a preimage in $F \in \mathscr{C}^1(\mathbb{R})$ (that is, it has an antiderivative). Moreover, by the Fundamental Theorem of Calculus, the inverse image of f is

$$D^{-1}(f) = \{F + c : c \in \mathbb{R}\}.$$

That is, the inverse images of f differ up to an element of the kernel!

Proposition 6: Kernel and image are subgroups

Let $\phi : G \to H$ be a group homomorphism. Then $\mathrm{Ker}(\phi)$ is a subgroup of G and $\mathrm{Im}(\phi)$ is a subgroup of H.

Proof. Since $\phi(e_G) = \phi(e_H)$ by properties of homomorphisms, then $\mathrm{Ker}(\phi) \neq \emptyset$. Let $s, t \in \mathrm{Ker}\phi$. Then by properties of homomorphisms,

$$\phi(st^{-1}) = \phi(s)\phi(t)^{-1} = e_H e_H^{-1} = e_H.$$

Thus, $st^{-1} \in \mathrm{Ker}(\phi)$ and so $\mathrm{Ker}(\phi)$ is a subgroup of G.

The proof for the image is left as a homework exercise. □

We now explain the second observation in the previous example.

Proposition 7: Kernel parameterizes inverse images

Let $\phi : G \to H$ be a group homomorphism. Let $s \in G$ and set $\phi(s) = t$. Then

$$\phi^{-1}(t) = \{sk : k \in \mathrm{Ker}(\phi)\}.$$

In additive notation, the above display is $\phi^{-1}(t) = \{s + k : k \in \mathrm{Ker}(\phi)\}$.

Proof. Let $k \in \mathrm{Ker}(\phi)$. Then

$$\phi(sk) = \phi(s)\phi(k) = te_H = t,$$

so $sk \in \phi^{-1}(t)$. That is, $\{sk : k \in \mathrm{Ker}(\phi)\} \subseteq \phi^{-1}(t)$.

On the other hand, suppose $u \in \phi^{-1}(t)$. Then $\phi(u) = t = \phi(s)$ and so

$$\phi(s^{-1}u) = \phi(s)^{-1}\phi(u) = t^{-1}t = e_H.$$

That is, $s^{-1}u \in \mathrm{Ker}(\phi)$. Now $u = s(s^{-1}u) \in \{sk : k \in \mathrm{Ker}(\phi)\}$. Hence,

$$\phi^{-1}(t) \subseteq \{sk : k \in \mathrm{Ker}(\phi)\}.$$

Equality now follows by definition of set equality. □

The kernel of a homomorphism must contain the identity. When the kernel only contains the identity we say it is *trivial*.

Corollary 8: Trivial kernel equivalent to injectivity

Let $\phi : G \to H$ be a group homomorphism. Then ϕ is injective if and only if $\text{Ker}(\phi)$ is trivial.

Proof. Suppose ϕ is injective. Then the preimage of any element can have at most one element. Since $\text{Ker}(\phi) = \phi^{-1}(e_H)$ and $e_G \in \phi^{-1}(e_H)$, then $\text{Ker}(\phi) = \{e_G\}$.

Now suppose $\text{Ker}(\phi)$ is trivial. Suppose $s, s' \in G$ such that $\phi(s) = \phi(s')$. But then $s^{-1}s' \in \text{Ker}(\phi) = \{e_G\}$. That is, $s^{-1}s' = e_G$ so $s' = s$. That is, ϕ is injective. □

3. Rings and Fields

We end this set of notes by (briefly) defining a ring and a field.

Definition: Ring,field

A *ring* is a set R along with two binary operations, typically denoted $+$ and $*$, satisfying the following properties:

(1) $(R, +)$ is an abelian group,
(2) $*$ is associative,
(3) the left and right distributive properties hold:

$$r * (s + t) = r * s + r * t$$
$$(s + t) * r = s * r + t * r$$

for all $s, t, r \in R$.

A *field* is a ring F in which $(F\backslash\{0\}, *)$ is an abelian group.

We typically write the additive identity of R as 0_R. A ring need not have a multiplicative identity. When it does, the ring is said to have *unity*. We say the ring R is *commutative* if the multiplication operation is commutative.

Example.
(1) The integers $\mathbb{Z}$ along with the operations of addition and multiplication is a ring but is not a field because not every element has a multiplicative inverse.
(2) The rationals $\mathbb{Q}$, the reals $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are all examples of fields.
(3) The set $\mathscr{C}^0(\mathbb{R})$ along with pointwise multiplication and addition is a ring. It is not a field, but it does have multiplicative identity: the constant function $f(x) = 1$.
(4) The set $\mathbb{Z}_3$ along with addition and multiplication mod 3 is a field. This works for any $\mathbb{Z}_p$ with prime p. In fact we denote this field $\mathbb{F}_p$. However, if n is composite, then $\mathbb{Z}_n$ is a ring but not a field.

Proposition 9: Properties of rings

Let A be a ring with unity 1. Then the following hold for all $a, b \in A$:

(1) $0a = 0 = a0$,
(2) $a(-b) = -(ab) = (-a)b$,
(3) $(-a)(-b) = ab$,
(4) $(-1)a = -a$,
(5) (-1)(-1) = 1.

Proof. (1) We have $0a = (0 + 0)a = 0a + 0a$. Subtracting $0a$ from both sides gives the result. Showing $a0 = 0$ is similar.

(2) To prove $a(-b) = -(ab)$ it suffices to show that $a(-b)$ is an additive inverse to ab. By the distributive properties and part (1),

$$a(-b) + ab = a(-b + b) = a0 = 0.$$

Properties (3)-(5) are special cases of (2). □

Now we state our final definition. Though we will not use this, it will put our study of vector spaces in a broader context.

Definition: Module

Let R be a commutative ring with unity 1. A *module* over R is an abelian group $(M, +)$ and a multiplication operation $R \times M \to M$ (called an *action*) such that for all $r, s \in R$ and $m, n \in M$ we have

(1) $r(sm) = (rs)m$,
(2) $(r + s)m = rm + sm$,
(3) $r(m + n) = rm + rn$,
(4) $1m = m$.

One can define modules for noncommutative rings and rings without unity, but that will not be necessary here.

Example. (1) Any ring R is a module over itself.

(2) Let R^n be the cartesian product of n copies of the ring R. Then R^n is a module over R with action defined by

$$r(r_1, r_2, \ldots, r_n) = (rr_1, rr_2, \ldots, rr_n).$$

(3) Let F be a field and let $F[x]$ be the set of polynomials with coefficients in F. Then $F[x]$ is a module over F with action defined by

$$c(a_0 + a_1x + \cdots + a_nx^n) = ca_0 + ca_1x + \cdots + ca_nx^n.$$

Definition: Vector space

A vector space is a module $(V, +)$ over a field F.

The Well-Ordering Principle

Every nonempty subset of $\mathbb{N}$ contains a least element.

We say that nonempty subsets of $\mathbb{N}$ are *well-ordered*.

The next theorem formalizes the idea of division that you learned in grade school. We will use this to classify subgroups of $\mathbb{Z}$. It is also our first example of an *existence and uniqueness proof*.

Theorem 10: The Division Algorithm

Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique integers q and r such that $a = bq + r$ with $0 \leq r < b$.

Proof. (Existence) Let

$$S = \{a - bk : k \in \mathbb{Z} \text{ and } a - bk \geq 0\} \subseteq \mathbb{N}.$$

If $0 \in S$, then choose $q = a/b$ and $r = 0$. Assume $0 \notin S$. If $a \geq 0$, then $a = a - b \cdot 0 \in S$. If $a < 0$, then $a - b(2a) = a(1 - 2b) \in S$. In either case, $S \neq \emptyset$ and so we may apply the Well-Ordering Principle to find a least member of S, say r. By definition of S, there exists an integer q such that $r = a - bq \geq 0$. We claim $r < b$. Suppose otherwise. Then

$$a - b(q + 1) = a - bq - b = r - b > 0.$$

But then $a - b(q + 1) \in S$ and $a - b(q + 1) < r$, contradicting the choice of r.

(Uniqueness) Suppose there exist r, r', q, q' such that $a = bq + r$ and $a = bq' + r'$ with $0 \leq r, r' < b$. Then $bq + r = bq' + r'$. Assume $r' \geq r$. Then $b(q - q') = r - r'$ so b divides $r' - r$ and $0 \leq r' - r \leq r' < b$. Thus, $r' - r = 0$ so $r = r'$ and $q = q'$. □

In light of this result, we say that b *divides* a if $a = bq + 0$ for some (unique) $q \in \mathbb{Z}$. We will often use the notation $b \mid a$ in place of b divides a. We will also say that b is a *factor* of a.

The *gcd* of integers m and n is defined as the largest common factor of m and n, denoted $\gcd(m, n)$. Note that by definition $\gcd(m, n)$ divides both m and n. We say m and n are *relatively prime* if $\gcd(m, n) = 1$.

Theorem 11: Bézout's identity

Let $a, b \in \mathbb{Z}$. There exist integers x, y such that $\gcd(a, b) = ax + by$. Furthermore, the gcd of a and b is unique.

Proof. Let $S = \{s > 0 : s = ax + by \text{ for } x, y \in \mathbb{Z}\}$. Then $|a| + |b| \in S$ so $S \neq \emptyset$. By the Well-Ordering Principle, there exists a least element, $t \in S$. Then $t = au + bv$ for some $u, v \in \mathbb{Z}$ by definition of S.

We claim t is a common divisor of a and b. By the division algorithm, there exists $q, r \in \mathbb{Z}$ such that $a = tq + r$ and $0 \leq r < t$. Then

$$r = a - tq = a - (au + bv)q = a(1 - uq) + b(-qv).$$

If $r > 0$, then $r \in S$ and $r < t$, a contradiction. Thus, $r = 0$ so t divides a. Similarly, t divides b.

Set $\gcd(a, b) = d$. Then $t \mid a$ and $t \mid b$, so $t \mid \gcd(a, b)$. But d divides a and b, so d divides $au+bv = t$. Since $d \mid t$ and $t \mid d$, then $t = d$. □

Note that there is no claim to uniqueness of the x, y in Bézout's identity. However, one can show that any linear combination $ax+by$ is a multiple of d. Thus, $ax+by = 1$ if and only if $\gcd(a, b) = 1$ (a, b are relatively prime).

The process of finding the gcd and the integers x, y is known as the *Euclidean Algorithm* and is demonstrated in the next example.

Example. Calculate $d = \gcd(471, 562)$ and find integers x and y such that $d = 471x + 562y$.

We repeatedly apply the division algorithm.

$$\begin{aligned}
562 &= 471 \cdot 1 + 91 \\
471 &= 91 \cdot 5 + 16 \\
91 &= 16 \cdot 5 + 11 \\
16 &= 11 \cdot 1 + 5 \\
11 &= 5 \cdot 2 + 1 \\
5 &= 1 \cdot 5 + 0.
\end{aligned}$$

Thus, $d = 1$. That is, 471 and 562 are relatively prime. Now by reversing:

$$\begin{aligned}
1 &= 11 + (-2) \cdot 5 = 11 + (-2)[16 + (-1) \cdot 11] \\
&= (3) \cdot 11 + (-2) \cdot 16 = (3) \cdot [91 + (-5) \cdot 16] + (-2) \cdot 16 \\
&= (3) \cdot 91 + (-17) \cdot 16 = (3) \cdot 91 + (-17) \cdot [471 + (-5) \cdot 91] \\
&= (88) \cdot 91 + (-17) \cdot 471 = (88) \cdot [562 + (-1) \cdot 471] + (-17) \cdot 471 \\
&= (88) \cdot 562 + (-105) \cdot 471
\end{aligned}$$

Hence, $x = -105$ and $y = 88$.

The next result classifies subgroups of $\mathbb{Z}$ using the division algorithm.

Proposition 12: Subgroups of $\mathbb{Z}$

Let H be a subgroup of $\mathbb{Z}$, then $H = d\mathbb{Z}$ for some integer d.

Proof. First consider the case that $H = \{0\}$ (the trivial subgroup). Then $H = 0\mathbb{Z}$ and we are done. Now we consider the case that $H \neq \{0\}$.

Because H is a subgroup, $H \neq \emptyset$. Moreover, since H is closed under inverses, then we may assume that H contains a positive integer. The set of positive integers in H is a subset of $\mathbb{N}$, and so by the Well-Ordering principle, choose the least positive integer d. We claim $H = d\mathbb{Z}$.

Since H is a subgroup and closed under the operations, then $d\mathbb{Z} \subseteq H$. We will show that the opposite inclusion holds. Let $a \in H$. We claim $a \in d\mathbb{Z}$.

If $a = 0$ then clearly $a \in d\mathbb{Z}$ so assume $a \neq 0$. By the division algorithm, there exists integers q, r such that $a = dq + r$ and $0 \leq r < d$. Write $r = a - dq$. Since $a \in H$ and $d \in H$, then $a - dq \in H$. But $r < d$, a contradiction unless $r = 0$. That is, $a = dq \in d\mathbb{Z}$, proving the claim. Thus, $H \subseteq d\mathbb{Z}$ and equality holds. $\square$

Vector spaces and linear transformations

Here we restate the definition of a vector space and study maps between vector spaces.

1. Vector spaces and subspaces

Throughout, k is a field. For most of our discussion, $\mathbb{C}$ and $\mathbb{R}$ are sufficient examples. However, it also helps to think about fields such as $\mathbb{Z}_p$ (p prime).

Definition: Vector space

A *vector space over a field* k is an additive abelian group $(V, +)$ along with an operation $k \times V \to V$ called scalar multiplication satisfying the following axioms for all $\lambda, \mu \in k$ and all $v, w \in V$:

(1) $(\lambda\mu)v = \lambda(\mu v)$,
(2) $(\lambda + \mu)v = \lambda v + \mu v$,
(3) $\lambda(v + w) = \lambda v + \lambda w$,
(4) $1v = v$.

The elements of V are called *vectors* and the elements of k are called *scalars*. We denote the zero element of V by $\mathbf{0}$ (the *zero vector*) and the additive identity in k by 0.

We now look at several examples of vector spaces, several of which should be very familiar.

Example. Any field k is a vector space over itself. This follows directly from the axioms of rings.

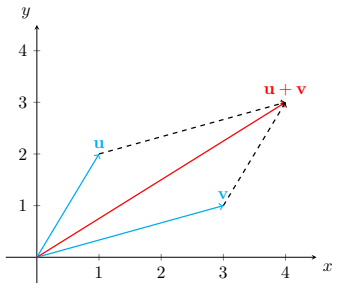
Example. For any field k, the set k^2 denotes ordered pairs (x_1, x_2), $x_1, x_2 \in k$. The addition and scalar multiplication operations are defined, respectively, by

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \text{ for all } (x_1, x_2), (y_1, y_2) \in k^2$$
$$\lambda(x_1, x_2) = (\lambda x_1, \lambda x_2) \text{ for all } (x_1, x_2) \in k^2, \lambda \in k.$$

The zero vector is $\mathbf{0} = (0, 0)$, and the additive identity of (x_1, x_2) is just $(-x_1, -x_2)$.

When $k = \mathbb{R}$, then $k^2 = \mathbb{R}^2$ can be represented as the usual Cartesian plane. We draw vectors as arrows from the origin (initial point/tail) to the corresponding point (terminal point/head). Addition corresponds to the *parallelogram law:*

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapter 3. Last Updated: September 30, 2021



Scalar multiplication scales a vector and reverses direction when $\lambda < 0$.

Example. We extend the previous example. Let k be a field and let n be a positive integer. Define k^n to be the set of all n-tuples of elements of k:

$$k^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in k\}.$$

Addition and scalar multiplication are component-wise:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n) \text{ for all } (x_1, \ldots, x_n), (y_1, \ldots, y_n) \in k^2$$
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n) \text{ for all } (x_1, \ldots, x_n) \in k^2, \lambda \in k.$$

The vector space $\mathbb{R}^n$ is called *real n-space* while $\mathbb{C}^n$ is called *complex n-space*. We will typically use bold notation to represent vectors:

$$\mathbf{x} = (x_1, \ldots, x_n)$$

though we may also use the notation $\overrightarrow{x}$.

Example. Recall that $\mathscr{C}^0(\mathbb{R})$ denotes the additive abelian group of continuous real-valued functions on $\mathbb{R}$. We define a scalar multiplication of $\mathbb{R}$ on $\mathscr{C}^0(\mathbb{R})$ by

$$(\lambda f)(x) = \lambda f(x).$$

One can similarly define $\mathscr{C}^n(\mathbb{R})$ as the set of real valued functions on $\mathbb{R}$ have continuous nth derivatives (defined everywhere). One can also replace $\mathbb{R}$ by $\mathbb{C}$.

Example. The set $\mathbb{Q}[x]$ of polynomial functions (or just polynomials) with rational coefficients is a vector space over $\mathbb{Q}$. Addition and scalar multiplication are the usual operations. Similarly, $\mathbb{R}[x]$ and $\mathbb{C}[x]$ are vector spaces over $\mathbb{R}$ and $\mathbb{C}$, respectively.

The following properties hold for all vector spaces.

Proposition 1: Vector space properties

Let V be a vector space over a field k. Then the following assertions hold for all $\lambda \in k$ and $v \in V$:

(1) $\lambda \mathbf{0} = \mathbf{0}$,
(2) $0v = \mathbf{0}$,
(3) $(-\lambda)v = -(\lambda v)$,
(4) $\lambda v = \mathbf{0}$ if and only if $\lambda = 0$ or $v = \mathbf{0}$.

Proof. (1) By right distributivity:

$$\lambda \mathbf{0} = \lambda(\mathbf{0} + \mathbf{0}) = \lambda \mathbf{0} + \lambda \mathbf{0}.$$

The result now follows by cancellativity in V.

(2) By left distributivity,

$$0v = (0 + 0)v = 0v + 0v.$$

Again, the result now follows by cancellativity in V.

(3) By left distributivity,

$$\lambda v + (-\lambda v) = (\lambda + (-\lambda))v = 0v = \mathbf{0}.$$

The result follows from uniqueness of inverses in V.

(4) Sufficiency follows from (i) and (ii). Suppose $\lambda v = \mathbf{0}$ and $\lambda \neq 0$. Then by associativity,

$$\mathbf{0} = \lambda^{-1}\mathbf{0} = \lambda^{-1}(\lambda v) = (\lambda^{-1}\lambda)v = 1v = v,$$

so $v = \mathbf{0}$. □

Definition: Subspace

A subset W of a vector space V over a field k is called a *subspace* of V if it constitutes a vector space over K in its own right with respect to the additive and scalar operations defined on V.

As with subgroups, there is an easy check for subspaces.

Theorem 2: Subspace Criterion

Let W be a nonempty subset of the vector space V. Then W is a subspace of V if and only if it is closed under addition and scalar multiplication.

Proof. If W is a subspace, then clearly it is closed under addition and scalar multiplication. On the other hand, suppose W is closed under those operations. Let $v, w \in W$. Then $v - w = v + (-w) \in W$, so W is an additive subgroup of V by the subgroup test. Since W is closed under scalar multiplication, then the remaining axioms for follow immediately because they hold in V. □

Example. The following are examples of subgroups:

(1) For any vector space V, The *trivial subspace* $\{\mathbf{0}\}$ and V itself are subspaces of V.

(2) Let $V = \mathbb{R}^2$. Consider the set $W = \{(x, 0) : x \in \mathbb{R}\}$. (Geometrically, W corresponds to the x-axis in $\mathbb{R}^2$.) Then for $(x_1, 0), (x_2, 0) \in W$ and $\lambda \in \mathbb{R}$, we have

$$(x_1, 0) + (x_2, 0) = (x_1 + x_2, 0) \in W \quad \text{and} \quad \lambda(x_1, 0) = (\lambda x_1, 0) \in W,$$

so W is a subspace by the Subspace Criterion.

One can similarly show that the set $\{(0, y) : y \in \mathbb{R}\}$ is a subspace of V (which corresponds to the y-axis). More generally, any line through the origin corresponds to a subspace of V.

(3) Let $V = \mathbb{Q}[x]$ and let W be the set of polynomials of degree at most n. Since the sum of two such polynomials has degree at most n, then W is closed under addition. Similarly, the scalar multiple of such a polynomial has degree at most n, so W is closed under scalar multiplication.

We now turn to two of the most important definitions in the study of linear algebra.

Definition: Linear combination, span

Let V be a vector space. A *linear combination* of vectors $v_1, v_2, \ldots, v_n$ is an expression of the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n$, $\lambda_i \in k$. The set of all such linear combinations is called the *span* of $v_1, v_2, \ldots, v_n$, denoted $\text{span}(v_1, v_2, \ldots, v_n)$.

Proposition 3: Span is a subspace

Let $v_1, \ldots, v_n$ be a family of vectors in a vector space V. Then $W = \text{span}(v_1, v_2, \ldots, v_n)$ is a subspace of V.

Proof. Let $\lambda_1 v_1 + \cdots + \lambda_n v_n$ and $\mu_1 v_1 + \cdots + \mu_n v_n$ be elements of W, that is, linear combinations of the v_i (so $\lambda_i, \mu_i \in k$ for all i). Then

$$(\lambda_1 v_1 + \cdots + \lambda_n v_n) + (\mu_1 v_1 + \cdots + \mu_n v_n) = (\lambda_1 + \mu_1)v_1 + \cdots + (\lambda_n + \mu_n)v_n \in W.$$

Thus, W is closed under addition. Let $\nu \in k$, then

$$\nu(\lambda_1 v_1 + \cdots + \lambda_n v_n) = (\nu\lambda_1)v_1 + \cdots + (\nu\lambda_n)v_n \in W,$$

so W is closed under scalar multiplication. Thus, W is a subspace by the Subspace Criterion. □

We say that $W = \text{span}(v_1, v_2, \ldots, v_n)$ is *generated* or *spanned* by $v_1, \ldots, v_n$, or that these vectors *span* W. If V is spanned by a finite collection of vectors, we say V is *finitely generated* or *finite dimensional*. However, the *dimension* is not necessarily the number of vectors that define the spanning set. We need to talk about efficient spanning sets, which is called a *basis*, in order to properly define dimension.

Example. (1) Let $V = \mathbb{R}^2$. The span of the vector $(1,1)$ is the set of multiples of $(1,1)$. Geometrically, this corresponds to the line through the origin and the point $(1,1)$. In fact, even in $\mathbb{R}^n$, the span of a single nonzero vector corresponds to the line through the origin and that point.

Now consider the span of the vectors $(1,1)$ and $(1,0)$. If (x_1, x_2) is in this span, then there are scalars $a, b \in \mathbb{R}$ such that $a(1,1) + b(1,0) = (x_1, x_2)$. This gives a system of linear equations:

$$a + b = x_1$$

$$a = x_2.$$

Note that for any choice of x_1, x_2 there is a solution given by $a = x_2$ and $b = x_1 - x_2$. That is, $(1,1)$ and $(1,0)$ span all of $\mathbb{R}^2$.

(2) Let $V = k^n$ where k is any field. We denote by $\mathbf{e}_i$ the vector whose ith component is 1 and the remaining components are zero. The vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are called the *canonical basis vectors* of V. Note that for any vector $(x_1, \ldots, x_n) \in V$ we have

$$(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j e_j.$$

Thus, the canonical basis vectors span V, so V is finite dimensional.

(3) Let $V = \mathbb{Q}[x]$ and let W be the set of polynomials of degree at most n. Then W is spanned by the vectors $1, x, x^2, \ldots, x^n$.

(4) The general solution to the ordinary differential equation $y'' - 4y = 0$ is $y = c_0 e^{2x} + c_1 e^{-2x}$. Hence, the set of solutions is the span of e^{2x} and e^{-2x} in $\mathscr{C}^2(\mathbb{R})$.

Exercise. If W_0, W_1 are subspaces of a vector space V. Set

$$W_0 + W_1 = \{w_0 + w_1 : w_0 \in W_0, w_1 \in W_1\}.$$

Show that $W_0 + W_1$ is a subspace of V.

2. Linear transformations

Just as a group homomorphism preserves the structure of a group, a linear transformation between vector spaces preserves the structure of the vector spaces.

Definition: Linear transformation

Let V and V' be vector spaces over a field k. A function $T : V \to V'$ is a *linear transformation* if it satisfies the following for all $v, w \in V$ and $\lambda \in k$:

$$T(v + w) = T(v) + T(w)$$
$$T(\lambda v) = \lambda T(v)$$

The first condition implies that T is a group homomorphism between the additive groups $(V, +)$ and $(V', +)$. This implies, by properties of group homomokrphisms:

(1) $T(\mathbf{0}) = \mathbf{0}$,
(2) $T(-v) = -T(v)$ for all $v \in V$, and
(3) $T(mv) = mT(v)$ for all $v \in V$ and $m \in \mathbb{Z}$.

Example. (1) The *zero map* $V \to V'$ that maps every element to $\mathbf{0}$ is a linear transformation.

(2) Let $V = k^n$. The *projection onto the jth coordinate* is the map $\rho_j : V \to k$ defined by $\rho_j(x_1, \ldots, x_n) = x_j$ for all $(x_1, \ldots, x_n) \in k^n$.

(3) Let $V = \mathbb{R}^2$ and let $a, b, c, d \in \mathbb{R}$. Define a map $T : V \to V$ by

$$T(x_1, x_2) = (ax_1 + bx_2, cx_1 + dx_2).$$

We claim this map is a linear transformation. Let $(x_1, x_2), (y_1, y_2) \in V$ and $\lambda \in k$. Then

$$\begin{aligned} T((x_1, x_2) + (y_1, y_2)) &= T(x_1 + y_1, x_2 + y_2) = (a(x_1 + y_1) + b(x_2 + y_2), c(x_1 + y_1) + d(x_2 + y_2)) \\ &= (ax_1 + bx_2, cx_1 + dx_2) + (ay_1 + by_2, cy_1 + dy_2) = T(x_1, x_2) + T(y_1, y_2) \\ T(\lambda(x_1, x_2)) &= T(\lambda x_1, \lambda x_2) = (a\lambda x_1 + b\lambda x_2, c\lambda x_1 + d\lambda x_2) \\ &= \lambda(ax_1 + bx_2, cx_1 + dx_2) = \lambda T(x_1, x_2). \end{aligned}$$

Question: When is this map invertible? (What are the conditions on a, b, c, d?)

(4) The differentiation operator $D : \mathscr{C}^1(\mathbb{R}) \to \mathscr{C}^0(\mathbb{R})$ is a linear transformation. This follows from basic facts of differentiation.

We now consider various properties of linear transformations. We will use extensively our properties of group homomorphisms.

Proposition 4: Composition of linear transformations

Let $T : V \to V'$ and $T' : V' \to V''$ be linear transformations. The $T' \circ T$ is a linear transformation.

Proof. The additivity property follows from properties of group homomorphisms. Let $v \in V$ and $\lambda \in k$. Then

$$(T' \circ T)(\lambda v) = T'(T(\lambda v)) = T'(\lambda T(v)) = \lambda T'(T(v)) = \lambda(T' \circ T)(v).$$

Thus, $T' \circ T$ is a linear transformation. □

We define kernel and image just as in the case of group homomorphisms.

Proposition 5: Kernel and image are subspaces

Let $T : V \to V'$ be a linear transformation. Then $\mathrm{Ker}(T)$ is a subspace of V and $\mathrm{Im}(T)$ is a subspace of V'.

Proof. By properties of group homomorphisms, $\mathrm{Ker}(T)$ is a subgroup of V. Let $v \in \mathrm{Ker}(T)$ and $\lambda \in k$. Then $T(\lambda v) = \lambda T(v) = \lambda \mathbf{0} = \mathbf{0}$. Thus, $\lambda v \in \mathrm{Ker}(T)$, so $\mathrm{Ker}(T)$ is a subspace of V.

Similarly, $\mathrm{Im}(T)$ is a subgroup of V'. Let $w \in \mathrm{Im}(T)$ and $\lambda \in k$. Then there exists $v \in V$ such that $T(v) = w$. Then $T(\lambda v) = \lambda T(v) = \lambda w$, so $\lambda w \in \mathrm{Im}(T)$. Thus $\mathrm{Im}(T)$ is a subspace of V'. □

Definition: Isomorphism

A bijective linear transformation is called an *isomorphism* (of vector spaces). If there is an isomorphism between vector spaces V and V', we write $V \cong V'$.

Proposition 6: Kernel parameterizes inverse images

Let $T : V \to V'$ be a linear transformation and let $v \in V$. If $T(v) = v'$, then

$$T^{-1}(v') = \{v + u : u \in \mathrm{Ker}(T)\}.$$

Corollary 7: Trivial kernel equivalent to injectivity

Let $T : V \to V'$ be a linear transformation. Then T is injective if and only if $\mathrm{Ker}(T) = \{\mathbf{0}\}$.

Example. Define a map $T : \mathbb{R}^2 \to R$ by $T(x_1, x_2) = x_1 + x_2$. This is a linear transformation (verify!). Then $\mathrm{Ker}(T) = \{(x, -x) : x \in \mathbb{R}\}$, which corresponds geometrically to a line through the origin (with slope -1.

Now if $a \in \mathbb{R}$. One preimage of a is $(a, 0)$. Then $T^{-1}(a) = \{(a, 0) + (x, -x) : x \in \mathbb{R}\}$. That is, it is a line through $(a, 0)$ parallel to the above line.

3. Direct products and internal direct sums

We have discussed the construction of a direct product for groups. We extend this to vector spaces.

Definition: Direct product

Let W_0, W_1 be vector spaces over a field k. The *direct product* of W_0, W_1 is defined as

$$W_0 \times W_1 = \{(w_0, w_1) : w_0 \in W_0, w_1 \in W_1\},$$

with addition and scalar multiplication defined by

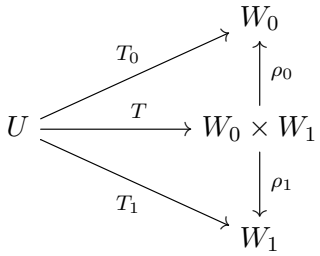
$$(w_0, w_1) + (u_0, u_1) = (w_0 + u_0, w_1 + u_1) \text{ for all } w_0, u_0 \in W_0, w_1, u_1 \in W_1$$
$$\lambda(w_0, w_1) = (\lambda w_0, \lambda w_1) \text{ for all } w_0 \in W_0, w_1 \in W_1, \lambda \in k.$$

By an earlier exercise, $W_0 \times W_1$ is an additive group. It is easy to verify that it is closed under scalar multiplication. Hence, $W_0 \times W_1$ is a vector space with identity $(\mathbf{0}, \mathbf{0})$.

Given a direct product $W_0 \times W_1$ of vector spaces, there are two maps, called *projection maps* by the rules:

$$\rho_0 : W_0 \times W_1 \to W_0 \qquad \rho_1 : W_0 \times W_1 \to W_1$$
$$(w_0, w_1) \mapsto w_0 \qquad (w_0, w_1) \mapsto w_1$$

The direct product satisfies a certain *universal property*. Suppose U is any vector space (over k) such that there exist linear transformations $T_0 : U \to W_0$ and $T_1 : U \to W_1$. Then there exists a unique linear transformation $T : U \to W_0 \times W_1$ such that the following diagram commutes:



In fact, it's not hard to define T. Just set $T(u) = (T_0(u), T_1(u))$ for all $u \in U$. It is easy to show that this is a linear transformation such that

$$\rho_0 \circ T = T_0 \quad \text{and} \quad \rho_1 \circ T = T_1.$$

It is left as an exercise to show that T is unique.

A general idea in mathematics is that to study complex objects we find ways to decompose them into simpler objects. The notion of an internal direct sum is one such idea.

Definition: Internal direct sum

Let W_0, W_1 be subspaces of a vector space V. We say that V is the *internal direct sum* of W_0 and W_1 if

(1) $V = W_0 + W_1$, and
(2) $W_0 \cap W_1 = \{\mathbf{0}\}$.

In this case, we write $V = W_0 \oplus W_1$.

Example. (1) Consider the following subspaces of $V = \mathbb{R}^2$:

$$W_0 = \{(x, 0) : x \in \mathbb{R}\} \quad \text{and} \quad W_1 = \{(0, x) : x \in \mathbb{R}\}.$$

Clearly any vector $(a, b) \in V$ can be written as $(a, 0) + (0, b)$ with $(a, 0) \in W_0$ and $(0, b) \in W_1$, so $V = W_0 + W_1$, Moreover, $W_0 \cap W_0 = \{(0, 0)\}$ and so $V = W_0 \oplus W_1$.

(2) Consider the following subspaces of $V = \mathbb{Q}[x]$:

$$W_0 = \text{span}\{1, x^2, x^4, \ldots\} \quad \text{and} \quad W_1 = \text{span}\{x, x^3, x^5, \ldots\}.$$

These are called, respectively, the subspaces of even and odd polynomials. Any polynomial can be decomposed into its even and odd parts, so $V = W_0 + W_1$. Furthermore, $W_0 \cap W_0 = 0$ (the zero polynomial) and so $V = W_0 \oplus W_1$.

Proposition 8: Equivalent conditions to IDP

Let V be a vector space. The following are equivalent:

(1) V is the internal direct sum of subspaces W_0 and W_1.
(2) Every element $v \in V$ can be written uniquely in the form $v = w_0 + w_1$ with $w_0 \in W_0$ and $w_1 \in W_1$, and
(3) $V = W_0 + W_1$ and whenever $w_0 + w_1 = \mathbf{0}$ for some $w_0 \in W_0$ and $w_1 \in W_1$, $w_0 = w_1 = \mathbf{0}$.

Proof. (1 $\Rightarrow$ 2) Let $v \in V$. Since V is the IDP of W_0 and W_1, then $v = w_0 + w_1$ for some $w_0 \in W_0$ and $w_1 \in W_1$. Now suppose $w_0 + w_1 = u_0 + u_1$ for some $u_0 \in W_0$ and $u_1 \in W_1$. Then $w_0 - u_0 = u_1 - w_1 \in W_0 \cap W_1 = \{\mathbf{0}\}$, so $w_0 = u_0$ and $w_1 = u_1$.

(2 $\Rightarrow$ 3) Since every element in V can be written in the form $v = w_0 + w_1$, then $V = W_0 + W_1$. Suppose $w_0 + w_1 = \mathbf{0}$ for some $w_0 \in W_0$ and $w_1 \in W_1$. Since $\mathbf{0} \in W_0$ and $\mathbf{0} \in W_1$, and $\mathbf{0} + \mathbf{0} = \mathbf{0}$, then we have $w_0 = w_1 = \mathbf{0}$.

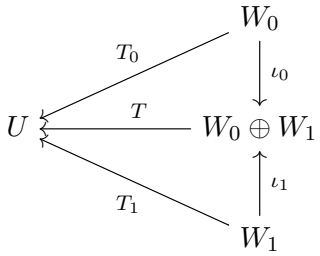
(3 $\Rightarrow$ 1) We are given $V = W_0 + W_1$, so we need only show that $W_0 \cap W_1 = \{\mathbf{0}\}$. Let $v \in W_0 \cap W_1$. Then $-v \in W_1$ and $v + (-v) = \mathbf{0}$. By the given condition, this implies $v = \mathbf{0}$. $\square$

The universal property satisfied by an IDP is *dual* to that of a direct product. Given an IDP $V = W_0 \oplus W_1$, there are two maps, called *inclusion maps*:

$$\begin{aligned} \iota_0 : W_0 &\to V & \qquad \iota_1 : W_1 &\to V \\ w_0 &\mapsto w_0 & \qquad w_1 &\mapsto w_1 \end{aligned}$$

These maps are obviously linear transformations.

Suppose U is any vector space (over k) such that there exist linear transformations $T_0 : W_0 \to U$ and $T_1 : W_1 \to U$. Then there exists a unique linear transformation $T : W_0 \oplus W_1 \to U$ such that the following diagram commutes:



Again, defining T is straightforward. Every element of $W_0 \oplus W_1$ can be written uniquely in the form $w_0 + w_1$ with $w_0 \in W_0$ and $w_1 \in W_1$. Thus, we set $T(w_0 + w_1) = T_0(w_0) + T_1(w_1)$. It is now easy to verify that

$$T \circ \iota_0 = T_0 \quad \text{and} \quad T \circ \iota_1 = T_1.$$

It is left as an exercise to show that T is unique.

We can extend the notion of an IDP to several subpaces. Let $W_0, W_1, \ldots, W_m$ be a collection of subspaces of a vector space V. Then V is an internal direct sum of the W_i if

(1) $V = W_0 + \cdots + W_m$, and
(2) for all $i = 0, \ldots, m$, $W_i \cap \left(\sum_{j \neq i} W_j\right) = \{\mathbf{0}\}$.

In this case we write $W = W_0 \oplus \cdots \oplus W_m$.

A digression on induction

Induction is a method for proving statements of the form: $\forall n \in \mathbb{N}, P(n)$. For example, consider the sum of the first n integers. We write out the first few cases below:

$$\begin{aligned} 1 &= 1 & 1+2+3+4 &= 10 \\ 1+2 &= 3 & 1+2+3+4+5 &= 15 \\ 1+2+3 &= 6 & 1+2+3+4+5+6 &= 21. \end{aligned}$$

For each computation, we can use the previous computation and just add the next number. In fact, the sum of the first n positive integers is $\frac{n(n+1)}{2}$. The question is how to prove this.

First Principle of Mathematical Induction

Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If for all integers with $k \geq n_0$, $S(k)$ implies $S(k+1)$ is true, then $S(n)$ is true for all integers greater than or equal to n_0.

There are two steps to any proof by induction:

(1) Prove the base case $S(n_0)$. This is often $n_0 = 0$ or 1, depending on the problem. Generally it is the least value that you want to prove the statement for.
(2) Assume the statement is true for some $n \geq n_0$ (assume $S(n)$) and prove the statement for $n+1$ (prove $S(n+1)$).

In step (2), the assumption that $S(n)$ is true is called the *inductive hypothesis*.

Theorem 9: Sum of the first n positive integers

The sum of the first n positive integers is $\frac{n(n+1)}{2}$.

Proof. The base case is $n = 1$. Since $1(1+1)/2 = 1$, then the statement is true for $n = 1$. Assume the statement is true for some $n \geq 1$. That is, $1+2+\cdots+n = n(n+1)/2$. Then at $n+1$ we have

$$\begin{aligned} 1+2+\cdots+n+(n+1) &= (1+2+\cdots+n)+(n+1) \\ &= \frac{n(n+1)}{2}+(n+1) \quad \text{by the inductive hypothesis} \\ &= \frac{1}{2}(n(n+1)+2(n+1)) \\ &= \frac{1}{2}(n+1)(n+2), \end{aligned}$$

which is the statement for $n+1$. Thus, the result is true by induction. □

For positive integers n and k, with $k \leq n$. The *binomial coefficient* is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This is so-called because these are the coefficients appearing in the *binomial theorem*:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Note that,

$$\begin{aligned}\binom{n}{k} + \binom{n}{k-1} &= \frac{n!(n+1-k)}{k!(n+1-k)!} + \frac{n!k}{k!(n+1-k)!} = \frac{n!(n+1-k)+n!k}{k!(n+1-k)!} \\ &= \frac{n!(n+1)}{k!(n+1-k)!} = \frac{n!}{k!(n+1-k)!} = \binom{n+1}{k}.\end{aligned}$$

Theorem 10: The Binomial Theorem

Let n be a positive integer and let $a, b \in \mathbb{R}$. Then

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Proof. When $n = 1$, we have $a + b = a^1 b^0 + a^0 b^1$, so the statement is true. Now assume the statement is true for some $n \geq 1$. Then

$$\begin{aligned}(a+b)^{n+1} &= (a+b)(a+b)^n \\ &= (a+b)\left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\right) \quad \text{by the inductive hypothesis} \\ &= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=1}^{n} \binom{n}{k} a^k b^{n+1-k} + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k}\right] a^k b^{n+1-k} + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^k b^{n+1-k} + b^{n+1} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}.\end{aligned}$$

□

The next version of induction is subtly different but is very useful in certain circumstances. It is called *strong induction*.

Second Principle of Mathematical Induction

Let $S(n)$ be a statement about integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If $S(n_0), S(n_0+1), \ldots, S(k)$ imply that $S(k+1)$ for $k \geq n_0$, then the statement is true for all integers $n \geq n_0$.

A positive integer $p > 1$ is *prime* if the only positive factors of p are 1 and p.

Lemma 11: Primes dividing a product

Let $a, b \in \mathbb{Z}$ and p a prime number. If $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof. Suppose $p \nmid a$. Since $\gcd(a, p) = 1$, then by Bézout's Identity, there exist integers r, s such that $ar + ps = 1$. Thus, $b = b(1) = b(ar + ps) = (ab)r + p(bs)$. Since $p \mid ab$ and $p \mid p$, then $p \mid b$. □

Strong induction is used in the uniqueness part of the next result.

Theorem 12: Fundamental Theorem of Arithmetic

Let n be an integer such that $n > 1$. Then $n = p_1 p_2 \cdots p_k$, where the p_i are prime. Moreover, this factorization is unique up to rearrangement.

Proof. (Existence) Let S be the set of all integers that cannot be written as the product of primes and suppose $S \neq \emptyset$. By the Well-Ordering Principle, S has a least element, say a. Then a is not prime and so $a = a_1 a_2$ where $1 < a_1 < a$ and $1 < a_2 < a$. By choice of a, a_1 and a_2 are not in S, so a_1 and a_2 can be written as products of primes. But then so can a, a contradiction.

(Uniqueness) The theorem is true for $n = 2$ because 2 is prime. Now assume the result holds for all integers m such that $1 < m < n$. Write

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_\ell$$

where the p_i, q_i are prime, and $p_1 \leq p_2 \leq \cdots \leq p_k$, $q_1 \leq q_2 \leq \cdots q_\ell$. By the Lemma, $p_1 \mid q_i$ for some i and $q_1 \mid p_j$ for some J. The p_i and q_i are prime, so $p_1 = q_i$ and $q_1 = p_j$. But then $p_1 \leq p_j = q_1 \leq q_i = p_1$, so $p_1 = q_1$. By the inductive hypothesis,

$$n' = p_1 \cdots p_k = q_2 \cdots q_\ell$$

has a unique factorization. Thus, $k = \ell$ and $q_i = p_i$ for $i = 1, \ldots, k$. □

Exercise. Prove the following statements:

(1) Prove that the sum of the first n cubes is $\frac{n^2(n+1)^2}{4}$.
(2) Prove that $10^{n+1} + 10^n + 1$ is divisible by 3 for $n \in \mathbb{N}$.
(3) Let X be a set with n elements. Show that X has exactly 2^n subsets. (The set of all subsets of X is called the *power set* of X, denoted $\mathcal{P}(X)$.)

Dimension

Vector spaces are often infinite, so how do we measure their dimension? The answer comes from looking at the smallest possible set such that every element is a linear combination of that set. Such a set is called a basis and the size of this basis will determine the dimension.

1. Bases and Dimension

Definition: Linearly independent

Let V be a vector space over a field k. Then a family of vectors $v_1, \ldots, v_n$ is called *linearly independent* if

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = \mathbf{0}$$

implies $\lambda_1 = \cdots = \lambda_n = 0$ ($\lambda_j \in k$). Otherwise, we say the set is *linearly dependent*.

Turning this definition around, we say the set is linearly dependent if there is some set of scalars $\lambda_1, \ldots, \lambda_n$ not all zero such that $\lambda_1 v_1 + \cdots + \lambda_n v_n = \mathbf{0}$.

Example. (1) Let S be any set in a vector space containing the zero vector $\mathbf{0}$. Then S is linearly dependent since we have the relation $1\mathbf{0} = \mathbf{0}$. Thus, a linearly independent set may not contain the zero vector.

(2) Consider the set $\{(1,0,0),(1,1,0),(1,1,1)\}$ in $\mathbb{R}^3$. Assume we have a linear combination

$$\lambda_1(1,0,0) + \lambda_2(1,1,0) + \lambda_3(1,1,1) = (0,0,0)$$

That is, $(\lambda_1 + \lambda_2 + \lambda_3, \lambda_2 + \lambda_3, \lambda_3) = 0$. Hence, from the last coordinate we see that $\lambda_3 = 0$. From the second coordinate now, $\lambda_2 = 0$. And from the first coordinate, $\lambda_1 = 0$. Thus, the set is linearly independent.

(3) Now consider the set $\{(1,0,0),(1,1,0),(2,1,0)\}$ in $\mathbb{R}^3$. This set is linearly dependent since

$$1 \cdot (1,0,0) + 1 \cdot (1,1,0) + (-1) \cdot (2,1,0) = (0,0,0).$$

This is called a *linear dependence relation* (and this set is linearly dependent).

(4) The canonical basis vectors $e_1, \ldots, e_n$ in k^n are linearly independent.

(5) Let v and w be nonzero vectors in a vector space V. If $\{v, w\}$ is linearly dependent, then $\lambda_1 v + \lambda_2 w = \mathbf{0}$ for some scalars λ_1, λ_2. Note this implies that both $\lambda_1, \lambda_2 \neq 0$. Thus, $v = -\frac{\lambda_2}{\lambda_1} w$. That is, v is a scalar multiple of w. In the case of $\mathbb{R}^2$, this means that v and w lie on the same line through the origin.

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapter 4. Last Updated: November 5, 2021

Proposition 1: Characterization of Linear Dependence

A collection of vectors $v_1, \ldots, v_n$ is linearly dependent if and only if one of the v_j can be expressed as a linear combination of the others.

Proof. ($\Rightarrow$) Suppose $v_1, \ldots, v_n$ is linearly dependent. Then there are scalars $\lambda_1, \ldots, \lambda_n$ not all zero such that

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = \mathbf{0}.$$

Suppose $\lambda_1 \neq 0$. Then,

$$v_1 = -\frac{1}{\lambda_1}\left(\lambda_2 v_2 + \cdots + \cdots + \lambda_n v_n\right).$$

This argument works for any v_j with $\lambda_j \neq 0$ by reordering.

($\Leftarrow$) Assume v_1 can be written as a linear combination of $v_2, \ldots, v_n$. Then there are scalars $\mu_2, \ldots, \mu_n$ such that

$$v_1 = \mu_2 v_2 + \cdots + \mu_n v_n.$$

Hence,

$$(-1)v_1 + \mu_2 v_2 + \cdots + \mu_n v_n = \mathbf{0}.$$

Hence, $v_1, \ldots, v_n$ is linearly dependent. □

The next lemma expresses a similar idea as above, but is useful in one of our main results.

Lemma 2: Removing linearly dependent vectors

Let $v_1, \ldots, v_n$ be a collection of vectors in V. If one of the v_i is a linear combination of the other vectors, then $\operatorname{span}(v_1, \ldots, v_n) = \operatorname{span}(v_1, \ldots, \widehat{v_i}, \ldots, v_n)$.

Proof. Any linear combination involving v_i can be replaced by a linear combination involving only the other vectors. □

Combining the last two results means that any span of a linearly dependent set of vectors can be written involving a span of fewer vectors.

Definition: Basis

A *basis* of a vector space V is a linearly independent collection of vectors that spans V.

Example. The set $e_1, \ldots, e_n$ is a basis of k^n, hence the name *canonical basis vectors*.

There is no requirement that a basis be finite, and a vector space may (and often does!) have many bases. However, we will show that if one basis of a vector space V is finite, then every other basis has the same number of elements. This number is known as the *dimension* of the vector space. Our next several results are designed to prove this fact.

Proposition 3: Basis is a coordinate system

A subset B of a vector space V is a basis if and only if every vector in V can be written uniquely as a linear combination of vectors in B.

Proof. ($\Rightarrow$) Suppose B is a basis and let $v \in V$. Since B spans V, then v can be written as a linear combination of vectors in B. Suppose there are two ways:

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = v = \mu_1 v_1 + \cdots + \mu_n v_n,$$

with $v_i \in B$. This implies that

$$(\lambda_1 - \mu_1)v_1 + \cdots + (\lambda_n - \mu_n)v_n = \mathbf{0}.$$

By linear independence, we have $\lambda_i = \mu_i$ for all i.

($\Leftarrow$) Suppose every vector in V can be written uniquely as a linear combination of vectors in B. Clearly, B spans V. Suppose $\lambda_1 v_1 + \cdots + \lambda_n v_n = \mathbf{0}$ is a linear combination of vectors in B. One such linear combination is obtained by taking $\lambda_i = 0$ for all i. By uniqueness, this is the only such linear combination. Thus, B is linearly independent and hence a basis. □

Let V be a vector space over a field k and let B be a basis with n elements. Define the *coordinate map* $\gamma_B : V \to k^n$ as follows: write $v \in V$ uniquely in terms of the basis vectors in B, so $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, then $\gamma_B(v) = (\lambda_1, \ldots, \lambda_n) \in k^n$.

Example. (1) The canonical basis of k^n is a basis and the coordinate map is just the identity.

(2) One basis for $\mathbb{R}^2$ is $B = \{v_1, v_2\}$ with $v_1 = (1, 0)$ and $v_2 = (1, 1)$. Since $(2, 1) = 1 \cdot (1, 0) + 1 \cdot (1, 1)$, so $\gamma_B(2, 1) = (1, 1)$.

(3) Let W be the subspace of $R[x]$ consisting of polynomials of degree at most n. Then a basis for W is $B = \{1, x, \ldots, x^n\}$. The coordinate map $\gamma_B : W \to k^{n+1}$ is given by

$$\gamma_B(a_0 + a_1 x + \cdots + a_n x^n) = (a_0, a_1, \ldots, a_n).$$

Let V be a vector space. A subset S of V is a maximally linearly independent set if S is not properly contained in another (larger) linearly independent set. Similarly, a subset S is a minimal generating set if S spans V and S does not properly contain a (smaller) set that spans V.

Proposition 4: Equivalent characterizations of basis

Let S be a subset of the vector space V. Then the following are equivalent:

(1) S is a maximally linearly independent set,
(2) S is a minimal generating set,
(3) S is a basis.

Proof. (1) $\Rightarrow$ (3) Suppose S is a maximally linearly independent set. We must show S spans V. That is, S does not span V. Then there is some vector $v \in V$ that is not a linear combination of the vectors in S. But then $S \cup \{v\}$ is linearly independent.

(3) $\Rightarrow$ (1) Suppose S is a basis and let $v \in V \backslash S$. Then v can be written as a linear combination of the elements of S. Thus, $S \cup \{v\}$ is linearly dependent.

(2) $\Rightarrow$ (3) Suppose S is a minimal generating set. If S is not linearly independent, then some vector $v \in S$ may be written as a linear combination of the other vectors in S. But then $S \backslash \{v\}$ has the same span as S, a contradiction.

(3) $\Rightarrow$ (2) Suppose S is a basis, so S is a generating set of V. If T is a proper subset of S that spans V, then there is some vector $v \in S \backslash T$ that can be written as a linear combination of the vectors in T. But this means that S is linearly dependent. A contradiction. □

This brings us to our first main result.

Theorem 5: The Exchange Theorem

Suppose that the collection $v_1, \ldots, v_n$ spans V and $w_1, \ldots, w_r$ is a linearly independent set. Then $r \leq n$.

Proof. Note by definition that the w_i are nonzero. Since the v_i span V, then

$$w_1 = \lambda_1 v_1 + \cdots + \lambda_n v_n.$$

Suppose, WLOG, that $\lambda_1 \neq 0$. Then v_1 is a linear combination of $v_2, \ldots, v_n, w_1$. Hence, $v_2, \ldots, v_n, w_1$ span V. Repeating, we can write

$$w_2 = \mu_1 w_1 + \mu_2 v_2 + \cdots + \mu_n v_n.$$

Since w_1, w_2 are linearly independent, then one of $\mu_2, \ldots, \mu_n$ is nonzero. Again, WLOG, assume $\mu_2 \neq 0$, so v_2 is a linear combination of $v_3, \ldots, v_n, w_1, w_2$. Thus, $v_3, \ldots, v_n, w_1, w_2$ spans V. We continue in this way, deleting a v_i and replacing the corresponding v_i with w_i without changing the span. For this to be possible, we require that $r \leq n$. □

Corollary 6: Existence of dimension

Suppose that B and B' are both bases for the finitely-generated vector space V. Then B and B' have the same number of elements.

Proof. Since V is finitely generated, then there is some finite set that spans V. Hence, by the Exchange Theorem, any linearly independent set (and hence any basis) must have finitely many elements. Let n be the number of elements in B and n' the number of elements in B'. Since B and B' both span V, then again by the Exchange Theorem, $n' \leq n$ and $n \leq n'$, so $n = n'$. □

Definition: Dimension

The number of elements in a basis for a finitely-generated vector space V is called the *dimension* of V, denoted $\dim(V)$. The zero vector space is assigned dimension 0.

Example. (1) The vector space k^n has dimension n since the canonical basis has n elements.

(2) The vector space $\mathbb{Q}[x]$ is infinite dimensional since any spanning set will have infinitely many elements. The subspace of polynomials of degree at most n has dimension $n+1$.

(3) The solution set in $\mathbb{R}^2$ to the linear equation $x_1+x_2=0$ is spanned by a single vector, $(-1,-1)$. Hence, its dimension is 1.

2. Vector spaces are free!

We have not yet proved that every vector space has a basis.

Proposition 7: Bases live between linearly independent and spanning sets

Let $S \subseteq S'$ be finite subsets of V such that S is linearly independent and S' spans V. Then there exists a basis B such that $S \subseteq B \subseteq S'$.

Proof. Suppose $S = \{v_1, \ldots, v_n\}$ and S' contains the additional vectors $w_1, \ldots, w_m$. If S' is linearly independent we are done. If not, one of the w_j depends on the other w_i (because the v_i are linearly independent). Thus, we may delete w_j without changing the span. Continuing in this way, we remove superfluous w_i until we arrive at a linearly independent set which is therefore a basis. □

The next statement we only prove for finite-dimensional vector spaces, but is in fact true in general.

Theorem 8: The Fundamental Theorem of Linear Algebra

The following hold in any vector space:

(1) Every linearly independent set may be extended to a basis.
(2) Every spanning set may be contracted to a basis.
(3) Every vector space has a basis. (Vector spaces are free!)

Proof. For (1), let S be a linearly independent set. Since $S \subset V$ and V spans itself, then we apply the previous proposition to obtain the result. For (2), let S' be a spanning set and consider $\{v\}$ for any nonzero $v \in V$. Then $\{v\} \subset S'$ and so again the previous proposition applies. Now (3) is clear from (1) and (2). □

The next result will either be very surprising, or very unsurprising depending on your perspective.

Corollary 9: Classification of finite-dimensional vector spaces

Every finite-dimensional vector space is isomorphic to k^n for some n.

Proof. Let V be a finite-dimensional vector space, say $\dim(V) = n$. Then V has a basis B by the fundamental theorem. Now we need only recall that the coordinate map $\gamma_B : V \to k^n$ is an isomorphism. □

Proposition 10: Size of linearly independent and spanning sets

Let V have dimension n. Then

(1) no subset of V of more than n vectors can be linearly independent, and
(2) no subset of V of fewer than n vectors can span V.

Proof. The first statement is just the Exchange Theorem. The second follows from the previous theorem that a spanning set of fewer than n elements can be contracted to a basis of fewer than n elements, a contradiction. □

The next result is essentially the pigeonhole principle. If we know the dimension of the space, it is enough to check linear independence or spanning to confirm that a set of n elements is a basis.

Theorem 11: Pigeonhole for vector spaces

Let V have dimension n and let S be a collection of n vectors in V. Then the following are equivalent:

(1) S is linearly independent
(2) S spans V
(3) S is a basis.

Proof. Of course, (3) $\Rightarrow$ (1) and (3) $\Rightarrow$ (2) are trivial. For (1) $\Rightarrow$ (3) we need only note that, since S is linearly independent, it can be extended to a basis of V. But if S were not already a basis, then S would extend to a basis with more than n elements, a contradiction. Similarly, for (2) $\Rightarrow$ (3), since S spans V, it can be contracted to a basis. But then the basis would have fewer than n elements, a contradiction. □

Suppose V has dimension n and W is a subspace of dimension n. Then W has a basis B of n elements. But then B is a linearly independent set of n elements in V, and hence a basis of V by the previous theorem. Hence, $W = V$. The next result generalizes this idea.

Proposition 12: Subspaces of finite-dimensional vector spaces

Every subspace of a finite-dimensional vector space is finite dimensional of smaller or equal dimension.

Proof. Let W be a subspace of V, with $\dim(V) = n < \infty$. Starting with the null set, we can extend one vector at a time to obtain a maximally linearly independent set in W. But such a set is also linearly independent in V, and hence has fewer than n elements. Hence, W has a basis of no more than n elements. □

Example. (1) Any set of three vectors in $\mathbb{R}^2$ is linearly dependent. More generally, any set of $n+1$ elements in k^n is linearly dependent.

(2) The vector space $\mathscr{C}^0(\mathbb{R})$ is infinite-dimensional. To see this, we need only observe that $\mathbb{R}[x]$ (or even $\mathbb{Q}[x]$) is an infinite-dimensional subspace. Note this is not true over finite fields $\mathbb{F}_p$.

(3) The set $\{(1,0,0),(1,1,0),(2,0,1)\}$ is linearly independent and hence a basis for $\mathbb{R}^3$.

Example. Consider the following linear system in the variables x_1, x_2, x_3 with coefficients and constants in k:

$$a_1x_1 + b_1x_2 + c_1x_3 = y_1$$
$$a_2x_1 + b_2x_2 + c_2x_3 = y_1$$
$$a_3x_1 + b_3x_2 + c_3x_3 = y_3$$

Then the following are equivalent:

(1) The system has at least one solution for all y_1, y_2, y_3.
(2) The corresponding homogeneous system ($y_i = 0$ for all i) has only the trivial solution,
(3) The system has exactly one solution for all y_1, y_2, y_3.

The first statement is simply that the vectors $a = (a_1, a_2, a_3)$, $b = (b_1, b_2, b_3)$, and $c = (c_1, c_2, c_3)$ span k^3. Statement (2) says that the vectors a, b, c are linearly independent and statement (3) says that they are a basis. But for any set of vectors, these three statements are equivalent in k^3.

3. Rank and Nullity

For a linear transformation, the rank and nullity are connected through dimension.

Definition: Rank, nullity

Let $T : V \to W$ be a linear transformations between finite-dimensional vector spaces. The number $\dim(\mathrm{Im}(T))$ is called the *rank* of T, and $\dim(\mathrm{Ker}(T))$ is called the *nullity*.

Theorem 13: The Rank-Nullity Theorem

Let $T : V \to W$ be a linear transformations between finite-dimensional vector spaces. Then

$$\dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)) = \dim(V).$$

Proof. Let $v_1, \ldots, v_n$ be a basis for $\mathrm{Ker}(T)$, which exists because $\mathrm{Ker}(T)$ is a subspace of V. (This also tells us that $n \leq \dim(V)$.) Extend this to a basis of V: $v_1, \ldots, v_n, w_1, \ldots, w_m$. Hence, $\dim(V) = n + m$. We claim $\dim(\mathrm{Im}(T)) = m$.

Since T is surjective onto its image, then T maps the spanning set $v_1, \ldots, v_n, w_1, \ldots, w_m$ of V to a spanning set $T(v_1), \ldots, T(v_n), T(w_1), \ldots, T(w_m)$ of $\mathrm{Im}(T)$. But $T(v_i) = \mathbf{0}$, so in fact $T(w_1), \ldots, T(w_m)$ is a spanning set of $\mathrm{Im}(T)$. We claim it is also linearly independent. Suppose that

$$\lambda_1 T(w_1) + \cdots + \lambda_m T(w_m) = \mathbf{0}.$$

By linearity,

$$T(\lambda_1 w_1 + \cdots + \lambda_m w_m) = \mathbf{0}.$$

But then $\lambda_1 w_1 + \cdots + \lambda_m w_m \in \mathrm{Ker}(T)$. Since $\mathrm{Ker}(T)$ is spanned by the v_i, then

$$\lambda_1 w_1 + \cdots + \lambda_m w_m = \mu_1 v_1 + \cdots + \mu_n v_n$$

for some scalars μ_i. Then we have the linear combination in V:

$$-\mu_1 v_1 - \cdots - \mu_n v_n + \lambda_1 w_1 + \cdots + \lambda_m w_m = \mathbf{0}.$$

Since $v_1, \ldots, v_n, w_1, \ldots, w_m$ is a basis of V, linear independence forces all coefficients to be zero. In particular, $\lambda_i = 0$ for all i, so the w_j are linearly independent and hence a basis for $\mathrm{Im}(T)$. It follows that $\dim(\mathrm{Im}(T)) = m$. □

Corollary 14: Pigeonhole for linear transformations

Let $T : V \to W$ be a linear transformations between finite-dimensional vector spaces of the same dimension. Then the following are equivalent:

(1) T is injective.
(2) T is surjective.
(3) T is an isomorphism.

Proof. The statements (3) $\Rightarrow$ (1) and (3) $\Rightarrow$ (2) are clear. Set $n = \dim(V) = \dim(W)$. To prove (1) $\Rightarrow$ (3), suppose T is injective. Then $\mathrm{Ker}(T) = \{\mathbf{0}\}$ and so by Rank-Nullity, $\dim(\mathrm{Ker}(T)) = 0$. Thus, $\dim(\mathrm{Im}(T)) = n = \dim(W)$. Since $\mathrm{Im}(T)$ is a subspace of W, this implies $\mathrm{Im}(T) = W$. That is, T is surjective and thus an isomorphism.

Similarly, to prove (2) $\Rightarrow$ (3), we assume T is surjective. That is, $\mathrm{Im}(T) = W$, so $\dim(\mathrm{Im}(T)) = \dim(W) = n$. By Rank-Nullity, $\dim(\mathrm{Ker}(T)) = 0$, so $\mathrm{Ker}(T) = \{\mathbf{0}\}$ and T is injective. Thus, T is an isomorphism. □

Corollary 15: Comparing dimensions

Let $T : V \to W$ be a linear transformations between finite-dimensional vector spaces. If T is injective, then $\dim(V) \leq \dim(W)$. If T is surjective, then $\dim(W) \leq \dim(V)$.

Proof. This is similar to the above. If T is injective, then $\dim(V) = \dim(\mathrm{Im}(T)) \leq \dim(W)$. If T is surjective, then

$$\dim(V) = \dim(\mathrm{Im}(T)) + \dim(\mathrm{Ker}(T)) = \dim(W) + \dim(\mathrm{Ker}(T)),$$

so $\dim(W) \leq \dim(V)$. □

Matrices

In this chapter we study one of our primary computational tools for solving systems of equations

1. Notation and Terminology

Let k be a field. An $m \times n$ matrix over a field k is an array of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

where each a_{ij} is an element of k. We write $A = (a_{ij})$ tto indiate that A is a matrix whose (i,j)-entry is denoted a_{ij}. The set of all such matrices is $\mathrm{Mat}_{m\times n}(k)$. A *square matrix* is one in which $m = n$ and we denote these by $M_n(k)$.

We denote the columns of A by $A^1, \ldots, A^n$, which are vectors in k^m. Similarly, we denote the rows of A by $A_1, \ldots, A_m$, which are vectors in k^n. Hence,

$$A = (A^1, \ldots, A^n).$$

Next we introduce operations on matrices that are reminiscent of operations on vectors. In fact, we will show that $\mathrm{Mat}_{m\times n}(k)$ forms a vector space over k.

Definition: Matrix sum, scalar product of a matrix

Let $A = (a_{ij})$ and $B = (b_{ij})$ lie in $\mathrm{Mat}_{m\times n}(k)$.

(1) Their *sum* $A + B$ is the $m \times n$ matrix whose (i,j) entry is $a_{ij} + b_{ij}$.

(2) For any $\lambda \in k$, the *scalar product* λA is the $m \times n$ matrix whose (i,j)-entry is λa_{ij}.

We may summarize these definitions by writing $A + B = (a_{ij} + b_{ij})$ and $\lambda A = (\lambda a_{ij})$.

Example. Let $A = \begin{pmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{pmatrix}$. Then

$$A + 3B = \begin{pmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{pmatrix} + 3\begin{pmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{pmatrix} + \begin{pmatrix} 3 & 15 \\ -6 & 9 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 17 \\ -5 & 14 \\ 2 & 3 \end{pmatrix}.$$

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapter 5. Last Updated: November 5, 2021

Proposition 1: $\mathrm{Mat}_{m\times n}(k)$ is a vector space

For all m and n, $\mathrm{Mat}_{m\times n}(k)$ is a vector space over k of dimension mn. In particular, the matrices of a given size form an additive group with respect to matrix addition.

Proof. It is clear that we have closure under addition. Let $A, B, C \in \mathrm{Mat}_{m\times n}(k)$ and write $A = (a_{ij})$, $B = (b_{ij})$, and $C = (c_{ij})$. Then

$$(A + B) + C = ((a_{ij} + b_{ij}) + c_{ij}) = a_{ij} + (b_{ij} + c_{ij}) = A + (B + C),$$

so matrix addition is associative. The $m \times n$ zero matrix O is the additive identity, since

$$A + O = (a_{ij} + 0) = (a_{ij}) = A.$$

Moreover, the additive identity of A is the scalar product $(-1)A$ since

$$A + (-1)A = (a_{ij}) + (-a_{ij}) = (a_{ij} - a_{ij}) = (0) = O.$$

Finally, we have

$$A + B = (a_{ij} + b_{ij}) = (b_{ij} + a_{ij}) = A + B,$$

so matrix addition is commutative. Thus, $\mathrm{Mat}_{m\times n}(k)$ is an additive abelian group. The remaining properties for a vector space are similarly checked and left as an exercise.

Let E_{ij} denote the matrix whose (i, j)-entry is 1 and the remaining entries are 0. It is easy to see that the set $\{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis for $\mathrm{Mat}_{m\times n}(k)$. Hence, $\dim(\mathrm{Mat}_{m\times n}(k)) = mn$. □

Definition: Matrix product

Let $A \in \mathrm{Mat}_{m\times n}(k)$ and $B \in \mathrm{Mat}_{n\times p}(k)$. Then the *product* AB is the $m \times p$ matrix $C = (c_{ij})$ defined by

$$c_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

This is known as the *row-column rule* for matrix multiplication. (For those "in the know", the entry c_{ij} is the *dot product* of the ith row of A with the jth column of B.) Note the definition requires that the number of columns of A match the number of rows of B. This should be reminiscent of the rule that when taking the composition of functions $g \circ f$, the codomain of f must live inside the domain of g.

Example. Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{pmatrix}$. Then

$$AB = \begin{pmatrix} 2(1) + 1(-3) & 2(2) + 1(1) & 2(4) + 1(6) \\ -1(1) + 5(-3) & -1(2) + 5(1) & -1(4) + 5(6) \end{pmatrix} = \begin{pmatrix} -1 & 5 & 14 \\ -16 & 3 & 26 \end{pmatrix}.$$

Note that the product BA is undefined. Even when both are defined, they are in general not equal. That is, matrix multiplication is *noncommutative*.

A *diagonal matrix* is a square matrix whose entries not on the diagonal (the (i,i)-entries) are zero. A special case of this is the $n \times n$ *identity matrix*, denoted I_n, which is defined as the matrix whose (i,j) entry is δ_{ij} where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

(This function is known as the *Kronecker delta* function.) Hence,

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots 0 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Note that for any $n \times n$ matrix, $AI_n = A = I_n A$.

Proposition 2: Algebraic properties of matrix multiplication

Suppose A, B, and C are matrices of appropriate sizes so that the given operations are defined, then

(1) $A(BC) = (AB)C$
(2) $A(B + C) = AB + AC$
(3) $(A + B)C = AC + BC)$
(4) $\lambda(AB) = (\lambda A)B = A(\lambda B)$ for all $\lambda \in k$.

In particular, $M_n(k)$ is a ring with unity.

Proof. We prove (2). Property (3) is similar and property (4) is left as an exercise. We will prove property (1) later in an indirect way.

Let A be $m \times n$, while both B and C are $n \times p$. Then the (i,j)-entry of $A(B + C)$ is

$$\sum_{\ell=1}^{n} a_{i\ell}(b_{\ell n} + c_{\ell n}) = \sum_{\ell=1}^{n} (a_{i\ell} b_{\ell n} + a_{i\ell} c_{\ell n}) = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell n} + \sum_{\ell=1}^{n} a_{i\ell} c_{\ell n},$$

which is the (i,j)-entry of $AB + AC$. □

We now introduce a new operation on matrices: the transpose. For now, this is simply defined algebraically, but it will play an important role later.

Definition: Matrix transpose, symmetric matrix

Let $A = (a_{ij})$ be an $m \times n$ matrix. The *transpose* of A, denoted ${}^tA$, is the $n \times m$ matrix whose (i, j)-entry is a_{ji}. The matrix A is *symmetric* if $A = {}^tA$.

A symmetric matrix is necessarily square. Note that the transpose is obtained by switching the rows and columns in A.

Example.

$$ {}^t\begin{pmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{pmatrix} = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 5 & -3 \end{pmatrix}. $$

Proposition 3: Algebraic properties of matrix transpose

Suppose A and B are matrices of appropriate sizes so that the given operations are defined, then

(1) ${}^t(A + B) = {}^tA + {}^tB$
(2) ${}^t(\lambda A) = \lambda({}^tA)$
(3) ${}^tAB = {}^tB\,{}^tA$

Properties (1) and (2) above are easy to verify. Property (3) is not difficult, but we will again see this in an indirect way later.

Definition: Invertible matrix, the general linear group

An element A of $M_n(k)$ is *invertible* (or *nonsingular* if there exists an element $B \in M_n(k)$ such that $AB = I_n = BA$. The set of all such invertible matrices is called the *general linear group of rank n matrices*, denoted $\mathrm{GL}_n(k)$.

Note that $GL_n(k)$ is just $k^\times = k\backslash\{0\}$.

Proposition 4: $\mathrm{GL}_n(k)$ is a group

For all $n > 1$, the group $\mathrm{GL}_n(k)$ is a (noncommutative) group under matrix multiplication.

Proof. Let $A, B \in \mathrm{GL}_n(k)$, so that there exist inverses A^{-1}, B^{-1}. Then

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = (AI_n)A^{-1} = AA^{-1} = A.$$

Similarly, $(B^{-1}A^{-1})(AB) = I_n$. Hence, AB is invertible, so $AB \in \mathrm{GL}_n(k)$. So, matrix multiplication is a binary operation on $\mathrm{GL}_n(k)$. Associativity follows from properties of matrix multiplication. The identity is I_n, and invertibility follows by definition. □

Note that uniqueness of inverses now follows from properties of groups.

2. Introduction to linear systems

Throughout this section, we regard k^n as column vectors ($n \times 1$ matrices). If A is an $m \times n$ matrix and $\mathbf{x} \in k^n$, then $A\mathbf{x} \in k^m$. Hence, matrix multiplication defines a map

$$\begin{aligned} T_A : k^n &\to k^m \\ \mathbf{x} &\mapsto A\mathbf{x} \end{aligned}$$

By properties of matrix multiplication, $A(\lambda\mathbf{x}) = \lambda A(\mathbf{x})$ and $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in k^n$ and $\lambda \in k$. That is, T_A is a linear transformation. (It turns out that *every* linear transformation is determined by a matrix.) The *rank* of the matrix A is the rank of T_A, that is, $\dim(\mathrm{Im}(T_A))$.

Let $\mathbf{e}_j$ be the jth canonical basis vector of k^n, then $A\mathbf{e}_j = A^j$, the jth column of A.

Observe that

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

Thus, given a linear system with m equations and n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_n \end{aligned}$$

we may express in matrix form as $A\mathbf{x} = \mathbf{y}$. We may also express it as the vector equation

$$(\star) \qquad A^1x_1 + \cdots + A^nx_n = \mathbf{y}.$$

When $\mathbf{y} = \mathbf{0}$ we call the linear system *homogeneous*.

The next few results tie the theory of linear systems to properties of linear transformations.

Proposition 5: Subspaces associated to $A\mathbf{x} = \mathbf{y}$

The set of all $\mathbf{y}$ for which there exists a solution to the linear system $A\mathbf{x} = \mathbf{y}$ of m equations and n unknowns is a subspace of k^m. The solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is a subspace of k^n.

Proof. The first subspace is simply the image of T_A and the second is the kernel. □

Proposition 6

Suppose that $n > m$. Then the homogeneous linear system of m equations in n unknowns represented by the matrix equation $A\mathbf{x} = \mathbf{0}$ always has a nontrivial solution. In fact, the solution set is a subspace of k^n of dimension at least $n - m$.

Proof. The image of T_A has dimension at most m. Then by Rank-Nullity, the kernel of T_A has dimension $n - \dim(\mathrm{Im}(T_A))$. □

The following result is stated slightly differently in the text.

Theorem 7: The Invertible Matrix Theorem (version 1)

Let $A \in M_n(k)$. The following are equivalent:

(1) The linear system $A\mathbf{x} = \mathbf{y}$ has at least one solution for all $\mathbf{y} \in k^n$.
(2) The columns of A span k^n.
(3) The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(4) The columns of A are linearly independent.
(5) The linear system $A\mathbf{x} = \mathbf{y}$ has exactly one solution for all $\mathbf{y} \in k^n$.
(6) The columns of A constitute a basis for k^n.

Proof. Statements (2), (4), and (6) are equivalent by Pigeonhole for finite-dimensional vector spaces. It is easy to see that (1) ⇔ (2) and (3) ⇔ (4) by considering the vector equation ($\star$). Moreover, (5) ⇔ (6) by uniqueness of coordinates for bases. □

If A is invertible, then we may solve $A\mathbf{x} = \mathbf{y}$ as $\mathbf{x} = A^{-1}\mathbf{y}$, so A invertible implies (1) above. Now suppose (1) holds. For each i, let $\mathbf{b}_i$ be a solution to the matrix equation $A\mathbf{x} = \mathbf{e}_i$. Let $B = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$. Then,

$$AB = A\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix} = \begin{pmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_n \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} = I_n.$$

This proves that B is a *right inverse* of A, but does not prove that it is also a *left inverse*. We will return to prove in full generality that these statements are equivalent to A invertible.

3. Solution techniques

We will write the linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_n \end{aligned}$$

as an *augmented matrix*

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & y_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & y_n \end{array}\right)$$

which we will abbreviate as $(A|\mathbf{y})$.

Elementary row operations

The following *elementary row operations* can be performed on any augmented matrix without changing the solution space of the associated linear system:

- Interchange two rows ($R_i \leftrightarrow R_j$)
- Multiply a row by a nonzero scalar ($R_i \leftarrow \lambda R_i$)
- Add a scalar multiple of one row to another ($R_i \leftarrow R_i + \lambda R_j$)

Example. Consider the system

$$\begin{aligned} x_1 - x_2 &= 4 \\ 2x_1 - x_2 &= 7 \end{aligned}$$

Then we form the augmented matrix and reduce using elementary row operations:

$$\left(\begin{array}{cc|c} 1 & -1 & 4 \\ 2 & -1 & 7 \end{array}\right) \xrightarrow{R_2 \leftarrow R_2 + (-2)R_1} \left(\begin{array}{cc|c} 1 & -1 & 4 \\ 0 & 1 & -1 \end{array}\right) \xrightarrow{R_1 \leftarrow R_1 + R_2} \left(\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & -1 \end{array}\right)$$

Translating back into a system tells us $x_1 = 3$ and $x_2 = -1$. In fact, this is the unique solution to the system.

Before looking at another example, we will discuss the general algorithm for reducing marices.

Definition: (Reduced) row echelon form

A matrix is in *row echelon form* (REF) if satisfies the following:

(1) any rows consisting entirely of zeros are at the bottom;
(2) the first nonzero entry of each nonzero row is 1 (called a *leading 1*);
(3) the leadings 1's of the nonzero rows move strictly to the right as we proceed down the matrix.

A matrix is in *reduced row echelon form* (RREF) if it is in REF and is satisfies the additional condition:

(4) each leading 1 is the only nonzero entry in its column.

The process of putting a matrix in (R)REF is called *Gauss-Jordan elimination*. A leading 1 is also sometimes called a *pivot*.

Example. Put the following matrix in RREF:

$$\left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{array}\right) \xrightarrow{R_2 \leftarrow R_2 + (-1)R_1} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{array}\right) \xrightarrow{R_3 \leftarrow R_3 + (-2)R_1} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -6 \\ 4 & 5 & 4 & 2 \end{array}\right)$$

$$\xrightarrow{R_4 \leftarrow R_4 + (-4)R_1} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -6 \\ 0 & -3 & -12 & -18 \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_4} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & -3 & -12 & -18 \\ 0 & 0 & -3 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2 \leftarrow -\frac{1}{3}R_2} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & -3 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{R_3 \leftarrow -\frac{1}{3}R_3} \left(\begin{array}{cccc|c} 1 & 2 & 4 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow[R_1 \leftarrow R_1 + (-4)R_3]{R_2 \leftarrow R_2 + (-4)R_3} \left(\begin{array}{cccc|c} 1 & 2 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_1 \leftarrow R_1 + (-2)R_2} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This example corresponds to a system with a unique solution, $(x_1, x_2, x_3) = (1, -2, 2)$.

Now we consider several examples of matrices in RREF and how these correspond to solutions of systems.

Example. (1) Consider the system with augmented matrix:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Any solution must satisfy $x_1 = 2$ and $x_3 = 7$. However, there are no conditions on x_2 and hence any choice of x_2 gives a solution. We say x_2 is a *free variable* (the variables x_1 and x_3 are called *basic*

variables). Since there is a solution, we say the system is *consistent*. The previous two examples were also consistent.

(2) Consider the system with augmented matrix:

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{array}\right).$$

The last row corresponds to the equation $0 = 4$, which his absurd. Hence, this system has no solution and we say it is *inconsistent*. This occurs precisely when there is a leading 1 (pivot) in the augmented column.

(3) Consider the system with augmented matrix:

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & -4 & 5 \\ 0 & 0 & 0 & 1 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This system is consistent. In this case, the variables x_1, x_3, x_4 are basic, while the variables x_2, x_5 are free. We can express the solution in parametric form as

$$x_1 = 1 - 2x_2 + x_5$$

$$x_3 = 5 + 4x_5$$

$$x_4 = 2 - 6x_5$$

$$x_2, x_5 \text{ arbitrary/free}$$

In vector form we might write:

$$(x_1, x_2, x_3, x_4, x_5) = (1, 0, 5, 2, 0) + (-2, 1, 0, 0, 0)x_2 + (1, 0, 4, -6, 1)x_5$$

so the solution space is two-dimensional. In particular, it is a plane through the point $(1, 0, 5, 2, 0)$.

4. Multiple systems and Matrix Inversion

Suppose we have two systems $A\mathbf{x}_1 = \mathbf{y}_1$ and $A\mathbf{x}_2 = \mathbf{y}_2$. We can solve them simultaneously as in the next example.

Example. Consider the two linear systems which have the same coefficient matrix,

$$\begin{aligned} x_1 - x_2 &= 4 & \qquad x_1 - x_2 &= 2 \\ 2x_1 - x_2 &= 7 & \qquad 2x_1 - x_2 &= 1 \end{aligned}$$

We form an extended augmented matrix as follows and row reduce the coefficient matrix:

$$\left(\begin{array}{cc|cc} 1 & -1 & 4 & 2 \\ 2 & -1 & 7 & 1 \end{array}\right) \xrightarrow{R_2 \leftarrow R_2 + (-2)R_1} \left(\begin{array}{cc|cc} 1 & -1 & 4 & 2 \\ 0 & 1 & -1 & -3 \end{array}\right) \xrightarrow{R_1 \leftarrow R_1 + R_2} \left(\begin{array}{cc|cc} 1 & 0 & 3 & -1 \\ 0 & 1 & -1 & -3 \end{array}\right)$$

Hence, we find that there are unique solutions to the two systems. In the first case, the solutions is $(3, -1)$, and in the second it is $(-1, -3)$.

The previous example extends easily to many systems with the same coefficient matrix. We will be most interested in the following equation for A and $n \times n$ matrix:

$$AX = I_n.$$

Here, X must be another $n \times n$ matrix. If A is invertible, then of course $X = A^{-1}$. On the other hand, suppose there exists a unique solution X. Then we can find this solution by row reducing $(A|I_n)$ to obtain $(I_n|X)$, so X is a *right inverse* of A. (It is also a *left inverse*, but this remains to be proven.)

Example. We find the inverse to the coefficient matrix from he previous problem using this method:

$$A = \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix}.$$

We form an extended augmented matrix $(A|I_2)$ and row reduce to find $(I_2|A^{-1})$:

$$\left(\begin{array}{cc|cc} 1 & -1 & 1 & 1 \\ 2 & -1 & 0 & 0 \end{array}\right) \xrightarrow{R_2 \leftarrow R_2 + (-2)R_1} \left(\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{array}\right) \xrightarrow{R_1 \leftarrow R_1 + R_2} \left(\begin{array}{cc|cc} 1 & 0 & -1 & 1 \\ 0 & 1 & -2 & 1 \end{array}\right)$$

Thus,

$$A^{-1} = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix},$$

which is easy to verify.

One can show (and in particular you will on the next homework) that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$, and in case this is nonzero, we have

$$A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Example. Find the inverse of $A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{pmatrix}$.

We form the augmented matrix $[A|I_3]$ and row reduce A to I_3.

$$[A|I_3] = \left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right).$$

Thus, $A^{-1} = \begin{pmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{pmatrix}$.

The column space and null space

In this brief section, we return to some of the ideas from above using some different terminology.

Definition: Column space, null space

Let $A \in \mathrm{Mat}_{m\times n}(k)$. The *column space of* A, denoted $\mathrm{col}(A)$, is the span of the columns of A. The *null space of* A, denoted $\mathrm{nul}(A)$, is the set of solutions of the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$.

Since $\mathrm{col}(A)$ is the span of a set of vectors in k^m, then it is a subspace of k^m. Since $\mathrm{nul}(A)$ is just the kernel of the map $T_A : k^n \to k^m$ given by $\mathbf{x} \mapsto A\mathbf{x}$, then $\mathrm{nul}(A)$ is a subspace of k^n. But we've seen this already! Look back at Proposition 5. Note that $\mathrm{col}(A)$ can also be defined as the set of $\mathbf{y}$ such that $A\mathbf{x} = \mathbf{y}$ for some $\mathbf{x} \in k^n$, which is also the image of T_A. By Rank-Nullity, we have

$$\dim(\mathrm{col}(A)) + \dim(\mathrm{nul}(A)) = n.$$

We say $\dim(\mathrm{col}(A))$ is the *rank* of the matrix A (which is just the rank of T_A) and $\dim(\mathrm{nul}(A))$ is the *nullity* of A (which is just the nullity of T_A).

Proposition 8: Basis of $\mathrm{col}(A)$

The pivot columns of $A \in \mathrm{Mat}_{m\times n}(k)$ are a basis for $\mathrm{col}(A)$.

Proof. Let R be the reduced row echelon form for A. The pivot columns are clearly linearly independent in R because they are coordinate basis vectors. On the other hand, the non-pivot columns depend linearly on the pivot columns. Hence, the pivot columns span $\mathrm{col}(R)$.

Now note that $A\mathbf{x} = \mathbf{0}$ and $R\mathbf{x} = \mathbf{0}$ have the same solution space (row operations do not change the solution set). Hence, the relations between the columns in A are the same as the relations between the columns in B. It follows that the pivot columns in A are linearly independent, while the non-pivot columns depend linearly on the pivot columns. The result follows. □

Example. Let $A = \begin{pmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & 4 \end{pmatrix}$.

We will find a basis for $\mathrm{col}(A)$ and $\mathrm{nul}(A)$. The matrix A row reduces to

$$R = \begin{pmatrix} 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The pivot columns are 1, 3, and 5. Hence, a basis for $\text{col}(A)$ is

$$\left\{ \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix} \begin{pmatrix} -7 \\ -1 \\ 4 \end{pmatrix}. \right\}.$$

Note that it's not actually necessary to put A in RREF to determine the pivot columns. Once we have identified our pivots we are done. However, to compute a basis for $\text{nul}(A)$ we need the *RREF*. The free variables in the corresponding linear system correspond to the non-pivot columns. Thus, the solution space of the matrix equation $A\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} x_2 + \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix} x_4.$$

These are linearly independent, and so a basis for $\text{nul}(A)$ is

$$\left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

This is useful for finding bases in k^n (really, any vector space). Say $W = \text{span}(v_1, \ldots, v_n)$ with $v_i \in k^m$. Now let A be the matrix whose columns are $v_1, \ldots, v_n$. A basis for $\text{col}(A)$ will then be a basis for W. In order to explain this for an arbitrary vector space, we first need the result below.

Proposition 9: Isomorphisms preserve bases

Let $T : V \to W$ be an isomorphism of vector spaces. If $\{v_1, \ldots, v_r\}$ is a basis for V, then $\{T(v_1), \ldots, T(v_r)\}$ is a basis for W.

Proof. Let $w \in W$. Since T is surjective, there is some $v \in V$ such that $T(v) = w$. Write $v = \lambda_1 v_1 + \cdots + \lambda_r v_r$ with $\lambda_i \in k$. Then

$$w = T(v) = T(\lambda_1 v_1 + \cdots + \lambda_r v_r) = \lambda_1 T(v_1) + \cdots + \lambda_r T(v_r).$$

Thus, $\{T(v_1), \ldots, T(v_r)\}$ spans W. Now suppose there are scalars μ_i such that

$$\mu_1 T(v_1) + \cdots + \mu_r T(v_r) = \mathbf{0}.$$

Hence, $T(\mu_1 v_1 + \cdots + \mu_r v_r) = \mathbf{0}$. Since T is injective, then $\mu_1 v_1 + \cdots + \mu_r v_r = \mathbf{0}$. The v_i are linearly independent, and so $\mu_i = 0$ for all i. Thus, $\{T(v_1), \ldots, T(v_r)\}$ is linearly independent and hence a basis. □

The particular isomorphism we will use is the coordinate map corresponding to a particular basis.

Example. Let W be the subspace of $\mathbb{R}[x]$ consisting of polynomials of degree at most two. One basis for W is $\mathcal{E} = \{1, x, x^2\}$. Let $\gamma_{\mathcal{E}}$ be the coordinate map corresponding to this basis. Note that $\gamma_{\mathcal{E}}(1) = e_1$, $\gamma_{\mathcal{E}}(x) = e_2$, and $\gamma_{\mathcal{E}}(x^2) = e_3$.

Consider the set $\mathcal{B} = \{1, 1 + x, 1 - x^2\}$. Then

$$\gamma_{\mathcal{E}}(1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \gamma_{\mathcal{E}}(1 + x) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \gamma_{\mathcal{E}}(1 - x^2) = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

The matrix A whose columns are these three vectors row reduces to the identity matrix. (Without row reducing, we can see this because there are clearly three pivots in the matrix.) It follows that the set of these vectors is a basis for $\mathbb{R}^3$. Thus, $\mathcal{B}$ is a basis for W (because $\gamma_{\mathcal{E}}$ is an isomorphism.

Representation of Linear Transformations

We have already learned the definition of a linear transformation. In this chapter we study them more closely, along with their connection to matrices.

Equivalence relations and similarity

We present this topic a little out of order. It will reappear in a natural way later in this chapter.

Definition: Similar matrices

We say matrices $A, B \in M_n(k)$ are *similar* if there exists $P \in \mathrm{GL}_n(k)$ such that $B = P^{-1}AP$.

Our notation for similar matrices is $A \sim B$.

Example. The matrices $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} -4 & -5 \\ 3 & 4 \end{pmatrix}$ are similar. To see this, just take $P = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ and check that $B = P^{-1}AP$.

The following proposition might be reminiscent of other relationships you have studied, such as congruent or similar triangles.

Proposition 1: Similarity is an equivalence relation

For all matrices $A, B, C \in M_n(k)$, we have

(1) $A \sim A$;
(2) If $A \sim B$, then $B \sim A$;
(3) If $A \sim B$ and $B \sim C$, then $A \sim C$.

Proof. (1) Note that $I_n^{-1}AI_n = A$, so $A \sim A$.

(2) Assume $A \sim B$, so there exists $P \in \mathrm{GL}_n(k)$ such that $B = P^{-1}AP$. But then $B \sim A$ since

$$A = PBP^{-1} = (P^{-1})^{-1}B(P^{-1}).$$

(3) Assume $A \sim B$ and $B \sim C$. Then there exists $P, Q \in \mathrm{GL}_n(k)$ such that $B = P^{-1}AP$ and $C = Q^{-1}BQ$. Substituting we have

$$C = Q^{-1}BQ = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) = (PQ)^{-1}A(PQ).$$

Hence, $A \sim C$. □

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapter 6. Last Updated: November 5, 2021

Now we discuss equivalence relations in more detail. Recall that a *relation* on sets A and B is a subset of $A \times B$. An *equivalence relation* is a special type of relation on $A \times A$.

Definition: Equivalence relation

Let A be a set. A relation $R \subset A \times A$ is an *equivalence relation* if it satisfies the following properties:

(1) (reflexive property) $(a, a) \in R$ for all $a \in A$;
(2) (symmetric property) If $(a, b) \in R$, then $(b, a) \in R$;
(3) (transitive property) If $(a, b), (b, c) \in R$, then $(a, c) \in R$.

Instead of the subset notation, we will often replace the subset notation $(a, b) \in R$ and use the notation $a \sim b$. In this way, the above properties become:

(1) (reflexive property) $a \sim a$ for all $a \in A$;
(2) (symmetric property) If $a \sim b$, then $b \sim a$;
(3) (transitive property) If $a \sim b$ and $b \sim c$, then $a \sim c$.

Definition: Equivalence classes

Let $\sim$ be an equivalence relation on a set A. For $x \in A$, the set

$$[a] = \{b \in A : a \sim b\}$$

is called the *equivalence class of a under $\sim$*.

We will return later to discuss the equivalence classes of similar matrices.

1. The space of linear transformations

Along with setting up some notation, the section is devoted to proving that the set of linear transformations between two vector spaces is itself a vector space.

Let V and W be vector spaces over k. Then $\text{Hom}(V, W)$ denotes the set of all linear transformations $V \to W$. Given $f, g \in \text{Hom}(V, W)$ and $\lambda \in k$, we can define functions

$$\lambda f : V \to W \text{ by } (\lambda f)(v) = \lambda(f(v)) \text{ for all } v \in V$$
$$f + g : V \to W \text{ by } (f + g)(v) = f(v) + g(v) \text{ for all } v \in V$$

It is easy to show that both are linear. For example, if $v, w \in W$ and $\mu \in k$, then

$$\begin{aligned}(f+g)(v+w) &= f(v+w) + g(v+w) = (f(v) + f(w)) + (g(v) + g(w)) \\ &= (f(v) + g(v)) + (f(w) + g(w)) = (f+g)(v) + (f+g)(w) \\ (f+g)(\mu v) &= f(\mu v) + g(\mu v) = \mu(f(v)) + \mu(g(v)) = \mu(f(v) + g(v)) = \mu(f+g)(v).\end{aligned}$$

It is left as an exercise to show that λf is linear.

Proposition 2: $\text{Hom}(V, W)$ is a vector space

The set $\text{Hom}(V, W)$ is a vector space over k with respect to the operations defined above.

Proof. This is similar to our proof that $\mathscr{C}^0(\mathbb{R})$ is a vector space. The zero element is the zero map $z : V \to W$ defined by $z(v) = \mathbf{0}$ for all $v \in V$. The additive inverse of an element $f \in \text{Hom}(V, W)$ is $(-1)f$. The details are left to the reader. □

There is additional structure here because we also have the composition operation. But we have to be careful about our domain and codomain and also because composition is noncommutative.

Proposition 3: Composition with addition and scalar multiplication

Let U, V, and W be vector spaces over k. Then

(1) $g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2$ for all $g \in \text{Hom}(V, W)$ and $f_1, f_2 \in \text{Hom}(U, V)$,
(2) $(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$ for all $g_1, g_2 \in \text{Hom}(V, W)$ and $f \in \text{Hom}(U, V)$,
(3) $(\lambda g) \circ f = g \circ (\lambda f) = \lambda(g \circ f)$ for all $g \in \text{Hom}(V, W)$, $f \in \text{Hom}(U, V)$ and $\lambda \in k$.

Proof. For (1), let $u \in U$. Then

$$\begin{aligned}(g \circ (f_1 + f_2))(u) &= g((f_1 + f_2)(u)) = g(f_1(u) + f_2(u)) \\ &= g(f_1(u)) + g(f_2(u)) = (g \circ f_1)(u) + (g \circ f_2)(u).\end{aligned}$$

Hence the result holds. The other properties are proved similarly. □

Corollary 4: $\mathrm{Hom}(V, V)$ is a k-algebra

For any vector space V over a field k, $\mathrm{Hom}(V, V)$ is both a vector space over k and a ring with unity. Moreover, the vector space and ring structures are related by the following law:

$$(\lambda g) \circ f = g \circ (\lambda f) = \lambda(g \circ f) \text{ for all } f, g \in \mathrm{Hom}(V, V) \text{ and } \lambda \in k$$

This corollary actually proves that $\mathrm{Hom}(V, V)$ is a k-algebra. We will not define this formally, except to say that it essentially is what is stated in the corollary. There are three operations: addition, "multiplication", and scalar multiplication. The addition and scalar operations make $\mathrm{Hom}(V, V)$ into a vector space. The addition and "multiplication" operation make it into a ring. Finally, to be a k-algebra, the scalar multiplication and "multiplication" operation must play nicely together. Another example of a k-algebra we have seen is $M_n(k)$.

Generally to prove that two functions are equal, we must check that they are equal on an arbitrary element of the domain (as we did in the proposition above). But for linear transformations, we need only check on basis elements.

Proposition 5: Values on a spanning set determine a linear transformation

Let $f, g \in \mathrm{Hom}(V, W)$. If S is a spanning set for V and $f(s) = g(s)$ for all $s \in S$, then $f = g$.

Proof. Let $v \in V$ and write $v = \sum_{j=1}^{n} \lambda_j v_j$ for $v_j \in S$. By linearity,

$$f(v) = f\left(\sum_{j=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j f(v_j) = \sum_{j=1}^{n} \lambda_j g(v_j) = g\left(\sum_{j=1}^{n} \lambda_j v_j\right) = g(v).$$

Hence, $f = g$. □

We end with a little digression. Since $\mathrm{Hom}(V, B)$ is a vector space, we can talk about linear transformations on it. But what is a linear transformation on the space of linear transformations? It helps to think about $f, g \in \mathrm{Hom}(V, V)$ diagramatically. We need a way to connect these, so in fact we need another map $V \to V$:

$$\begin{array}{ccc} V & \xrightarrow{f} & V \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\phi} \\ V & \xrightarrow[g]{} & V \end{array}$$

such that this diagram commutes[1]. So, formally, a linear transformation on $\mathrm{Hom}(V, V)$ is some $\phi \in \mathrm{Hom}(V, V)$ such that $g \circ \phi = \phi \circ f$.

[1] If this is interesting to you, then you might want to look into (higher) category theory.

2. The representation of $\mathrm{Hom}(k^n, k^m)$

Previously, we showed that the map $T_A : k^n \to k^m$ defined by $T_A(\mathbf{x}) = A\mathbf{x}$ is an element of $\mathrm{Hom}(k^n, k^m)$. In this section we show the converse: that every element in $\mathrm{Hom}(k^n, k^m)$ can be presented this way.

Definition: Standard matrix

The *standard matrix* of $T \in \mathrm{Hom}(k^n, k^m)$, denoted $M(T)$, is the $m \times n$ matrix over k whose jth column is the vector $T(\mathbf{e}_j)$.

Valenza calls $M(T)$ the *matrix of T with respect to the canonical bases for k^n and k^m*, which isn't wrong, but it doesn't exactly roll right off the tongue.

Example. Suppose $T : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} 2x_1 + x_3 \\ x_2 - x_3 \end{pmatrix}.$$

Then

$$T(\mathbf{e}_1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad T(\mathbf{e}_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad T(\mathbf{e}_3) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Hence,

$$M(T) = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

Note that

$$M(T) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_3 \\ x_2 - x_3 \end{pmatrix}.$$

Here is a more geometric example.

Example. Let $V = \mathbb{R}^2$ and let T_θ be the map which rotates a point in the plane counterclockwise around the origin by the angle θ. This map is linear (check!). We compute its standard matrix:

$$T_\theta(\mathbf{e}_1) = (\cos\theta, \sin\theta)$$
$$T_\theta(\mathbf{e}_2) = (-\sin\theta, \cos\theta).$$

Hence,

$$M_{T_\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

By the nature of rotation, $T_{\theta+\psi} = T_\theta \circ T_\psi$, so

$$M(T_{\theta+\psi}) = M(T_\theta)M(T_\psi).$$

As matrices

$$\begin{pmatrix} \cos(\theta+\psi) & \sin(\theta+\psi) \\ -\sin(\theta+\psi) & \cos(\theta+\psi) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}.$$

This is much easier to remember than the formulas,

$$\cos(\theta+\psi) = \cos\theta\cos\psi - \sin\theta\sin\psi$$
$$\sin(\theta+\psi) = \cos\theta\sin\psi + \sin\theta\cos\psi.$$

Proposition 6: Linear transformations are matrices are linear transformations

The mappings

$$\begin{aligned} \mathrm{Mat}_{m\times n}(k) \to \mathrm{Hom}(k^n, k^m) \qquad & \qquad \mathrm{Hom}(k^n, k^m) \mapsto \mathrm{Mat}_{m\times n}(k) \\ A \mapsto T_A \qquad & \qquad T \mapsto M(T) \end{aligned}$$

are mutually inverse isomorphisms of vector spaces. In particular, given any linear transformation $T \in \mathrm{Hom}(k^n, k^m)$, $M(T)$ is the unique matrix such that

$$T(\mathbf{x}) = M(T)\mathbf{x} \text{ for all } \mathbf{x} \in k^n.$$

Proof. Let $A \in \mathrm{Mat}_{m\times n}(k)$. We claim $A = M(T_A)$. By definition, the jth column of $M(T_A)$ is $T_A(\mathbf{e}_j) = A\mathbf{e}_j = A^j$, the jth column of A.

On the other hand, consider $T \in \mathrm{Hom}(k^n, k^m)$. We claim $T = T_{M(T)}$. By definition, $T_{M(T)}(\mathbf{e}_j)$ is the matrix product $M(T)\mathbf{e}_j$, which is the jth column of $M(T)$. Hence, $T(\mathbf{e}_j) = T_{M(T)}(\mathbf{e}_j)$ for all canonical basis vectors. Thus, they are equal linear transformations.

We need only prove linearity for one of the maps, since the other is the inverse and inverses of linear transformations are linear transformations. We leave this easy check as an exercise. □

It turns out that these maps also respect the operation of composition. That is, the corresponding maps $M_n(k) \leftrightarrow \mathrm{Hom}(k^n, k^n)$ are ring homomorphisms. Thus, they are k-algebra isomorphisms.

The next result is immediate because we showed previously that $\dim(\mathrm{Mat}_{m\times n}(k)) = mn$.

Corollary 7: Dimension of $\mathrm{Hom}(k^n, k^m)$

The dimension of $\mathrm{Hom}(k^n, k^m)$ is mn.

We now relate matrix operations to our linear transformations. As implied earlier, this makes proving several of the matrix properties we skipped much easier.

Proposition 8: Matrix multiplication is associative

(1) If $A \in \mathrm{Mat}_{m\times n}(k)$ and $B \in \mathrm{Mat}_{n\times p}(k)$, then $T_{AB} = T_A \circ T_B$.
(2) Matrix multiplication is associative.
(3) If $T : k^p \to k^n$ and $S : k^n \to k^m$, then $M(S \circ T) = M(S)M(T)$.

Proof. (1) To prove $T_{AB} = T_A \circ T_B$, it suffices to prove that they have the same effect on the canonical basis vectors. First note that

$$T_{AB}(\mathbf{e}_j) = (AB)\mathbf{e}_j = (AB)^j$$

which is the jth column of AB. On the other hand

$$(T_A \circ T_B)(\mathbf{e}_j) = T_A(T_B(\mathbf{e}_j)) = T_A(B^j) = A \cdot B^j,$$

which is A times the jth column of B. But by definition of matrix multiplication, this is exactly the jth column of AB.

(2) Suppose $A(BC)$ and $(AB)C$ are defined. Then by (1) and properties of function composition,

$$T_{(AB)C} = T_{AB} \circ T_C = (T_A \circ T_B) \circ T_C = T_A \circ (T_B \circ T_C) = T_A \circ T_{BC} = T_{A(BC)}.$$

Taking standard matrices of both sides gives the result.

(3) Let $\mathbf{x} \in k^p$. Then

$$M(S \circ T)\mathbf{x} = (S \circ T)(\mathbf{x}) = S(T(\mathbf{x})) = S(M(T)\mathbf{x}) = M(S)(M(T)\mathbf{x}) = (M(S)M(T))\mathbf{x}.$$

Thus, $M(S \circ T) = M(S)M(T)$. □

3. The representation of $\mathrm{Hom}(V, V')$

We will be assuming here, as we often do, that all vector spaces are finite dimensional. Recall that a linear transformation $T : V \to V'$ can be represented by a matrix, but this definition depends on our choice of coordinate systems (bases) for V and V'.

Suppose $\dim(V) = n$ and let $B = \{v_1, \ldots, v_n\}$ be a basis for V. Note that the order of the elements is important here. Recall that the coordinate map $\gamma_B : V \to k^n$ is an isomorphism. Similarly, if $\dim(V') = m$ and V' has basis $B' = \{v'_1, \ldots, v'_m\}$, then we have a coordinate map $\gamma_{B'} : V' \to k^m$. Thus, we have a commutative diagram:

$$\begin{array}{ccc} V & \xrightarrow{T} & V' \\ {\scriptstyle \gamma_B}\downarrow & & \downarrow{\scriptstyle \gamma_{B'}} \\ k^n & \xrightarrow[\gamma_{B'} \circ T \circ \gamma_B^{-1}]{} & k^m \end{array}$$

The bottom map is in fact the *unique* map making the diagram commute. Since $\gamma_{B'} \circ T \circ \gamma_B^{-1}$ is a linear transformation $k^n \to k^m$, then it is represented by a matrix. We denote this matrix by $M_{B,B'}(T)$. This is the unique matrix such that

$$\gamma_B(T(v)) = M_{B,B'}(T)\gamma_B(v) \quad \text{for all } v \in V.$$

We say that $M_{B,B'}$ *represents* T. We can then rewrite the previous commutative diagram as

$$\begin{array}{ccc} V & \xrightarrow{T} & V' \\ {\scriptstyle \gamma_B}\downarrow & & \downarrow{\scriptstyle \gamma_{B'}} \\ k^n & \xrightarrow[M_{B,B'}(T)]{} & k^m \end{array}$$

Theorem 9: Matrix representation is an isomorphism

Let V and V' be finite-dimensional vector spaces over the field k, with respective bases B and B' and dimensions n and m. then the map

$$\begin{aligned} \mathrm{Hom}(V, V') &\to \mathrm{Mat}_{m \times n}(k) \\ T &\mapsto M_{B,B'}(T) \end{aligned}$$

is an isomorphism of vector spaces.

Proof. Let $S, T \in \mathrm{Hom}(V, V')$. Note that

$$\gamma_{B'} \circ (S + T) \circ \gamma_B^{-1} = \gamma_{B'} \circ S \circ \gamma_B^{-1} + \gamma_{B'} \circ T \circ \gamma_B^{-1}.$$

The map on the left is represented by $M_{B,B'}(S)$ and the one of the right by $M_{B,B'}(T)$. One checks that λT is a linear transformation for any $\lambda \in k$. Thus the given map is a linear transformation.

Since the zero matrix is only represented by the zero map, then it follows that the map is injective. It is left to prove surjectivity. Let $A \in \mathrm{Mat}_{m\times n}(k)$ and let $T = \gamma_{B'}^{-1} \circ T_A \circ \gamma_B$ where $T_A : k^n \to k^m$ is the usual map defined by left multiplication by the matrix A. Then

$$\gamma_{B'} \circ T \circ \gamma_B^{-1} = \gamma_{B'} \circ \left(\gamma_{B'}^{-1} \circ T_A \circ \gamma_B\right) \circ \gamma_B^{-1} = T_A,$$

which has standard matrix A. Hence, $M_{B,B'}(T) = A$, so the map is surjective. □

If we specialize to the case that $V = V'$ and choose a basis B, then we can abbreviate $M_{B,B}(T)$ to $M_B(T)$. Then one can in fact prove that there is an isomorphism of k-algebras:

$$\begin{aligned} \mathrm{Hom}(V, V) &\to M_n(k) \\ T &\mapsto M_B(T) \end{aligned}$$

The next result generalizes what we have already shown for standard matrices and matrix products.

Proposition 10: Composition corresponds to products

Let V, V', and V'' be finite-dimensional vector spaces over the field k, with bases B, B', and B'', respectively. Let $T : V \to V'$ and $T' : V' \to V''$ be linear transformations. Then

$$M_{B,B''}(T' \circ T) = M_{B',B''}(T')M_{B,B'}(T).$$

Proof. We could prove this directly, or we could use commutative diagrams:

$$\begin{array}{ccccc} V & \xrightarrow{T} & V' & \xrightarrow{T} & V'' \\ {\scriptstyle\gamma_B}\downarrow & & \downarrow{\scriptstyle\gamma_{B'}} & & \downarrow{\scriptstyle\gamma_{B''}} \\ k^p & \xrightarrow[M_{B,B'}(T)]{} & k^n & \xrightarrow[M_{B',B''}(T')]{} & k^m \end{array}$$

On one hand, the diagram shows that

$$(\gamma_{B''} \circ T' \circ T)(v) = M_{B,B'}(T)M_{B',B''}(T')\gamma_B(v).$$

On the other hand, $M_{B,B''}(T' \circ T)$ is the unique map making the outer diagram commute. The result follows. □

Let v_j be the jth element of the basis B (remember that order matters). Then since $\gamma_B(v_j) = \mathbf{e}_j$, we have

$$M_{B,B'}(T)\mathbf{e}_j = M_{B,B'}(T)\gamma_B(v_j) = \gamma_{B'}(T(v_j)).$$

That is, the jth column of $M_{B,B'}(T)$ is the jth coordinate vector of $T(v_j)$.

Example. (1) Let V be the subspace of $\mathcal{C}^0(\mathbb{R})$ spanned by $\sin x$ and $\cos x$. Hence, $\{\sin x, \cos x\}$ is a basis for W. Let $D : V \to V$ be the differentiation operator. Then

$$D(\sin x) = 0 \cdot \sin x + 1 \cdot \cos x$$
$$D(\cos x) = (-1) \cdot \sin x + 0 \cdot \cos x.$$

Thus, the matrix of D is

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Now if we take $f(x) = 2\sin x + 5\cos x \in W$, then $D(f)$ can be computed using only linear algebra by

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} -5 \\ 2 \end{pmatrix}$$

which matrices what we know from calculus that $f'(x) = -5\sin x + 2\cos x$. Now taking the second power of A we have

$$A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is the matrix of $D^2 = D \circ D$, the second derivative operator. Note that $A^4 = I_2$.

(2) Let W be the subspace of $\mathcal{C}^0(\mathbb{R})$ consisting of polynomials of degree at most 4. Then one basis for W is $\{1, x, x^2, x^3, x^4\}$. The differentiation operator has matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

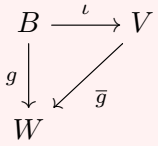
Note that A^5 is the zero matrix.

4. The dual space

Duality is an important part of mathematics, especially in algebra. Here we will discuss formally the dual space of a vector space.

Theorem 11: Extending a linear transformation

Let V be a vector space over k with basis B. Given any vector space W over k and any function $g : B \to W$, there is a unique linear transformation $\overline{g} : V \to W$ such that the following diagram commutes:



Proof. Let $v \in V$. Then v can be written uniquely as a linear combination of the basis elements:

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

with $v_i \in B$ and $\lambda_i \in k$. Now define $\overline{g}$ by

$$\overline{g}(v) = \lambda_1 g(v_1) + \cdots + \lambda_n g(v_n).$$

It is easy to check that $\overline{g} \circ \iota = g$. That is, $\overline{g}$ extends g. If $v_i \in B$, then $\iota(v_i) = v_i$ so

$$(\overline{g} \circ \iota)(v) = (\overline{g}(v) = g(v_i).$$

It remains to show that $\overline{g}$ is linear. Let $v, w \in V$. Write,

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$
$$w = \mu_1 v_1 + \cdots + \mu_n v_n,$$

with $\lambda_i, \mu_i \in k$. Then

$$v + w = (\lambda_1 + \mu_1)v_1 + \cdots + (\lambda_n + \mu_n)v_n$$

so by definition of $\overline{g}$ we have

$$\begin{aligned}
\overline{g}(v + w) &= (\lambda_1 + \mu_1)g(v_1) + \cdots + (\lambda_n + \mu_n)g(v_n) \\
&= \lambda_1 g(v_1) + \mu_1 g(v_1) + \cdots + \lambda_n g(v_1) + \mu_n g(v_n) \\
&= (\lambda_1 g(v_1) + \cdots + \lambda_n g(v_n)) + (\mu_1 g(v_1) + \cdots + \mu_n g(v_n)) \\
&= \overline{g}(v) + \overline{g}(w).
\end{aligned}$$

A similar argument shows $\overline{g}(cv) = c\overline{g}(v)$ for any scalar c. □

We say $\overline{g}$ in the above theorem is an *extension by linearity*. What this result says is that we are *free* to send the basis elements anywhere in W and this determines the linear transformation $\overline{g}$. This result depends on the linear independence of the elements of the basis B. Hence, we say the vector space is *free on its basis*, explaining the phrase, *vector spaces are free!*

Definition: Dual space

Let V be a vector space over k. Then $V^* = \mathrm{Hom}(V, k)$ is called the *dual space of* V.

Note that $\mathrm{Hom}(V, k)$ is the set of all vector space homomorphisms (linear transformations) from V to k and is itself a vector space over k. These homomorphisms are the *dual* elements to those in V, but also there is a duality between the linear transformations with domain T. Let $T : V \to W$ be a linear transformation. For any $f \in W^* = \mathrm{Hom}(W, k)$, the composed map $f \circ T$ is a homomorphism from V to k. We define the *transpose map of* T as

$$T^* : W^* \to V^*$$
$$f \mapsto f \circ T$$

which is itself a linear transformation.

Proposition 12: Properties of transposition

(1) For any vector space V over k, $(1_V)^* = 1_{V^*}$.
(2) If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations of vector spaces over k, then $(T_2 \circ T_1)^* = T_1^* \circ T_2^*$.

Proof. (1) Note that 1_V is a map from V to V, so $(1_V)^*$ is a map from V^* to V^*. Let $f \in V^*$, then $(1_V)^*(f) = f \circ 1_V = f = 1_{V^*}(f)$.

(2) Here, $T_2 \circ T_1 : U \to W$, so $(T_2 \circ T_1)^* : W^* \to U^*$. (Note that $T_1^* \circ T_2^*$ has the same domain and codomain.) Let $f \in W^*$. Then by associativity of function composition we have

$$(T_2 \circ T_1)^*(f) = f \circ (T_2 \circ T_1) = (f \circ T_2) \circ T_1 = T_2^*(f) \circ T_1 = T_1^*(T_2^*(f)) = (T_1^* \circ T_2^*)(f). \qquad \square$$

This type of order-reversing relationship is known as *contravariance*.

Proposition 13: Injectivity and surjectivity of transpose maps

Let $T : V \to W$ be a linear transformation of vector spaces over k. Then

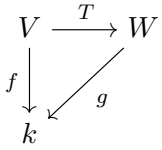
(1) if T is surjective, then T^* is injective;
(2) if T is injective, then T^* is surjective.

Proof. (1) Let $f, g \in W^*$ such that $T^*(f) = T^*(g)$. That is, $f \circ T = g \circ T$. Let $w \in W$. By surjectivity of T there exists $v \in V$ such that $T(v) = w$. Then

$$f(w) = f(T(v)) = (f \circ T)(v) = (g \circ T)(v) = g(T(v)) = g(w).$$

Hence, $f = g$ so T^* is injective.

(2) Let $f \in V^* = \mathrm{Hom}(V, k)$. We claim there exists $g \in W^*$ such that $f = T^*(g) = g \circ T$. That is, we mush find a function g that makes the following diagram commute:



Let B be a basis for V. Then T is injective so $T(B)$ is a linearly independent set in W. We can extend this to a basis of W, say B'. Define a map $B' \to k$ by the rule:

$$w \mapsto \begin{cases} f(v) & \text{if } w = T(v) \text{ for some } v \in B \\ 0 & \text{otherwise} \end{cases}$$

By the result above, this set map extends uniquely to a linear transformation $g : W \to k$. Then by construction, $f = g \circ T$. □

We will use this to prove that rank of a linear transformation is preserved by duality. Before doing this, we need a preparatory result, but one that is important in its own right. This will also explain the usage of the term transpose map for T^*.

Let V be a finite-dimensional vector space with basis $v_1, \ldots, v_n$. In V^*, define the elements $v_1^*, \ldots, v_n^*$ to be the unique linear maps from $V \to k$ defined by

$$(\star) \qquad v_i^*(v_j) = \delta_{ij}.$$

Since we have defined these maps on basis elements, they extend to maps on the whole of V. Furthermore, by extension of linearity, if $f \in V^*$, then

$$(\star\star) \qquad f = \sum_{j=1}^{n} f(v_j) v_j^*.$$

It follows that $v_1^*, \ldots, v_n^*$ is a basis for V^*, which we call the *dual basis*.

Proposition 14: Isomorphism between V and V^*

Keep the above notation. The linear transformation $V \to V^*$ defined by

$$v_j \mapsto v_j^*$$

is an isomorphism of vector spaces. Hence, $V \cong V^*$.

Theorem 15: Rank is preserved by duality

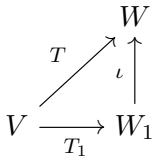
Let $T : V \to W$ be a linear transformation of a finite-dimensional vector spaces over k. Then

$$\mathrm{rk}(T) = \mathrm{rk}(T^*).$$

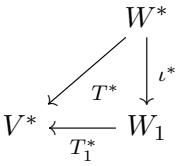
Proof. First suppose T is surjective. Then $\text{Im}(T) = W$ and $\text{rk}(T) = \dim(W)$. Hence, T^* is injective and from Rank-Nullity we have $\text{rk}(T^*) = \dim(W^*)$. But $\dim(W^*) = \dim(W)$ so

$$\text{rk}(T) = \dim(W) = \dim(W^*) = \text{rk}(T^*).$$

Now suppose T is arbitrary and let $\text{Im}(T) = W_1$. Let $\iota : W_1 \to W$ be the inclusion map and let $T_1 : V \to W_1$ be the same as T (but with codomain restricted), so $\text{rk}(T) = \text{rk}(T_1)$. Then we have the commutative diagram:



Taking duals we have



Since ι is injective, then ι^* is surjective. Moreover, $\text{Im}(T^*) = \text{Im}(T_1^*)$ and hence $\text{rk}(T^*) = \text{rk}(T_1^*)$. But T_1 is surjective, so using our special case we have $\text{rk}(T_1) = \text{rk}(T_1^*)$. Putting these together gives the result. □

We are now ready to highlight the importance of the transpose map.

Theorem 16: Matrix of dual map

Let $M(T)$ denote the matrix of T with respect to the bases B_1 and B_2 and let $M(T^*)$ denote the matrix of T^* with respect to the dual bases B_2^* and B_1^*. Then

$$M(T^*) = {}^tM(T).$$

Proof. Let $B_1 = \{v_1, \ldots, v_n\}$ and $B_2 = \{w_1, \ldots, w_m\}$. Suppose $M(T) = (a_{ij})$ with respect to bases B_1 and B_2. Then

$$\begin{aligned} T^*(w_j^*) = w_j^* \circ T &= \sum_{i=1}^{n} [w_j^* \circ T](v_i) v_i^* \quad (\text{by } (\star\star)) \\ &= \sum_{i=1}^{n} w_j^*(T(v_i)) v_i^* = \sum_{i=1}^{n} w_j^* \left(\sum_{\ell=1}^{m} a_{\ell i} w_\ell \right) v_i^* \\ &= \sum_{i=1}^{n} a_{ji} v_i^* \quad (\text{by } (\star)). \end{aligned}$$

Hence, the jth column of $M(T*)$ is a_{ji} as the entries range over i. The result now follows. □

We have previously introduced the column space of a matrix. Similarly we can define the *row space* of a matrix $A \in \mathrm{Mat}_{m\times n}(k)$, denoted $\mathrm{row}(A)$ as the subspace of k^n spanned by the rows of A.

Corollary 17: Rank of a matrix equals rank of its transpose

Let $A \in \mathrm{Mat}_{m\times n}(k)$. Then $\mathrm{rk}(A) = \mathrm{rk}(A^T)$. Consequently, $\dim(\mathrm{col}(A)) = \dim(\mathrm{row}(A))$.

We are now ready to make a contribution to the Invertible Matrix Theorem, which we will add in formally later. Recall that the statements of the IMT are equivalent to an $n \times n$ matrix A having a *right* inverse B. Now note that

$$(x_1, \ldots, x_n) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = x_1 A_1 + \cdots + x_n A_n.$$

where the A_i denote the rows of A. If A is invertible, then $\mathrm{rk}(A) = n$, so $\mathrm{rk}(A^T) = n$. Hence, the rows span k^n. It follows that for each $\mathbf{e}_i$ (now written as a row vector) we can find a solution for the equation $x_1 A_1 + \cdots + x_n A_n = \mathbf{e}_i$. In this way we construct a *left* inverse for A, say C. But then we have

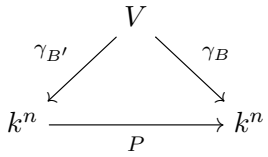
$$C = CI_n = C(AB) = (CA)B = I_n B = B.$$

That is, the left inverse and right inverse are the same. Hence, A is invertible.

5. Change of basis

We will now see how to easily switch between two bases of a given vector space using matrices.

Assume $\dim(V) = n$ and let $B = \{v_1, \ldots, v_n\}$ and $B' = \{w_1, \ldots, w_n\}$ be (ordered) bases for V. Let γ_B and $\gamma_{B'}$ be the respective coordinate isomorphisms from V to k^n. Then $P = \gamma_B \circ (\gamma_{B'})^{-1} \in M_n(k)$ is the unique isomorphism that makes the following diagram commute:



Here we abuse notation a bit because P really stands for T_P. We call P the *transition matrix* from B' to B. By construction,

$$P\gamma_{B'}(w_j) = \gamma_B(w_j).$$

Hence, the left-hand side is $P\mathbf{e}_j = P^j$, while the right-hand side is the coordinate vector of w_j with respect to the basis B. We use this idea to easily work out the next example.

Example. Let V the space of real polynomials of degree at most 2. Let $B = \{1, x, x^2\}$ and $B' = \{1+x, 1-x, 1+x^2\}$. You should verify that B' is indeed a basis of V. The transition matrix is then

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

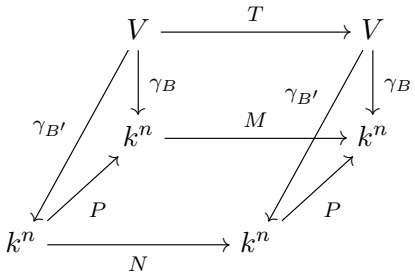
Theorem 18: Change of basis formula

Let $T : V \to V$ be an endomorphism of a finite-dimensional vector space. Let M denote the matrix of T with respect to the basis B and let N denote the matrix of T with respect to the basis B'. Then

$$N = P^{-1}MP$$

where P is the transition matrix from B' to B.

Proof. We have the following diagram:



We use commutativity of various diagrams within this to prove that the bottom face is commutative, which proves the result.

The triangles on either side are commutative because these define the transition matrix P. Similarly, the front slanted face and the back face are commutative because these define the matrices N and M, respectively. We have

$$\begin{aligned} N\gamma_{B'} &= \gamma_{B'}T \quad \text{(commutativity of front face)} \\ PN\gamma_{B'} &= P\gamma_{B'}T \quad \text{(multiplying left sides by } P\text{)} \\ &= \gamma_B T \quad \text{(commutativity of front face)} \\ &= M\gamma_B \quad \text{(commutativity of back face)} \\ &= MP\gamma_{B'} \quad \text{(commutativity of left triangle).} \end{aligned}$$

Multiplying both sides by $\gamma_{B'}^{-1}$ gives $PN = MP$. □

Example. Continuing the previous example, we have

$$P^{-1} = \begin{pmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let D be the matrix of differentiation with respect to B. Then the matrix of differentiation with respect to B' is

$$P^{-1}DP = \begin{pmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 & 1 \\ 1/2 & -1/2 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that this matrix is similar to D.

Proposition 19: Similarity corresponds to change of basis

Given $A, B \in M_n(k)$, A is similar to B if and only if there exists a vector space V and an endomorphism T of V such that both A and B represent T with respect to (possibly) different bases.

Proof. One direction is just the change of basis formula.

Suppose $B = P^{-1}AP$. Let $V = k^n$ so A represents T_A, which is left multiplication by A with respect to the canonical basis. Since P is invertible, then the columns of P are also a basis for k^n. The transition matrix from $P^1, \ldots, P^n$ to the canonical basis is just P itself. By the theorem, the matrix of T_A with respect to the new basis is $P^{-1}AP$. □

Inner product spaces

This chapter allows us to discuss further geometric properties of vectors such as lengths and angles.

1. Real inner product spaces

We define an inner product space to be a vector space along with extra structure. However, it is possible to have different structures on the same vector space.

Definition: Inner product space

A *real inner product space* V is a real vector space together with a map

$$V \times V \to \mathbb{R}$$
$$(v, w) \mapsto \langle v|w\rangle$$

called a *real inner product*, satisfying the following properties:

(1) (positive definiteness) $\langle v|v\rangle \geq 0$ for all $v \in V$, with equality if and only if $v = \mathbf{0}$
(2) (symmetry) $\langle v|w\rangle = \langle w|v\rangle$ for all $v, w \in V$
(3) (bilinearity) $\langle u + v|w\rangle = \langle u|w\rangle + \langle v|w\rangle$ and $\langle \lambda v|w\rangle = \lambda \langle v|w\rangle$ for all $u, v, w \in V$ and $\lambda \in R$.

By (2), the bilinearity properties hold equally well in the second coordinate. These properties imply that for any $v_0 \in V$, the following maps are linear:

$$V \to \mathbb{R} \qquad\qquad V \to \mathbb{R}$$
$$v \mapsto \langle v|v_0\rangle \qquad\qquad v \mapsto \langle v_0|v\rangle$$

The first example below will seem familiar to anyone who has had a little multivariable calculus.

Example. (1) Let $V = \mathbb{R}^n$. The *dot product* on V is defined by

$$\langle \mathbf{x}|\mathbf{y}\rangle = \sum_{i=1}^{n} x_i y_i.$$

Checking the inner product properties is an easy exercise, but we do it here anyways. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and let $\lambda \in \mathbb{R}$. Then

$$\langle \mathbf{x}|\mathbf{x}\rangle = \sum_{i=1}^{n} x_i x_i = \sum_{i=1}^{n} x_i^2.$$

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapter 7. Last Updated: November 8, 2021

A sum of squares is necessarily nonnegative. Moreover, it will be zero if and only if each $x_i = 0$. This proves positive definiteness. For symmetry,

$$\langle \mathbf{x}|\mathbf{y}\rangle = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i = \langle \mathbf{y}|\mathbf{x}\rangle .$$

Finally, for bilinearity we have

$$\langle \mathbf{x}+\mathbf{z}|\mathbf{y}\rangle = \sum_{i=1}^{n}(x_i+z_i)y_i = \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} z_i y_i = \langle \mathbf{x}|\mathbf{y}\rangle + \langle \mathbf{z}|\mathbf{y}\rangle$$
$$\langle \lambda\mathbf{x}|\mathbf{y}\rangle = \sum_{i=1}^{n}(\lambda x_i)y_i = \lambda \sum_{i=1}^{n} x_i y_i = \lambda \langle \mathbf{x}|\mathbf{y}\rangle .$$

(2) Let $\mathscr{C}^0([a,b])$, the vector space of continuous functions with domain $[a,b]$. Define

$$\langle f|g\rangle = \int_a^b f(x)g(x)dx.$$

It is left as an exercise to check that this is an inner product.

We can now use inner products to define length for vectors. This recovers our intuitive notion for length in $\mathbb{R}^2$ using the dot product.

Definition: Length/norm

Let V be an inner product space. Then given a vector $v \in V$, we define the *length* (or *norm*) of v, denoted $|v|$ by

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

A vector of length one is called a *unit vector*.

The definition of the inner product implies immediate the following two properties:

(1) $|v| = 0$ if and only if $v = \mathbf{0}$ for all $v \in V$
(2) $|\lambda v| = |\lambda| \cdot |v|$ for all $v \in V$ and $\lambda \in \mathbb{R}$.

Example. (1) Taking the dot product, we have

$$|v| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2},$$

which agrees with length obtained through the Pythagorean Theorem.

(2) For the second example, we have

$$|f| = \left(\int_a^b f(x)^2\right)^{1/2}.$$

The next two results tell us how length behaves with respect to the dot product and to vector addition.

Theorem 1: The Cauchy-Schwarz Inequality

Let V be an inner product space. Then for all vectors $v, w \in V$,

$$|\langle v|w\rangle| \leq |v| \cdot |w|.$$

Proof. Let $x \in \mathbb{R}$ and $v, w \in V$. Then

$$\begin{aligned} 0 &\leq \langle v + xw | v + xw\rangle \\ &= \langle v|v\rangle + 2x\langle v|w\rangle + x^2\langle w|w\rangle \\ &= |v|^2 + 2x\langle v|w\rangle + x^2|w|^2. \end{aligned}$$

This is a polynomial in the variable x. Hence, it has at most one real root, so the discriminant is at most 0. That is,

$$4\langle v|w\rangle^2 - 4|v|^2|w|^2 \leq 0,$$

and the result follows from basic algebra. □

Corollary 2: The Triangle Inequality

For all $v, w \in V$,

$$|v + w| \leq |v| + |w|.$$

Proof. We have

$$\begin{aligned} |v+w|^2 &= \langle v+w|v+w\rangle \\ &= \langle v|v\rangle + 2\langle v|w\rangle + \langle w|w\rangle \\ &\leq |v|^2 + 2|\langle v|w\rangle| + |w|^2 \\ &\leq |v|^2 + 2(|v| \cdot |w|) + |w|^2 \quad \text{(by Cauchy-Schwarz)} \\ &= (|v| + |w|)^2. \end{aligned}$$

Taking square roots gives the result. □

We can use inner products to define angles between vectors. This agrees with our notion of angles in $\mathbb{R}^2$.

Definition: Angle between vectors, orthogonal

Let v and w be nonzero vectors in the inner product space V. Then the *angle between v and w* is the number $\theta \in [0, \pi]$ defined by the equation

$$\cos\theta = \frac{\langle v|w\rangle}{|v||w|}.$$

We say v is *orthogonal* to w if $\langle v|w\rangle = 0$.

If v and w are orthogonal, we write $v \perp w$. This is because, in $\mathbb{R}^2$, vectors are orthogonal if and only if the lines through those vectors are perpendicular. (That is, the angle between them is $\pi/2 = 90°$.)

Definition: Orthogonal set, orthonormal set

A set $S = \{v_1, \ldots, v_m\}$ of nonzero vectors in V is called an *orthogonal set* if it satisfies $\langle v_i|v_j\rangle = 0$ whenever $i \neq j$. If, in addition, each of the v_i is a unit vector, then we say S is an *orthonormal set*.

When the set S above is orthonormal, then the vectors satisfy

$$\langle v_i|v_j\rangle = \delta_{ij} \quad \text{for all } 1 \leq i, j \leq m.$$

Example. (1) The canonical basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal set and hence an *orthonormal basis* for $\mathbb{R}^n$.

(2) An orthonormal basis for $\mathbb{R}^2$ other than the canonical one is the set

$$\left\{\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)\right\}$$

There are many other such bases.

(3) The set of vectors consisting of

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x$$

in $\mathscr{C}^0([-\pi, \pi])$ is orthogonal. For example, making the substitution $u = \sin x$ gives

$$\int_{-\pi}^{\pi} \cos x \sin x dx = \int_0^0 u du = 0.$$

It is not difficult to verify this in general, though there are several cases to consider. Note this set is not orthonormal, but we can make it orthonormal simply by scaling each vector.

Orthogonality is much easier to check than linear independence (in general), and so the next result is useful.

Proposition 3: Orthogonality implies linear independence

Let $\{v_1, \ldots, v_m\}$ be an orthogonal set and suppose

$$\sum_{i=1}^{m} a_i v_i = \mathbf{0}.$$

Using bilinearity of the inner product, we have for any v_j,

$$0 = \left\langle \sum_{i=1}^{m} a_i v_i | v_j \right\rangle = \sum_{i=1}^{m} a_i \langle v_i | v_j \rangle = a_j \langle v_j | v_j \rangle .$$

Since $\langle v_j | v_j \rangle \neq 0$, this implies $a_j = 0$.

We end this section with a generalization of the Pythagorean Theorme to more dimensions (and for any inner product).

Theorem 4: The Pythagorean Theorem

Suppose that $\{v_1, \ldots, v_m\}$ is an orthogonal set in the inner product space V. Then

$$\left| \sum_{j=1}^{m} v_j \right|^2 = \sum_{j=1}^{m} |v_j|^2$$

Proof. This is a straightforward calculation:

$$\begin{aligned} \left| \sum_{j=1}^{m} v_j \right|^2 &= \left\langle \sum_{j=1}^{m} v_j | \sum_{j=1}^{m} v_j \right\rangle \\ &= \sum_{1 \leq i,j \leq m} \langle v_i | v_j \rangle \\ &= \sum_{j=1}^{m} \langle v_j | v_j \rangle \\ &= \sum_{j=1}^{m} |v_j|^2 . \end{aligned}$$

□

2. Orthogonal bases and orthogonal projections

In this section we assume that V is a (real) inner product space[1]. First we show a major utility of having an orthonormal basis for an inner product space. It is then established that every inner product space has an orthonormal basis through an algorithm known as the *Gram-Schmidt Process*.

Proposition 5: Coordinates relative to an orthonormal basis

Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis for V. For all $v \in V$, the following hold:

(1) $v = \sum_{j=1}^{n} \langle v|u_j\rangle \, u_j$
(2) $|v|^2 = \sum_{j=1}^{n} \langle v|u_j\rangle$

Proof. Write $v = a_1u_1 + \cdots + a_nu_n$. Then

$$\langle v|u_i\rangle = \left\langle \sum_{j=1}^{n} a_j u_j | u_i \right\rangle = a_i \langle u_i|u_i\rangle = a_i.$$

The second statement is now just the Pythagorean Theorem. □

We can think about the components of v as the projection of v onto the coordinate axes defined by the u_i. The next definition formalizes this.

Definition: Orthogonal projection

Let $v \in V$ and let $u \in V$ be a unit vector. Then

$$\mathrm{pr}_u(v) = \langle v|u\rangle \, u$$

is the *orthogonal projection of v onto u*. More generally, if W is a subspace of V with orthonormal basis $u_1, \ldots, u_m$, then

$$\mathrm{pr}_W(v) = \sum_{j=1}^{m} \langle v|u_j\rangle \, u_j$$

is the *orthogonal projection of v onto W*.

Note that $\mathrm{pr}_W(v)$ necessarily lies in W since it is a linear combination of elements of W. We should think of this projection as a shadow of the vector v in the subspace W. In particular, in the case of $\mathbb{R}^3$ this is preceisely what happens. It is not hard to verify that if $v \in W$, then $\mathrm{pr}_W(v) = v$.

[1]We will skip a discussion of *complex* inner product spaces for the time being. If time allows we will return to this.

Lemma 6: Normal vectors

Let W be a subspace of V with orthonormal basis $u_1, \ldots, u_m$. For $v \in V$,

$$(v - \mathrm{pr}_W(v)) \perp u_i \quad (i = 1, \ldots, m).$$

Hence, $v - \mathrm{pr}_W(v)$ is orthogonal to every vector in W.

Proof. Again, this is an easy calculation:

$$\begin{aligned}
\langle v - \mathrm{pr}_W(v) | u_i \rangle &= \langle v | u_i \rangle - \langle \mathrm{pr}_W(v) | u_i \rangle \\
&= \langle v | u_i \rangle - \left\langle \sum_{j=1}^{m} \langle v | u_j \rangle \, u_j | u_i \right\rangle \\
&= \langle v | u_i \rangle - \sum_{j=1}^{m} \langle v | u_j \rangle \, \langle u_j | u_i \rangle \\
&= \langle v | u_i \rangle - \langle v | u_i \rangle \, \langle u_i | u_i \rangle = 0.
\end{aligned}$$

□

In the language of Calc II, suppose W is a two-dimensional subspace of $\mathbb{R}^3$ (so it represents a plane in $\mathbb{R}^3$). Then we say that $v - \mathrm{pr}_W(v)$ is a normal vector to the plane W in $\mathbb{R}^3$.

We now come to our main result for this section which shows us how to construct an orthonormal basis for V. In some sense, the proof of this theorem is more important that the theorem itself.

Theorem 7: The Gram-Schmidt Orthonormalization Process

Every finite-dimensional inner product space V has an orthonormal basis.

Proof. If V is the zero space, then the empty set is vacuously an orthonormal basis.

Suppose V is nontrivial and let $v_1, \ldots, v_n$ be a basis for V. For $j = 1, \ldots, n$, set $W_j = \mathrm{span}(v_1, \ldots, v_j)$. Then we have $W_1 \subseteq W_2 \subseteq \cdots \subseteq W_n = V$. Now set

$$u_1 = \frac{v_1}{|v_1|}.$$

Hence, u_1 is a unit vector and thus $\{u_1\}$ is an orthonormal basis for W_1. We continue by induction. Suppose we have constructed an orthonormal basis $\{u_1, \ldots, u_k\}$ for W_k for $1 \leq k < n$. Now set

$$u_{k+1} = \frac{v_{k+1} - \mathrm{pr}_{W_k}(v_{k+1})}{|v_{k+1} - \mathrm{pr}_{W_k}(v_{k+1})|}.$$

Now u_{k+1} is orthogonal to W_k, so that $\{u_1, \ldots, u_k, u_{k+1}\}$ is linearly independent, and clearly u_{k+1} is a unit vector. Thus by induction we obtain an orthonormal basis $\{u_1, \ldots, u_n\}$ for W_n. □

Here is a basic example.

Example. Consider the basis $\{v_1, v_2, v_3\}$ of $\mathbb{R}^3$ where

$$v_1 = (2,0,0), \quad v_2 = (1,5,0), \quad v_3 = (1,2,2).$$

Let $W_1 = \text{span}(v_1)$, $W_2 = \text{span}(v_1, v_2)$, and $W_3 = \text{span}(v_1, v_2, v_3)$.

Take $u_1 = v_1/|v_1| = (1,0,0)$. Then $W_1 = \text{span}(u_1)$. Now we construct u_2. First note that

$$\pi_{W_1}(v_2) = \langle v_2 | u_1 \rangle \, u_1 = 1 \cdot (1,0,0).$$

Then

$$u_2 = \frac{v_2 - \pi_{W_1}(v_2)}{|v_2 - \pi_{W_1}(v_2)|} = \frac{(0,5,0)}{|(0,5,0)|} = (0,1,0).$$

Then $W_2 = \text{span}(u_1, u_2)$. Finally we construct u_3. Note that

$$\pi_{W_2}(v_3) = \langle v_3 | u_1 \rangle \, u_1 + \langle v_3 | u_2 \rangle \, u_2 = 1 \cdot (1,0,0) + 2 \cdot (0,1,0) = (1,2,0).$$

Then

$$u_3 = \frac{v_3 - \pi_{W_2}(v_3)}{|v_3 - \pi_{W_2}(v_3)|} = \frac{(0,0,2)}{|(0,0,2)|} = (0,0,1).$$

Hence, $V = W_3 = \text{span}(u_1, u_2, u_3)$. Note that this constructed nothing more than the canonical basis.

Order matters in Gram-Schmidt. Here is the same example but with the order of vectors changed, with some computational support from Maple.

Example. Consider the basis $\{v_1, v_2, v_3\}$ of $\mathbb{R}^3$ where

$$v_1 = (1,2,2), \quad v_2 = (1,5,0), \quad v_3 = (2,0,0).$$

Let $W_1 = \text{span}(v_1)$, $W_2 = \text{span}(v_1, v_2)$, and $W_3 = \text{span}(v_1, v_2, v_3)$.

Take $u_1 = v_1/|v_1| = (1/3, 2/3, 2/3)$, and then $W_1 = \text{span}(u_1)$. Now we construct u_2. First note that

$$\text{pr}_{W_1}(v_2) = \langle v_2 | u_1 \rangle \, u_1 = \frac{11}{3}(1/3, 2/3, 2/3) = (11/9, 22/9, 22/9).$$

Now

$$u_2 = \frac{v_2 - \text{pr}_{W_1}(v_2)}{|v_2 - \text{pr}_{W_1}(v_2)|} = \frac{(-2/9, 23/9/ - 22/9)}{|(-2/9, 23/9/ - 22/9)|} = \left(-\frac{2\sqrt{113}}{339}, \frac{23\sqrt{113}}{339}, -\frac{22\sqrt{113}}{339} \right).$$

Then $W = \text{span}(u_1, u_2)$. Finally, we construct u_3. First,

$$\text{pr}_{W_2}(v_3) = \langle v_3 | u_1 \rangle \, u_1 + \langle v_3 | u_2 \rangle \, u_2 = (26/113, 40/113, 60/113)$$

Now

$$u_3 = \frac{(220/113, -40/113, -60/113)}{|(220/113, -40/113, -60/113)|} = \left(\frac{10\sqrt{113}}{113}, -\frac{2\sqrt{113}}{113}, -\frac{3\sqrt{113}}{113} \right)$$

Then $V = W_3 = \text{span}(u_1, u_2, u_3)$ and $\{u_1, u_2, u_3\}$ is indeed an orthonormal basis for V.

Here is one more example using polynomials instead of $\mathbb{R}^3$.

Example. Let V be the subspace of $\mathscr{C}^0([-1,1])$ spanned by $\{1, x, x^2\}$. We will construct an orthonormal basis for V. Recall that the inner product is given by

$$\langle f|g\rangle = \int_{-1}^{1} f(x)g(x)dx.$$

We note that these vectors are orthogonal to each other

So in the case of 1, we have

$$|1| = \sqrt{\langle 1|1\rangle} = \left(\int_{-1}^{1} (1)^2 dx\right)^{1/2} = \sqrt{2}.$$

Hence, set $u_1 = 1/\sqrt{2}$ and $W_1 = \text{span}(u_1)$. Now we construct u_2. We have

$$\text{pr}_{W_1}(x) = \langle x|u_1\rangle\, u_1 = \frac{1}{2}\int_{-1}^{1} x dx = 0.$$

This result makes sense because 1 and x are already orthogonal. Since

$$|x| = \left(\int_{-1}^{1} x^2\right)^{1/2} = \sqrt{2}/\sqrt{3},$$

then

$$u_2 = \frac{x}{\sqrt{2}/\sqrt{3}} = \frac{\sqrt{6}}{2}x.$$

Set $W_2 = \text{span}(u_1, u_2)$. Finally, we construct u_3. We have

$$\text{pr}_{W_2}(v_3) = \langle v_3|u_1\rangle\, u_1 + \langle v_3|u_2\rangle\, u_2 = \frac{\sqrt{2}}{3}u_1 + 0 = \frac{1}{3}.$$

Then

$$u_3 = \frac{x^2 - (1/3)}{|x^2 - (1/3)|} = \frac{3x^2 - 1)}{|3x^2 - 1|} = \frac{3x^2 - 1}{\sqrt{8}/\sqrt{5}} = \frac{\sqrt{10}}{4}(3x^2 - 1)$$

The basis we have constructed are the so-called *Legendre polynomials.*

We finish this section by discussing the orthogonal complement of a vector subspace. This gives a pleasant way of decomposing a vector space as a direct sum.

Definition: Orthogonal complement

Let V be a finite-dimensional inner product space with subspace W. The *orthogonal complement of* W is defined as

$$W^{\perp} = \{v \in V : \langle v|w\rangle = 0 \text{ for all } w \in W\}.$$

That is, $W^{\perp}$ is the set of vectors in V that are orthogonal to *every* vector in W.

Proposition 8: Orthogonal complements

Let V be a finite-dimensional inner product space with subspace W.

(1) $W^\perp$ is a subspace of W

(2) $W \cap W^\perp = \{\mathbf{0}\}$

(3) For all $v \in V$, there exists unique $w \in W$ and $w^\perp \in W^\perp$ such that $v = w + w^\perp$.

Proof. (1) Since the zero vector is orthogonal to any vector, then $\mathbf{0} \in W^\perp$. Let $v, v' \in W^\perp$ and $\lambda \in k$. Then for any $w \in W$, we have

$$\langle v + v' | w \rangle = \langle v | w \rangle + \langle v | w' \rangle = 0 + 0 = 0$$
$$\langle \lambda v | w \rangle = \lambda \langle v | w \rangle = \lambda 0 = 0.$$

Hence, $v + v', \lambda v \in W^\perp$. Hence, $W^\perp$ is a subspace.

(2) Let $w \in W \cap W^\perp$. Then $\langle w | w \rangle = 0$, so $w = \mathbf{0}$.

(3) Let $v \in V$. We may assume that W has an orthonormal basis by the Gram-Schmidt process. Set $w = \mathrm{pr}_W(v) \in W$. Let $w^\perp = v - w = v - \mathrm{pr}_W(v)$, which is orthogonal to W, so $w^\perp \in W^\perp$. Clearly $v = w + w^\perp$. If $v = w' + (w')^\perp$ with $w' \in W$ and $(w')^\perp \in W^\perp$, then $w + w^\perp = w' + (w')^\perp$ implies $w - w' = (w')^\perp - w^\perp \in W \cap W^\perp = \{\mathbf{0}\}$, so $w = w'$ and $w^\perp = (w')^\perp$. □

In the language of direct sums, we have shown that $V = W \oplus W^\perp$. The uniqueness result above now implies the following.

Corollary 9

The orthogonal projection $\mathrm{pr}_W(v)$ is independent of the choice of orthonormal basis for W.

The following result has some significant applications which we will discuss later.

Corollary 10: Best Approximation Theorem

For all $v \in V$, $\mathrm{pr}_W(v)$ is the point of W closest to V.

Proof. Let $v = w + w^\perp$ as above so that $w = \mathrm{pr}_W(v)$. Then for any $w' \in W$, we have by the Pythagorean Theorem,

$$|w' - v|^2 = |(w' - w) - w^\perp|^2 = |w' - w|^2 + |w^\perp|^2.$$

This is minimal when $w' = w$. □

Determinants and Eigenstuff

The determinant is a fundamental invariant of a matrix. It is useful in a host of problems. We have already seen this in the case of a 2×2 matrix. Recall that for a 2×2 matrix A, $\det(A) \neq 0$ if and only if A is invertible. We will see how to defined a determinant in general and see that this relationship is true for $n \times n$ matrices.

1. Existence and basic properties

Throughout, let k be a field. For $A \in M_n(k)$, $n > 1$, we set

$$\partial_{ij} A \in M_{n-1}(k), \quad 1 \leq i, j \leq n,$$

to be the matrix obtained from A by deleting the ith row and the jth column. So, for example,

$$\partial_{23} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix}.$$

The determinant is defined *recursively*. That is, to find the determinant of an $n \times n$ matrix we first need to know how to compute the determinant of an $(n-1) \times (n-1)$ matrix. For a 1×1 matrix $A = (a)$, we set $\det A = (a)$. Now for an $n \times n$ matrix $A = (a_{ij})$, $n > 1$, we set

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \partial_{1j} A. \tag{1}$$

This is called *cofactor expansion along the first row*. There is nothing special about the first row, but this will suffice for the time being[1]. Later we will see how to expand on any row or column.

Example. (1) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then (1) gives

$$\det(A) = ad - bc.$$

(2) Let $A = \begin{pmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{pmatrix}$. Then (1) gives

$$\begin{aligned} \det(A) &= a_{11}(\partial_{11} A) - a_{12}(\partial_{12} A) + a_{13}(\partial_{13} A) \\ &= 2(-9) - (-4)(-5) + (3)(11) = -5. \end{aligned}$$

These notes are partially derived from *Linear Algebra: An introduction to Abstract Mathematics* by Robert J. Valenza Most of this material is drawn from Chapters 8 and 9. Last Updated: December 2, 2021

[1]Another method for computing this is known as the *butterfly method*. (Not to be confused with the *The Butterfly Effect*, the grand masterpiece of Ashton Kutcher's acting career.) But this method is only useful in the 3×3 case.

The next result shows that the determinant has several important properties. The uniqueness portion of this theorem will be reserved for later.

Theorem 1: The Fundamental Theorem of Determinants

For each $n \geq 1$, there exists a unique map

$$\det : M_n(k) \to k$$

called the *determinant*, satisfying the following rules:

(1) (Multilinearity) Let $A = (A^1, \ldots, A^n)$ and suppose that $A^j = \lambda_1 C_1 + \lambda_2 C_2$ for column vectors $C_1, C_2 \in k^n$ and scalars λ_1, λ_2. Then

$$\det(A) = \lambda_1(\det A^2, \ldots, \underset{\text{col } j}{C_1}, \ldots, A^n) + \lambda_2 \det(A^2, \ldots, \underset{\text{col } j}{C_2}, \ldots, A^n)$$

That is, the determinant is linear in every column.

(2) (Alternation of sign) Suppose that $A = (A^1, \ldots, A^n)$ and that $A^j = A^{j+1}$ for some j, so that two adjacent columns of A are identical, then $\det(A) = 0$.

(3) (Normalization) The determinant of the $n \times n$ identity matrix is the unity of k; that is, $\det(I_n) = 1$ for all n.

Proof. We will show that our definition of the determinant satisfies the above properties by induction. The given properties hold vacuously in the case $n = 1$. Assume they hold for some $n \geq 1$.

(1) We argue for multilinearity in the first column, but the same argument applies to any column. Write $A^1 = \lambda_1 C_1 + \lambda_2 C_2$, so that $a_1 = \lambda_1 c_1 + \lambda_2 c_2$ where c_1, c_2 are the first entries in C_1, C_2, respectively. Now by (1),

$$\det(A) = (\lambda_1 c_1 + \lambda_2 c_2) \det(\partial_{11} A) + \sum_{j=2}^{n+1} (-1)^{j+1} a_{1j} \det(\partial_{1j} A)$$

We apply the inductive hypothesis to the $n \times n$ matrices $\partial_{1j} A$ to obtain

$$\begin{aligned} \det(A) = {} & \lambda_1 c_1 \det(\partial_{11} A) + \lambda_2 c_2 \det(\partial_{11} A) \\ & + \sum_{j=2}^{n+1} (-1)^{j+1} a_{1j} \lambda_1 \det(\partial_{1j}(C_1, A^2, \ldots, A^n)) \\ & + \sum_{j=2}^{n+1} (-1)^{j+1} a_{1j} \lambda_2 \det(\partial_{1j}(C_2, A^2, \ldots, A^n)). \end{aligned}$$

Since $\partial_{11} A$ eliminates the first column of A, then it is clear that

$$\partial_{11} A = \partial_{11}(C_1, A^2, \ldots, A^{n+1}) = \partial_{11}(C_2, A^2, \ldots, A^{n+1}).$$

Combining this observation with our prior computation, we ave

$$\begin{aligned}\det(A) &= \lambda_1 \left(c_1 \partial_{11}(C_1, A^2, \ldots, A^{n+1}) + \sum_{j=2}^{n+1} (-1)^{j+1} a_{1j} \det(\partial_{1j}(C_1, A^2, \ldots, A^{n+1})) \right) \\ &\quad + \lambda_2 \left(c_2 \partial_{11}(C_2, A^2, \ldots, A^{n+1}) + \sum_{j=2}^{n+1} (-1)^{j+1} a_{1j} \det(\partial_{1j}(C_2, A^2, \ldots, A^{n+1})) \right) \\ &= \lambda_1 \det(C_1, A^2, \ldots, A^{n+1}) + \lambda_2 \det(C_2, A^2, \ldots, A^{n+1}),\end{aligned}$$

where the last equality comes from observing that $c_1 \partial_{11}(C_1, A^2, \ldots, A^n)$ is the first term in the expansion of $\det(C_1, A^2, \ldots, A^n)$ (and similarly for c_2, C_2).

(2) Assume $A^1 = A^2$ (the general case follows similarly). Then $a_{11} = a_{12}$ and $\partial_{11} A = \partial_{12} A$ and so

$$\det(A) = a_{11} \det(\partial_{11} A) - a_{12} \det(\partial_{12} A) + \sum_{j=3}^{n+1} (-1)^{j+1} a_{1j} \det(\partial_{1j} A) = \sum_{j=3}^{n+1} (-1)^{j+1} a_{1j} \det(\partial_{1j} A).$$

Now observe that for each $j > 2$, the first and second columns of $\partial_{1j} A$ are identical. Hence, by the inductive hypothesis, $\det(\partial_{1j} A) = 0$. It follows that $\det(A) = 0$.

(3) Expanding on the first row of I_n, along with our inductive hypothesis, immediately gives,

$$\det(I_{n+1}) = 1 \cdot \det(I_n) = 1 \cdot 1 = 1.$$ □

Corollary 2

Let A be an $n \times n$ matrix. If A' is obtained from A by interchange of two columns, then $\det(A') = -\det(A)$. Consequently, if any two columns of A are identical, then $\det(A) = 0$.

Proof. If the first statement holds and two columns of A are identical, then we can swap them to obtain $A' = A$. Hence, $\det(A) = \det(A') = -\det(A)$. Thus $2\det(A) = 0$, so $\det(A) = 0$.

Now we prove the first statement. First, assume that any two adjacent columns of A are identical, say A^1 and A^2 (again, the general case follows similarly). Then $A = (A^1, A^2, A^3, \ldots, A^n)$ and $A' = (A^2, A^1, A^3, \ldots, A^n)$. Let $A'' = (A^1 + A^2, A^1 + A^2, A^3, \ldots, A^n)$. By alternation of sign, $\det(A'') = 0$. By multilinearity,

$$\begin{aligned}0 = \det(A'') &= \det(A^1 + A^2, A^1 + A^2, A^3, \ldots, A^n) \\ &= \det(A^1, A^1 + A^2, A^3, \ldots, A^n) + \det(A^2, A^1 + A^2, A^3, \ldots, A^n) \\ &= \det(A^1, A^1, A^3, \ldots, A^n) + \det(A^1, A^2, A^3, \ldots, A^n) \\ &\quad + \det(A^2, A^1, A^3, \ldots, A^n) + \det(A^2, A^2, A^3, \ldots, A^n) \\ &= 0 + \det(A) + \det(A') + 0.\end{aligned}$$

Finally, recall that any transposition is the product of an odd number of adjacent transpositions. Each adjacent transposition changes the sign by a factor of (-1). $\square$

Corollary 3

Suppose that the columns of A are linearly dependent. Then $\det(A) = 0$.

Proof. Suppose that A^1 is a linear combination of the other columns. That is

$$A^1 = \sum_{j=2}^{n} \lambda_j A^j.$$

Then by multilinearity,

$$\begin{aligned}
\det(A) &= \det(A^1, A^2, \ldots, A^n) \\
&= \det\left(\sum_{j=2}^{n} \lambda_j A^j, A^2, \ldots, A^n\right) \\
&= \sum_{j=2}^{n} \lambda_j \det\left(A^j, A^2, \ldots, A^n\right).
\end{aligned}$$

In each summand, the column A^j is repeated and so the determinant is zero. It follows that the sum is zero. $\square$

2. A nonrecursive formula, uniqueness

In this section, we present new formulas for the determinant and prove uniqueness.

Recall that $\mathcal{S}_n$ denotes the symmetric group on n letters. There is a homomorphism $\sigma : \mathcal{S}_n \to \{\pm 1\}$ called the *sign homomorphism*, where $\sigma(\pi) = 1$ is π is an even permutation and -1 if it is odd.

Theorem 4: Nonrecursive formula for determinant

For all $A \in M_n(k)$,

$$\det(A) = \sum_{\pi \in \mathcal{S}_n} \sigma(\pi) a_{\pi(1)1} \cdots a_{\pi(n)n}.$$

Proof. Regard the canonical basis vectors $\mathbf{e}_j$ as column vectors. Then by multilinearity,

$$\begin{aligned}
\det(A) = \det(A^1, \ldots, A^n) &= \det\left(\sum_{i=1}^n a_{i1}\mathbf{e}_i, A^2, \ldots, A^n\right) \\
&= \sum_{i=1}^n a_{i1} \det\left(\mathbf{e}_i, A^2, \ldots, A^n\right) \\
&= \sum_{i=1}^n a_{i1} \det\left(\mathbf{e}_i, \sum_{j=1}^n a_{j2}\mathbf{e}_j, \ldots, A^n\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n a_{i1}a_{j2} \det\left(\mathbf{e}_i, \mathbf{e}_j, \ldots, A^n\right) \\
&\vdots \\
&= \sum_{\phi} a_{\phi(1)1} \cdots a_{\phi(n)n} \det(\mathbf{e}_{\phi(1)}, \ldots, \mathbf{e}_{\phi(n)})
\end{aligned}$$

where ϕ ranges over all functions $\{1, \ldots, n\} \to \{1, \ldots, n\}$. But by a previous argument,

$$\det(\mathbf{e}_{\phi(1)}, \ldots, \mathbf{e}_{\phi(n)}) = 0 \quad \text{if} \quad \phi(i) = \phi(j) \quad \text{for any } i \neq j.$$

Thus, ϕ is injective. By pigeonhole, ϕ is surjective. That is, the sum ranges over all bijections. □

We note that this formula depended only on the defining properties of the determinant (multilinearity, alteration of sign, normality) and not on the given definition. Hence, any map $M_n(k) \to k$ satisfying these properties must also satisfy this formula. It follows that the determinant formula is unique.

We can now derive several facts about the determinant.

Corollary 5: Determinant of a transpose

The determinant of a matrix is equal to the determinant of its transpose.

Proof. Let $A \in M_n(k)$. We make a replacement of π by π^{-1} and note that as π varies over all of $\mathcal{S}_n$, then so does π^{-1}. Also note $\sigma(\pi) = \sigma(\pi^{-1})$ (exercise). Hence,

$$\begin{aligned}\det(A) &= \sum_{\pi\in\mathcal{S}_n} \sigma(\pi) a_{\pi(1)1}\cdots a_{\pi(n)n} = \sum_{\pi\in\mathcal{S}_n} \sigma(\pi^{-1}) a_{1\pi^{-1}(1)}\cdots a_{n\pi^{-1}(n)} \\ &= \sum_{\pi\in\mathcal{S}_n} \sigma(\pi) a_{1\pi(1)}\cdots a_{n\pi(n)} = \det({}^tA).\end{aligned}$$

□

Previously we defined the determinant in terms of expansion along the first row. However, our proof of the Fundamental Theorem is easily modified so as to work for any row. Thus, by uniqueness, we may obtain the determinant by expanding on *any* row. However, since transpose does not change the determinant, this must also apply to *any* column.

Proposition 6: Expansion by rows and columns

Let $A \in M_n(k)$. Then

(1) For any fixed row index i, $1 \le i \le n$,

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \partial_{ij} A.$$

(2) For any fixed column index j, $1 \le j \le n$,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \partial_{ij} A.$$

Example. Let $A = \begin{pmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{pmatrix}$. Expanding on the second column gives

$$\begin{aligned}\det(A) &= (-1)^{1+2} a_{12}(\partial_{12}A) + (-1)^{2+2} a_{22}(\partial_{22}A) + (-1)^{3+2} a_{32}(\partial_{32}A) \\ &= (-1)(-4)(-5) + (1)(1)(-5) + (-1)(4)(-5) = -5.\end{aligned}$$

Recall that a square matrix $A = (a_{ij})$ is *upper triangular* if $a_{ij} = 0$ for $i > j$ and *lower triangular* if $a_{ij} = 0$ for $i < j$.

Corollary 7: Determinant of triangular matrices

The determinant of a triangular matrix is the product of the triangular entries.

Proof. For a lower triangular matrix, use expansion along the first row and induction. Then note that the transpose of an upper triangular matrix is lower triangular. □

3. The determinant of a product; invertibility

We now establish how invertibility is equivalent to nonzero determinant. This is an easy consequence of the next fact, which is important in its own right.

Theorem 8: Determinant of a product

Let $A, B \in M_n(k)$. Then

$$\det(AB) = \det(A)\det(B).$$

Proof. Recall that $AB = (A \cdot B^1, \dots, AB^j \cdots, AB^n)$ and

$$AB^j = b_{1j}A^1 + \cdots + b_{nj}A^n.$$

Hence, by multilinearity,

$$\det(AB) = \sum_{\phi} b_{\phi(1)1} \cdots b_{\phi(n)n} \det(A^{\phi(1)}, \dots, A^{\phi(n)})$$

where the sum is taken over all function $\phi : \{1, \dots, n\} \to \{1, \dots, n\}$. As before, the determinant in the summand will be zero unless ϕ is bijective. Thus,

$$\begin{aligned}
\det(AB) &= \sum_{\pi \in \mathcal{S}_n} b_{\pi(1)1} \cdots b_{\pi(n)n} \det(A^{\pi(1)}, \dots, A^{\pi(n)}) \\
&= \sum_{\pi \in \mathcal{S}_n} b_{\pi(1)1} \cdots b_{\pi(n)n} \left(\sigma(\pi) \det(A^1, \dots, A^n)\right) \\
&= \det(A) \sum_{\pi \in \mathcal{S}_n} \sigma(\pi) b_{\pi(1)1} \cdots b_{\pi(n)n} = \det(A)\det(B).
\end{aligned}$$

□

It now follows that the determinant is a group homomorphism $\det : \mathrm{GL}_n(k) \to k^*$. The kernel of this map is precisely the subgroup of determinant 1 matrices, called the *special linear group*.

We can now make some additions to the Invertible Matrix Theorem.

Theorem 9: The Invertible Matrix Theorem (version 2)

Let $A \in M_n(k)$. The following are equivalent:

(1) The linear system $A\mathbf{x} = \mathbf{y}$ has at least one solution for all $\mathbf{y} \in k^n$.
(2) The columns of A span k^n.
(3) The rows of A span k^n.
(4) The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(5) The columns of A are linearly independent.
(6) The rows of A are linearly independent.
(7) The linear system $A\mathbf{x} = \mathbf{y}$ has exactly one solution for all $\mathbf{y} \in k^n$.
(8) The columns of A constitute a basis for k^n.
(9) The rows of A constitute a basis for k^n.
(10) The matrix A is invertible (i.e., $A \in \mathrm{GL}_n(k)$).
(11) The determinant of A is nonzero.
(12) The determinant of ${}^tA$ is nonzero.

Proof. We have already established many of these equivalences. Namely, 1,2,4,5,7, and 8. The statements 3, 6, and 9 now follow because $\mathrm{rk}(A) = \mathrm{rk}({}^tA)$. Moreover, we showed that 10 is equivalent by using duality to establish a two-sided inverse for A. It is clear that 11 and 12 are equivalent to each other. It remains only to show that 11 is equivalent to the other statements.

If A is invertible with inverse B, then

$$\det(A)\det(B) = \det(AB) = \det(I_n) = 1,$$

so clearly $\det(A) \neq 0$. On the other hand, if A is not invertible, then the columns of A must be linearly dependent and so $\det(A) = 0$. □

9.1 Eigenstuff - Definitions and elementary properties

This is one of the most important topics in terms of applications. However, much of the computation will seem very routine at this point.

Definition: Eigenvalue, eigenvector

Let $T : V \to V$ be a linear transformation on a vector space V over k, and suppose that for some $\lambda \in k$ there exists a nonzero $v \in V$ such that

$$T(v) = \lambda v.$$

Then λ is called an *eigenvalue* of T, and every vector v (including $\mathbf{0}$) that satisfies the equation above is called an *eigenvector belonging to* λ.

Note that while $\mathbf{0}$ is considered an eigenvector for λ, in order for λ to be an eigenvalue in the first place, there must exist a *nonzero* eigenvector. Otherwise, every scalar would be an eigenvalue with the eigenvector $\mathbf{0}$.

Example. (1) Let $V = \mathbb{R}^2$ and let $T_A : V \to V$ be the map defined by multiplication by the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since

$$T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad T\begin{pmatrix} 1 \\ -1 \end{pmatrix} = (-1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

then ± 1 are eigenvalues of T. We will see an algorithm to compute the eigenvalues for a general map of the form T_A.

(2) Let $\mathscr{C}^\infty(\mathbb{R})$ denote the vector space of infinitely differentiable functions $\mathbb{R} \to \mathbb{R}$. Since for $\lambda \in \mathbb{R}$, the differentiation map gives $D(e^{\lambda x}) = \lambda e^{\lambda x}$, then every $\lambda \in \mathbb{R}$ is an eigenvalue for D with eigenvectors $Ce^{\lambda x}$.

Given an eigenvalue λ for a linear transformation $T : V \to V$, the set of eigenvectors belonging to λ,

$$E_\lambda = \{v \in V : T(\lambda) = \lambda v\}$$

is a subspace of V (exercise) called the *eigenspace corresponding to* λ. A basis of V consisting of eigenvectors for some T is called an *eigenbasis of* V *with respect to* T. Not all vector spaces have an eigenbasis for a particular T. When they do, T takes a simple form on this eigenbasis.

Definition: Diagonalizable transformation

A linear transformation $T : V \to V$ is *diagonalizable* if it is representable by a diagonal matrix.

The next theorem characterizes diagonalizable transformations.

Theorem 10: Diagonalizable transformation

Let $T : V \to V$ be a linear transformation on a finite-dimensional vector space. Then T is diagonalizable if and only if there exists an eigenbasis for V with respect to T. In this case, the diagonal entries of the standard matrix are precisely the eigenvalues of T.

Proof. Suppose that $B = \{v_1, \ldots, v_n\}$ is a basis of V consisting of eigenvectors with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $T(v_j) = \lambda_j v_j$ for all j and so the standard matrix of T is diagonal with respect to B, with diagonal entries $\lambda_1, \ldots, \lambda_n$.

Conversely, if the standard matrix of T is diagonal with respect to some some basis. Then clearly each basis vector is an eigenvector for T. Thus, this basis is an eigenbasis. □

We freely refer to the eigenvalues and eigenvectors of a (standard) matrix in place of the corresponding linear transformation.

Corollary 11: Diagonalizable matrices

A matrix $A \in M_n(k)$ is similar to a diagonal matrix if and only if there exists an eigenbasis for k^n with respect to T_A.

Proof. Recall simply that if B and B' are basis for the finite-dimensional vector space V, and $T : V \to V$ is a linear transformation, then $M_{B'}(T) = P^{-1} M_B(T) P$ where P is the transition matrix from B' to B. □

Example. (1) In our previous example (1), V has eigenbasis $(1, 1)$ and $(1, -1)$. Note also that these vectors are orthogonal.

(2) Consider the linear transformation corresponding to rotation counterclockwise in $\mathbb{R}^2$ by the angle θ. This is the map $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If θ is not a scalar multiple of π, then no nonzero vector can be rotated to a scalar multiplication of itself. Hence, in this case T_θ is not diagonalizable.

The algorithm for finding eigenvalues is surprisingly simple.

Definition: Characteristic polynomial

Let $A \in M_n(k)$ and let t be an indeterminate. Then

$$p(t) = \det(tI_n - A)$$

is called the *characteristic polynomial of* A.

Note that the characteristic polynomial has degree n.

Theorem 12: Eigenvalues of A

The eigenvalues of $A \in M_n(k)$ are the roots of the characteristic polynomial of A.

Proof. Let $\lambda \in k$. Then $p(\lambda) = \det(\lambda I_n - A)$. Then $\det(\lambda I_n - A) = 0$ if and only if $\lambda I_n - A$ is singular (non-invertible). That is, the null space of $\lambda I_n - A$ is nontrivial. That is, there exists $\mathbf{x} \in k^n$ such that $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$. Equivalently, $A\mathbf{x} = \lambda\mathbf{x}$ for some $\mathbf{x} \neq 0$. □

To find the eigenvalues of a matrix A we need to find the roots of the characteristic polynomial. The eigenvectors for a particular eigenvalue λ are just the elements of the null space of $\lambda I_n - A$.

Example. (1) Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then the characteristic polynonmial is

$$p(t) = (tI_2 - A) = \det \begin{pmatrix} \lambda & -1 \\ -1 & \lambda \end{pmatrix} = \lambda^2 - 1.$$

Hence, the roots/eigenvalues are ± 1.

(2) Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then $p(t) = t^2 + 1$, which has no real roots and hence no real eigenvalues. It does, however, have two complex eigenvalues $\pm i$.

(3) Consider $A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix}$. Then the characteristic polynomial is

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{pmatrix} \lambda - 1 & 0 & -2 \\ -3 & \lambda + 1 & -3 \\ -2 & 0 & \lambda - 1 \end{pmatrix} \\ &= (\lambda - 1)((\lambda + 1)(\lambda - 1) - 0) - 2(0 + 2(\lambda + 1)) \\ &= (\lambda + 1)(\lambda^2 - 2\lambda - 3) = (\lambda + 1)^2(\lambda - 3). \end{aligned}$$

Hence, the eigenvalues are -1 and 3. Now we find a basis of each eigenspace:

$$E_{-1}(A) : (-1)I - A = \begin{pmatrix} -2 & 0 & -2 \\ -3 & 0 & -3 \\ -2 & 0 & -2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{basis of } E_{-1}\text{: } \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$$E_3(A) : 3I - A = \begin{pmatrix} 2 & 0 & -2 \\ -3 & 3 & -3 \\ -2 & 0 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -3/2 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{basis of } E_3\text{: } \left\{ \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\}.$$

The eigenvalues do depend on the field, but they do not depend on the choice of basis.

Proposition 13: Similar matrices and eigenvalues

If two matrices are similar, then they have the same characteristic polynomial.

Proof. Suppose $A \sim B$, so $B = P^{-1}AP$ for some $P \in \mathrm{GL}_n(k)$. Then

$$\begin{aligned} \det(tI_n - B) &= \det(tP^{-1}I_nP - P^{-1}AP) = \det(P^{-1}) \det(tI_n - A) \det(P) \\ &= \det(P)^{-1} \det(P) \det(tI_n - A) = \det(tI_n - A). \end{aligned}$$

□

It follows from this proposition that similar matrices have the same eigenvalues. We now see one case of how to form an eigenbasis.

Proposition 14: Linear independence of eigenvectors

Let $\lambda_1, \ldots, \lambda_r$ be distinct eigenvalues of $T : V \to V$ with corresponding nonzero eigenvectors $v_1, \ldots, v_r$. Then $v_1, \ldots, v_r$ are linearly independent.

Proof. Suppose $v_1, \ldots, v_r$ are linearly dependent, so there is some dependence relation among them. After reordering, we may assume v_1 is a linear combination of $v_2, \ldots, v_s$, where $v_2, \ldots, v_s$ are linearly independent. That is, there are nonzero scalars μ_i such that

$$v_1 = \mu_2 v_2 + \cdots + \mu_s v_s.$$

Applying T gives,

$$\begin{aligned} T(v_1) &= \mu_2 T(v_2) + \cdots + \mu_s T(v_s) \\ \lambda_1 v_1 &= \mu_2 \lambda_2 v_2 + \cdots + \mu_s \lambda_s v_s \\ \lambda_1(\mu_2 v_2 + \cdots + \mu_s v_s) &= \mu_2 \lambda_2 v_2 + \cdots + \mu_s \lambda_s v_s \\ 0 &= \mu_2(\lambda_2 - \lambda_1)v_2 + \cdots + \mu_s(\lambda_s - \lambda_1)v_s. \end{aligned}$$

Since the μ_i are nonzero and the v_i are linearly independent, then $\lambda_i = \lambda_1$ for some i, a contradiction.

□

Theorem 15: Diagonalization

Let $A \in M_n(k)$ and assume that the characteristic polynomial of A has n distinct roots. Then A is diagonalizable.

Proof. Each root of the the characteristic polynomial corresponds to a (distinct) eigenvalue, and each eigenvalue has a corresponding nonzero eigenvector. The eigenvectors corresponding to distinct eigenvalues form a linearly independent set. By pigeonhole, this set is a basis for k^n. □

In general, to diagonalize T, we need n linearly independent eigenvectors. Eigenvectors corresponding to distinct eigenspaces are linearly independent, as we saw above. Within a given eigenspace, the question becomes whether or not there are *enough* vectors in the eigenspace.

Definition: Algebraic and geometric multiplicty

Let λ be an eigenvalue of a matrix $A \in M_n(k)$. The *algebraic multiplicity* of λ is the multiplicity of λ as a root of the characteristic polynomial. The *geometric multiplicity* is the dimension of the eigenspace corresponding to λ.

Example. (1) Consider $A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix}$.

The algebraic and geometric multiplicity of -1 is two, while the algebraic and geometric multiplicity of 3 is one. Thus, A is diagonalizable. Let B be the standard basis of $\mathbb{R}^n$, the eigenbasis is

$$B' = \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} \right\}.$$

The transition matrix from B' to B is then

$$P = \begin{pmatrix} 0 & -1 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 2 \end{pmatrix}$$

Hence, $D = P^{-1}AP$ where $D = \text{diag}(-1, -1, 3)$.

(2) Consider $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Because A is upper triangular, the only eigenvalue is 1. In fact, the characteristic polynomial is $(t-1)^2$, so 1 has algebraic multiplicity two. Now the null space has basis $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$, so 1 has geometric multiplicity 1. Hence, A is not diagonalizable.

Lemma 16: Multiplicity

Let $A \in M_n(k)$ with eigenvalue λ. The geometric multiplicity of λ is no more than the algebraic multiplicity of λ.

Proof. Let λ be an eigenvalue of A with geometric multiplicity m and let $\{v_1, \ldots, v_m\}$ be a basis for $E_\lambda(A)$. We extend this to a basis $\{v_1, \ldots, v_n\}$ of k^n. Since $v_{m+1}, \ldots, v_n$ are not elements of $E_\lambda(A)$, then they are not eigenvectors for λ. Let $P = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$, so P is invertible by the IMT. Set $B = P^{-1}AP$. Then B is a block matrix of the form

$$B = \left(\begin{array}{cccc|c} \lambda & 0 & \cdots & 0 & \\ 0 & \lambda & \cdots & 0 & \\ \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \lambda & \\ \hline \multicolumn{4}{c|}{0} & \star \end{array}\right)$$

Since B is similar to A, they have the same characteristic equation and hence the same algebraic multiplicity for each eigenvalue. It is clear that the algebraic multiplicity of λ in B is at least m. $\square$

The issue here is that the elements $v_{m+1}, \ldots, v_n$ may not be eigenvectors at all. Thus, we arrive at our main theorem on diagonalization.

Theorem 17: The Diagonalization Theorem

Let $A \in M_n(k)$. Then A is diagonalizable if and only if for each eigenvalue λ, the algebraic multiplicity and geometric multiplicity of λ are equal.

Proof. The statement on equality of multiplicities is equivalent to k^n having an eigenbasis with respect to A, which is equivalent to A having n linearly independent eigenvectors. $\square$

There is one case where a matrix is *always* diagonalizable, as we discuss in the next section.

4. Orthogonal diagonalization of symmetric matrices

In this section we assume $k = \mathbb{R}$. There is a similar theory over $\mathbb{C}$ but we do not discuss it here. Recall that for $u, v \in k^n = \mathbb{R}^n$, the inner product (dot product) is given by $\langle u|v\rangle = {}^t uv$.

Definition: Orthogonal matrix

A matrix $Q \in M_{m,n}(k)$ is *orthogonal* if its columns form an orthonormal set.

There are several equivalent ways to view orthogonal matrices.

Theorem 18: Alternate characterizations of orthogonal matrices

Let $Q \in M_n(k)$. The following statements are equivalent:

(1) Q is orthogonal.
(2) $|Qx| = |x|$ for all $x \in \mathbb{R}^n$.
(3) $\langle Qx|Qy\rangle = \langle x|y\rangle$ for all $x, y \in \mathbb{R}^n$.
(4) ${}^tQQ = I$.

Proof. (1) $\Rightarrow$ (2) Suppose Q is orthogonal. Write $Q = \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix}$. Then

$$\begin{aligned} |Qx|^2 = \langle Qx|Qx\rangle &= \langle x_1q_1 + \cdots x_nq_n | x_1q_1 + \cdots x_nq_n\rangle \\ &= \sum_{i,j} \langle x_iq_i|x_jq_j\rangle = \sum_{i,j} x_ix_j \langle q_i|q_j\rangle = \sum_i x_i^2 \langle q_i|q_j\rangle = \sum_i x_i^2 = |x|^2. \end{aligned}$$

(2) $\Rightarrow$ (3) We have

$$\begin{aligned} {}^txy &= \frac{1}{4}\left(|x+y|^2 - |x-y|^2\right) = \frac{1}{4}\left(|Q(x+y)|^2 - |Q(x-y)|^2\right) \\ &= \frac{1}{4}\left(|Qx+Qy|^2 - |Qx-Qy|^2\right) = \langle Qx|Qy\rangle . \end{aligned}$$

(3) $\Rightarrow$ (4) We have

$${}^tQQ = \begin{pmatrix} {}^tq_1 \\ \cdots \\ {}^tq_n \end{pmatrix} \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix} = (\langle q_i|q_j\rangle) = (\langle Qe_i|Qe_j\rangle) = (\langle e_i|e_j\rangle) = (\delta_{ij}).$$

Hence, (4) holds.

(4) $\Rightarrow$ (1) As in the previous part, $I = {}^tQQ = (\langle q_i|q_j\rangle)$. Hence, the columns of Q are orthonormal. □

Property (2) above says that orthogonal matrices are *length preserving*. On the other hand, property (4) implies that a square orthogonal matrix Q is invertible and $Q^{-1} = {}^tQ$.

Example. The following matrices are easily seen to be orthogonal:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Hence,

$$A^{-1} = {}^tA = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad B^{-1} = {}^tB = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Both matrices in the previous example are examples of an *isometry*, a length-preserving transformation in $\mathbb{R}^3$. In fact, this length-preserving property characterizes orthogonal matrices.

Definition: Orthogonally diagonlizable

An square matrix A is *orthogonally diagonalizable* if there exists an orthogonal matrix Q and a diagonal matrix D such that ${}^tQAQ = D$.

Example. The eigenvalues of $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ are $\lambda_1 = -2$, $\lambda_2 = 4$ with corresponding eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Hence, A is diagonalizable. Set $P = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$, then $P^{-1}AP = D$ where $D = \begin{pmatrix} -2 & 0 \\ 0 & 4 \end{pmatrix}$.

Note that v_1 and v_2 are orthogonal. Normalizing these vectors we have

$$u_1 = \frac{1}{|v_1|}v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1\sqrt{2} \end{pmatrix}, \qquad u_2 = \frac{1}{|v_1|}v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1\sqrt{2} \end{pmatrix}.$$

Set $Q = \begin{pmatrix} u_1 & u_2 \end{pmatrix}$. Since Q is an orthogonal matrix, then $Q^{-1} = {}^tQ$, so we have ${}^tQAQ = D$. This is adventageous since computing the transpose is much easier than computing the inverse.

The next theorem is easy to prove. Its converse is also true, but this will take more work to prove.

Proposition 19: Orthogonally diagonalizable implies symmetric

If A is orthogonally diagonalizable, then A is symmetric.

Proof. Since A is orthogonally diagonalizable, then ${}^tQAQ = D$ for some orthogonal matrix Q and diagonal matrix D. But then $A = ({}^tQ)^{-1}D\,{}^tQ = QD\,{}^tQ$. Hence,

$${}^tA = (QD\,{}^tQ) = {}^t({}^tQ)\,{}^tD\,{}^tQ = QD\,{}^tQ = A. \qquad \square$$

We write complex numbers in $\mathbb{C}$ as $a+bi$ where a and b are real numbers and $i=\sqrt{-1}$. The *complex conjugate* of $z=a+bi$ is $\overline{z}=\overline{a+bi}=a-bi$. We can extend this notion to complex matrices. The conjugate of a matrix $A=(a_{ij})$ with complex entries is $\overline{A}=(\overline{a_{ij}})$.

The set of eigenvalues of a matrix A is called the *spectrum of* A and is denoted σ_A.

Theorem 20: The (real) Spectral Theorem

Let $A\in M_n(\mathbb{R})$ be symmetric.

(1) The eigenvalues of A are real.

(2) Any two eigenvectors from different eigenspaces of A are orthogonal.

(3) A is orthogonally diagonalizable.

Proof. (1) Suppose λ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}$. Then

$$A\overline{v}=\overline{A}\overline{v}=\overline{Av}=\overline{\lambda v}=\overline{\lambda}\overline{v}.$$

Using the fact that A is symmetric we have

$${}^t\overline{v}A={}^t(A\overline{v})={}^t(\overline{\lambda}\overline{v})=\overline{\lambda}\,{}^t\overline{v}.$$

Therefore,

$$\lambda({}^t\overline{v}v)={}^t\overline{v}(\lambda v)={}^t\overline{v}(Av)=({}^t\overline{v}A)v=(\overline{\lambda}\,{}^t\overline{v})v=\overline{\lambda}({}^t\overline{v}v).$$

This shows that $(\lambda-\overline{\lambda})({}^t\overline{v}v)=0$. Write

$$v=\begin{pmatrix}a_1+b_1i\\ \vdots\\ a_n+b_ni\end{pmatrix}\quad\text{so}\quad\overline{v}=\begin{pmatrix}a_1-b_1i\\ \vdots\\ a_n-b_ni\end{pmatrix}.$$

Then

$${}^t\overline{v}v=\langle\overline{v}|\overline{v}\rangle=(a_1^2+b_1)^2+\cdots+(a_n^2+b_n^2)\neq 0$$

because $v\neq 0$. We conclude that $\lambda-\overline{\lambda}=0$. That is $\lambda=\overline{\lambda}$ so λ is real.

(2) Let v_1,v_2 be eigenvectors for A with corresponding eigenvalues λ_1,λ_2, $\lambda_1\neq\lambda_2$. Then

$$\lambda_1(\langle v_1|v_2\rangle)={}^t(\lambda_1v_1)v_2={}^t(Av_1)v_2={}^tv_1\,{}^tAv_2={}^tv_1Av_2={}^tv_1(\lambda_2v_2)=\lambda_2(\langle v_1|v_2\rangle).$$

Hence, $(\lambda_1-\lambda_2)(\langle v_1|v_2\rangle)=0$. Since $\lambda_1\neq\lambda_2$, then we must have $\langle v_1|v_2\rangle=0$.

(3) Clearly the result holds for all 1×1 matrices. Assume all $(n-1)\times(n-1)$ symmetric matrices are orthogonally diagonalizable. Let A be $n\times n$, let λ_1 be an eigenvalue of A, and let u_1 a (unit) eigenvector for λ_1. Set $W=\text{span}\{u_1\}$.

By Gram-Schmidt, we may extend u_1 to an orthonormal basis $\{u_1, \ldots, u_n\}$ for $\mathbb{R}^n$ where $\{u_2, \ldots, u_n\}$ is a basis for $W^\perp$. Set $Q_1 = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$. Then

$$ {}^tQ_1AQ_1 = \begin{pmatrix} {}^tu_1Au_1 & \cdots & {}^tu_1Au_n \\ \vdots & \ddots & \vdots \\ {}^tu_nAu_1 & \cdots & {}^tu_nAu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & * \\ 0 & B \end{pmatrix}. $$

The first column is as indicated because ${}^tu_iAu_1 = {}^tu_i(\lambda u_1) = \lambda({}^tu_iu_1) = \lambda\delta_{i1}$. As ${}^tQ_1AQ_1$ is symmetric, $* = 0$ and B is a symmetric $(n-1)\times(n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \ldots, \lambda_n$ (by the inductive hypothesis). Because A and ${}^tQ_1AQ_1$ are similar, then the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q_2 such that ${}^tQ_2BQ_2 = D$, where the diagonal entries of D are $\lambda_2, \ldots, \lambda_n$. Now

$$ {}^t\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & {}^tQ_2BQ_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & D \end{pmatrix}. $$

Note that $\begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix}$ is orthogonal. Set $Q = Q_1 \begin{pmatrix} 1 & 0 \\ 0 & Q_2 \end{pmatrix}$. As the product of orthogonal matrices is orthogonal, Q is itself orthogonal and ${}^tQAQ$ is diagonal. □

Example. Let $A = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$. The eigenvalues of A are -2 and 7 with eigenspace bases:

$$ E_7 : \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix} \right\}, \qquad E_{-2} : \left\{ \begin{pmatrix} -1 \\ -1/2 \\ 1 \end{pmatrix} \right\}. $$

It is easy to verify that the basis vector for E_{-2} is orthogonal to those of E_7. However, the two basis vectors for E_7 *are not* orthogonal. In order to orthogonally diagonalize A, we need an orthogonal basis for E_7. To do this, we use Gram-Schmidt:

$$ u_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \quad u_2 = \begin{pmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{pmatrix}, \quad u_3 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}. $$

Now $Q = \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix}$ is orthogonal and ${}^tQAQ = \operatorname{diag}(7, 7, -2)$.

Let A be orthogonally diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $A = QD\,{}^tQ$ with $Q = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ and and $D = \operatorname{diag}\lambda_1, \ldots, \lambda_n$. Then

$$ A = QD\,{}^tQ = \lambda_1 u_1\,{}^tu_1 + \cdots + \lambda_n u_n\,{}^tu_n. $$

This is known as the *spectral decomposition of A*, or *projection form of the Spectral Theorem.* Each ${u_i}^t u_i$ is called a *projection matrix* because $({u_i}^t u_i)x$ is the projection of x onto $\text{span}\{u_i\}$.

Example 21. Let $A = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$. An orthonormal basis of the column space is given above. Setting Q and D as above, the projection matrices are

$$u_1{}^t u_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, u_2{}^t u_2 = \frac{1}{18}\begin{pmatrix} 1 & -4 & -1 \\ -4 & 16 & 4 \\ -1 & 4 & 1 \end{pmatrix}, u_3{}^t u_3 = \frac{1}{9}\begin{pmatrix} 4 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 4 \end{pmatrix}.$$

The spectral decomposition is

$$A = 7u_1{}^t u_1 + 7u_2{}^t u_2 - 2u_3{}^t u_3.$$