Christopher Heil

Lecture Notes for MATH 2406 (Abstract Vector Spaces)

April 4, 2013

3

Linear Spaces

3.1 Introduction

Now we hook up with the text by Apostol. Chapters 1 and 2 of Apostol review some facts about linear algebra in $\mathbf{R}^n$. You should have seen much of that material (though perhaps not all) when you took MATH 1502 (Calculus II) or MATH 1522 (Linear Algebra for Calculus). It would be a good idea for you to skim through Chapters 1 and 2 of Apostol now, but don't worry about reading every tiny detail. Our course is MATH 2406, whose title is *Abstract Vector Spaces*, so we are going to be doing linear algebra in all kinds of abstract spaces, not just in $\mathbf{R}^n$. There will be a lot of similarities to how things work in $\mathbf{R}^n$, but also a lot of interesting differences. The notes below will parallel the sections of Chapter 3 of Apostol's text but will also include some extra discussion and exercises.

3.2 Axiomatic Definition of a Vector Space

There are many different types of objects that we work with—numbers, functions, functions whose domain is a set of functions, and so on and so forth. And yet, when we try to prove things in these various and seemingly different contexts, we often find a lot of similarities. For example, we can add numbers, and we can also add functions, and these addition operations have a lot of properties in common. If we can see the commonalities, then we can prove facts that hold not just in one setting, but in many different settings. That is one reason why we introduce "abstract" definitions, like the following definition of a vector space.

Definition 3.1. Let V be a nonempty set. We will call the elements of V *vectors* (regardless of whether they are numbers, functions, apples, operators, whatever). The set V is a *vector space* if the following 10 axioms are simultaneously satisfied.

Closure Axioms

(1) Vector addition: There has to be an operation that takes two vectors x, $y \in V$ and gives us another vector in V. We call this operation *vector addition* (even though it might not have anything to do with "addition" in the usual sense), and we write $x+y$ for the vector that we get when we apply the operation to x and y. In other words,

$$\forall\, x, y \in V, \quad \exists \text{ a unique element } x+y \in V.$$

(2) Scalar multiplication: There has to be another operation that takes a vector $x \in V$ and a real number $c \in \mathbf{R}$ and gives us another vector in V. We call this operation *scalar multiplication*, and we write cx for the vector that we get when we apply this operation to x and c. In other words,

$$\forall\, x \in V, \quad \forall\, c \in \mathbf{C}, \quad \exists \text{ a unique element } cx \in V.$$

In the context of vector spaces, we call $\mathbf{R}$ the *set of scalars* or *field of scalars*, and we refer to a real number $c \in \mathbf{R}$ as a *scalar*.

Addition Axioms

(3) Commutativity: We must have $x+y = y+x$ for all x, $y \in V$.

(4) Associativity: We must have $(x+y)+z = x+(y+z)$ for all x, $y \in V$.

(5) Additive Identity: There must exist an element $0 \in V$ that satisfies

$$x + 0 \;=\; x \qquad \text{for all } x \in V.$$

We call this vector 0 the *zero element* or the *zero vector* for the vector space.

(6) Additive Inverses: For each vector $x \in V$, there must be a vector $(-x) \in V$ that satisfies

$$x + (-x) \;=\; 0.$$

Multiplication Axioms

(7) Associativity: We must have $(ab)x = a(bx)$ for all a, $b \in \mathbf{R}$ and $x \in V$.

(8) Multiplicative Identity: Scalar multiplication by the number 1 must satisfy $1x = x$ for every $x \in V$.

Distributive Axioms

(9) We must have

$$c(x+y) \;=\; cx + cy, \qquad \text{for all } x, y \in V \text{ and } c \in \mathbf{R}.$$

(10) We must have

$$(a+b)x \;=\; ax + bx, \qquad \text{for all } x \in V \text{ and } a, b \in \mathbf{R}.$$

We call V a vector if *all ten* of these axioms are satisfied. If any one axiom fails, then V is not called a vector space. ◇

Although Apostol prefers the term "linear space," I like "vector space" better so I will use that term in these notes. However, these two terms are entirely equivalent and mean exactly the same thing.

3.3 Examples of Vector Spaces

Example 3.2. You should check that $\mathbf{R}^n$ is a vector space. The operation of vector addition on $\mathbf{R}^n$ is simply componentwise addition, i.e.,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Scalar multiplication is also defined componentwise, i.e.,

$$c \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}.$$

The zero element of $\mathbf{R}^n$ (i.e., the additive identity) is the vector

$$0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that we use the same symbol 0 to denote both the *zero vector* and the *number zero*. You have to tell by context what 0 is supposed to mean. What is the additive inverse of a vector x? ♢

Exercise 3.3. We will use this example a lot. Consider the set of all functions that map real numbers to real numbers. We've seen this set before, and we called it

$$\mathcal{F}(\mathbf{R}) = \{f \;:\; f\colon \mathbf{R} \to \mathbf{R}\}.$$

For example, the function f whose rule is $f(x) = \sin x$ is one vector in this space. Every function whose domain and codomain is the real line $\mathbf{R}$ is a vector in this space. You should think of each such function as being one "point" in the set $\mathcal{F}(\mathbf{R})$.

There are many operations that we can perform on functions. For example, we know how to add functions, multiply a function by a scalar, multiply two functions together, compose two functions, and so forth. In order to make $\mathcal{F}(\mathbf{R})$ into a vector space, we are going to focus on just two of these operations. The first is addition of functions; this will be our *vector addition* in this space.

Vector addition in $\mathcal{F}(\mathbf{R})$

Let f and g be two vectors in $\mathcal{F}(\mathbf{R})$. That this just means that $f, g \in \mathcal{F}(\mathbf{R})$, i.e., we have two functions f and g that map real numbers to real numbers. We define *vector addition* by declaring that $h = f + g$ is the function whose rule is

$$h(x) \;=\; (f+g)(x) \;=\; f(x)+g(x), \qquad x \in \mathbf{R}.$$

Scalar multiplication in $\mathcal{F}(\mathbf{R})$

Let f be a vector in $\mathcal{F}(\mathbf{R})$, and let c be a scalar. This just means that $f \in \mathcal{F}(\mathbf{R})$ and $c \in \mathbf{R}$. We define *scalar multiplication* by declaring that $k = cf$ is the function whose rule is

$$k(x) \;=\; (cf)(x) \;=\; cf(x), \qquad x \in \mathbf{R}.$$

Verify that $\mathcal{F}(\mathbf{R})$ is a vector space

Now that we have defined the operations on $\mathcal{F}(\mathbf{R})$, the point of this exercise is to show that all ten of the vector space axioms are satisfied. When you do this, be careful to distinguish between elements of $\mathcal{F}(\mathbf{R})$, which are *functions*, and the value $f(x)$, which is the *number* that is output by f at the input x. For example, to prove associativity, we must show that if f, g, h are any three vectors (functions) in $\mathcal{F}(\mathbf{R})$, then $f+(g+h) = (f+g)+h$. How do you show that the function $f + (g + h)$ equals the function $(f + g) + h$? You have to show that they have the same *rule*. So, you let x be any real number, and then you have to verify that $(f + (g + h))(x)$ and $(f + g) + h)(x)$ are the same *real number*. Now, you get to use the fact that these are numbers, and we know that *addition of numbers* is associative. We can argue as follows:

$$\begin{aligned}
(f+(g+h))(x) &= f(x)+(g+h)(x) && \text{definition of vector addition}\\
&= f(x)+(g(x)+h(x)) && \text{definition of vector addition}\\
&= (f(x)+g(x))+h(x) && \text{associativity of NUMBERS}\\
&= (f+g)(x)+h(x) && \text{definition of vector addition}\\
&= ((f+g)+h)(x) && \text{definition of vector addition.}
\end{aligned}$$

This shows that $f + (g + h)$ and $(f + g) + h$ output the same value for every input x, so they have the same rule and therefore are the *same function*.

Now, that only takes care of one of the axioms, you still have to verify that all of the other axioms hold. Let's look at the additive identity axiom. We need to show that there is a vector 0 in $\mathcal{F}(\mathbf{R})$ that satisfies $0 + f = f$ for every $f \in \mathcal{F}(\mathbf{R})$. This *vector* 0 has to be a *function* in $\mathcal{F}(\mathbf{R})$. We use the same symbol for the function 0 and the number 0, so you have to use context to tell what is meant. The *function* 0 is the function whose rule is

$$0(x) \;=\; 0, \qquad x \in \mathbf{R}.$$

The 0 on the left-hand side of the equation on the preceding line represents the *function* 0, while the 0 on the right-hand side represents the *number* 0. Now you need to verify that if f is any vector in $\mathcal{F}(\mathbf{R})$, then $f + 0 = f$. To do this, you have to check that $f + 0$ and f have the *same rule*, so you have to show that $(f+0)(x)$ and $f(x)$ are equal for every x. Do it! And also verify that all of the other axioms hold! ♢

Exercise 3.4. Even though we refer to the vector addition operation as "addition" and we represent it with the symbol +, the actual vector addition operation on a given vector space doesn't have to have anything to do with "addition" as such. I'll try to give an example where the operation isn't an "addition," although I have to admit that this example isn't perfect because we're not going to end up with a vector space. But the lesson to take away is that the vector addition operation can be *any kind* of operation—as long as it satisfies the axioms in the definition of a vector space.

For this example, we keep $V = \mathcal{F}(\mathbf{R})$, but instead of using addition of functions as our vector addition, we're going to use *composition.* That is, we *declare* that *for this example* $f + g$ means the *composition* of f and g, i.e.,

$$f + g \;=\; f \circ g, \qquad f, g \in \mathcal{F}(\mathbf{R}).$$

On the other hand, we will leave scalar multiplication alone, so cf still means the function whose rule is $(cf)(x) = cf(x)$ for $x \in \mathbf{R}$. Your task is to determine *exactly which of the axioms hold and which fail*, using these definitions of vector addition and scalar multiplication.

To prove that an axiom fails you generally just need to give a single counterexample. For example, we have seen before that composition is not always commutative. So, to show that Axiom (3) in the definition of a vector space fails, you just have to show me two *specific* functions f and g such that $f + g = f \circ g$ is not the same as $g + f = g \circ f$. You need to give the rules for two specific functions, and then you have to show that there is a specific point x where $(f + g)(x) = f(g(x))$ is not equal to $(g + f)(x) = g(f(x))$. You just need one value of x for which equality fails, because then you know that these two functions $f + g$ and $g + f$ don't have the same rule and therefore aren't the same function.

Once you've done that, you know that the commutativity axiom fails, so you know that $\mathcal{F}(\mathbf{R})$ is NOT a vector space using these operations. However, for this exercise you need to determine EXACTLY which axioms are true and which fail. For example, we proved before that composition of functions is associative, so we have already proved that Axiom (4) *does* hold (you don't need to prove this again since we've already done it).

Here's a hint on Axiom (5): there IS an "additive identity" for this definition of vector addition. It's not the zero function that we used in Exercise 3.3. If we keep the symbol 0 for the function whose rule is $0(x) = 0$ for every x, then we have

$$(f + 0)(x) \;=\; (f \circ 0)(x) \;=\; f(0(x)) \;=\; f(0), \qquad x \in \mathbf{R}.$$

So we're not going to get $f + 0 = 0$ for every f (give a specific example of an f for which this fails!). The zero function is NOT the additive identity when vector addition is defined to be composition. But there is one—what is it? Let's give it a different name, let's call it z instead of 0. What is the rule for the function z that satisfies $f + z = f$ for every f, when vector addition means *composition*? $\diamondsuit$

Exercise 3.5. Again let $V = \mathcal{F}(\mathbf{R})$, but let "vector addition" mean *multiplication of functions*, i.e., we let $f + g$ be the function whose rule is $(f + g)(x) = f(x)g(x)$ for $x \in \mathbf{R}$. Determine which of the ten axioms hold, and which ones fail. $\diamondsuit$

Exercise 3.6. Fix positive integers m and n, and let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices. Define "vector addition" to be ordinary addition of matrices, and "scalar multiplication" is the usual multiplication of a matrix by a scalar. Show that $\mathcal{M}_{m\times n}$ is a vector space! $\diamondsuit$

We will see many more examples of vector spaces as we go along.

3.4 Some Consequences of the Vector Space Axioms

We will prove a few facts that follow immediately from the definition of a vector space. As we go along, we will prove many other facts, many of which require us to think more carefully and be more clever than we are in these proofs.

First we'll prove that there is only one zero vector in a vector space.

Lemma 3.7. *A vector space can have only one zero element.*

Proof. Let V be a vector space. One of the vector space axioms tells us that there has to be a zero vector in V, but it doesn't tell us that there couldn't be two of them (or more). The standard way to show that there can be only one something with a given property is to say "suppose that a and b both have this property" and then prove that a must equal b. This tells us that there's only one thing that has the property.

So, suppose that we had two things that were both zero vectors. Let's call them a and b. Note that I'm not assuming that a is not equal to b. If I wanted to do a proof by contradiction, then I would add that as the assumption to be contradicted, but I'm going to a direct proof instead of a proof by contradiction.

So, we've got these two vectors a and b that are both zero vectors. This means that

$$a + x = x, \qquad \text{for all } x \in V, \tag{3.1}$$

and similarly

$$b + x = x, \qquad \text{for all } x \in V. \tag{3.2}$$

Since equation (3.1) holds for every vector x, it holds for the specific vector $x = b$. Plugging $x = b$ into equation (3.1), we see that

$$a + b = b.$$

Similarly, taking $x = a$ in equation (3.2), it follows that

$$b + a = a.$$

Since vector addition is commutative, we therefore have

$$a = b + a = a + b = b.$$

Hence a and b are actually the same element, and therefore there's only one zero vector in V. □

Remember that we have to be careful to only use facts that are either given in the definitions are that we have already proved. Just because something seems "obvious," it might not be true. We have to prove that things *are* true. For example, it seems "obvious" that if x is a vector then $0x$, the scalar product of the number zero with the vector x, must be the zero vector. But how do we *know* this? We have to give a proof. Here it is.

Lemma 3.8. *If x is a vector in a vector space V, then $0x$ is the zero vector in V.*

Proof. To make the notation simpler, let $h = 0x$. We must show that $h = 0$ (note the difference between this 0 and the 0 in the symbols $0x$). First let's do a little calculation. We have

$$\begin{aligned} h + h &= (0x) + (0x) && \text{definition of } h \\ &= (0 + 0)x && \text{Distributive Law} \\ &= 0x && \text{Arithmetic of numbers} \\ &= h && \text{definition of } h. \end{aligned}$$

So $h + h = h$. We want to show that this implies that $h = 0$, but we can only use the axioms and things we've already proved. One of the axioms tells us that h has an inverse element $-h$. (I suspect that h and $-h$ are both the zero vector, but I haven't proved that yet, so I can't just assume that it true.) Adding $-h$ to both sides of the equation $h + h = h$, we obtain

$$(h + h) + (-h) = h + (-h).$$

On the right-hand side, we can apply the additive inverse axiom, which tells us that $h+(-h) = 0$. On the left-hand side, we can use associativity to rearrange parentheses. Doing these two things, we obtain

$$h + (h + (-h)) = 0.$$

Applying the additive inverse axiom again, we get

$$h + 0 = 0.$$

The additive identity axiom tells us that $h + 0 = h$, so it follows that

$$h = 0.$$

Since $h = 0x$, we've shown that $0x = 0$. □

Now let's prove that $(-1)x$, which is the scalar product of the number -1 with the vector x, is the same as the additive inverse $-x$.

Lemma 3.9. *If x is a vector in a vector space V, then $(-1)x$ is an additive inverse for x.*

Proof. One of the vector space axioms tells us that $1x = x$. Therefore we have

$$\begin{aligned} x + (-1)x &= 1x + (-1)x && \text{Multiplicative identity axiom} \\ &= (1 + (-1))x && \text{Distributive axiom} \\ &= 0x && \text{Arithmetic of numbers} \\ &= 0 && \text{by Lemma 3.8.} \end{aligned}$$

Hence $(-1)x$ is an additive inverse for x. (Note that I said "an"—for all we know right now, there might be more than one additive inverse. There isn't, but we haven't proved that yet.) □

Here are some more facts. I'm going to state these as exercises—you should try to prove them yourself, without looking at the textbook. But if you get stuck, you can look in the text, as several of these are worked out there (though not always with complete details).

Exercise 3.10. Let V be a vector space, and let x be any vector in V. Show that there is only one additive inverse for x. To do this, you have to assume that there were two vectors, say y and z, that both satisfied $x + y = 0$ and $x + z = 0$, and from this you somehow have to deduce that $y = z$. Be very careful—you can only use what is given in the definition of a vector space! For example, we have NOT defined a "subtraction" operation, and we have NOT proved that you can subtract a term from both sides of an equation! On the other hand, you CAN add things (like an additive inverse) to both sides of an equation!

Also prove the following facts.

(a) $c0 = 0$ for every scalar $c \in \mathbf{R}$. (Note that both 0's in this equation represent the zero vector.)

(b) $(-c)x = -(cx) = c(-x)$ for all $x \in V$ and $c \in \mathbf{R}$.

(c) Given $c \in \mathbf{R}$ and $x \in V$, if we have $cx = 0$ then either $c = 0$ or $x = 0$.

(d) If $cx = cy$ for some x, $y \in V$ and scalar $c \neq 0$, then $x = y$.

(e) If $ax = bx$ for some $x \in V$ and scalars $a \neq 0$, then $a = b$.

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

(g) $x + x = 2x$, $x + x + x = 3x$, and so forth (use induction). ♢

Notation 3.11 (Subtraction). From now on, we will write $x-y$ for $x+(-y)$. That is, we *declare* that $x - y$ means

$$x - y \ = \ x + (-y).$$

We call $x - y$ the *difference* of x and y. ♢

Subtraction is a good example of an operation that isn't associative. You check this, just with real numbers—show that subtractive of real numbers is neither associative nor commutative.

3.5 Exercises

Section 3.5 in Apostol's text consists of problems for you to work. A selection of suggested problems from his list is given on the class website.

Below are some additional exercises for you to work.

Definition 3.12. . Given real numbers $a < b$, we let

$$C[a, b] \ = \ \Big\{ f : f \text{ is a continuous function mapping } [a, b] \text{ into } \mathbf{R} \Big\}.$$

For each f, $g \in C[a, b]$, we define their *inner product* to be the number

$$\langle f, g \rangle \ = \ \int_a^b f(x)\, g(x)\, dx.$$

The L^1, L^2, and L^∞ *norms* of $f \in C[a, b]$ are defined to be

$$\|f\|_1 \ = \ \int_a^b |f(x)|\, dx,$$

$$\|f\|_2 \ = \ \left(\int_a^b |f(x)|^2\, dx \right)^{1/2},$$

$$\|f\|_\infty \ = \ \max_{x \in [a,b]} |f(x)|,$$

respectively. ♢

Here are a couple of facts about continuous functions that you can use without proof.

(a) Each function $f \in C[a,b]$ is bounded. Hence, if $f \in C[a,b]$ there is a number $M \geq 0$ such that

$$|f(x)| \leq M, \qquad \text{all } x \in [a,b].$$

(b) A continuous function f on the interval $[a,b]$ must achieve a max and min at some point. The absolute value of a continuous function is also continuous, so $|f|$ must also achieve a maximum at some point. That is, there is at least one point $t \in [a,b]$ such that

$$|f(t)| = \max_{x \in [a,b]} |f(x)|.$$

For this t we have

$$|f(x)| \leq |f(t)|, \qquad \text{all } x \in [a,b].$$

(c) If $f \in C[a,b]$ is continuous and $f(x) \neq 0$ for some point x, then there exists a $\delta > 0$ such that

$$\forall\, y \in [a,b], \quad |x-y| < \delta \implies f(y) \neq 0.$$

Now we state several exercises.

3.1. Prove that $C[a,b]$ is a vector space (the operations are the usual ones of addition of functions and multiplication of a function by a scalar).

3.2. Prove the following facts about the inner product of functions in $C[a,b]$. These facts show that the inner product of functions has properties that are analogous to those satisfied by the dot product of vectors in $\mathbf{R}^n$.

(a) $0 \leq \langle f, f \rangle < \infty$ for each $f \in C[a,b]$.

(b) $\langle f, f \rangle = 0$ if and only if $f = 0$. Note that $f = 0$ means that f is the *zero function.*

(c) $\langle f, g \rangle = \langle g, f \rangle$ for all $f,\ g \in C[a,b]$.

(d) $\langle f+g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$ for all $f,\ g,\ h \in C[a,b]$.

(e) The *Cauchy–Schwarz Inequality*:

$$\forall\, f, g \in C[a,b], \quad |\langle f, g \rangle| \leq \|f\|_2\, \|g\|_2.$$

3.3. Prove the following facts about the L^∞ norm of functions in $C[a,b]$.

(a) $0 \leq \|f\|_\infty < \infty$ for all $f \in C[a,b]$.

(b) $\|f\|_\infty = 0$ if and only if $f = 0$.

(c) $\|cf\|_\infty = |c|\,\|f\|_\infty$ for all $f \in C[a,b]$ and $c \in \mathbf{R}$.

(d) $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$ for all $f,\, g \in C[a,b]$.

Note: Be careful in this part, I want to see a correct proof that the maximum on the left-hand side of the inequality above is less than or equal to the sum of the maxima on the right-hand side of the inequality.

3.4. Prove that the L^1 norm of functions in $C[a,b]$ satisfies the same properties (a)–(d) that are given in the preceding problem (i.e., show that if we replace $\|f\|_\infty$ by $\|f\|_1$, then (a)–(d) still hold).

3.5. Prove the following facts about the relationship between the L^1 and the L^∞ norms on $C[a,b]$.

(a) $\forall\, f \in C[a,b], \quad \|f\|_1 \le (b-a)\,\|f\|_\infty$.

(b) Give an example of a function f such that equality holds in part (a), and give another example of a function such that strict inequality holds in part (a).

(c) Show that no matter what real number $C > 0$ we choose, the following statement is FALSE:

$$\forall\, f \in C[a,b], \quad \|f\|_\infty \;\le\; C\,\|f\|_1.$$

Compare this to the statement made in part (a) of this problem!

3.6 Subspaces

3.6.1 The Definition of a Subspace

It can be rather painful to prove that a given set is a vector space, because we have to prove that all ten of the axioms given in the definition of a vector space are satisfied. However, many times we are interested in sets that are part of some larger set that we already know is a vector space. We will see that if we know that V is a vector space and we have a set $S \subseteq V$, then we can determine with only a little work whether or not S is itself a vector space (assuming that we use the same operations that we used on V). First, however, we give a name to subsets that are themselves vector spaces.

Definition 3.13 (Subspace). Let V be a vector space. A subset S of V is called a *subspace* of V if S is a vector space. That is, S must satisfy all ten of the vector space axioms (using the same operations on S that we used on V). $\diamondsuit$

We will show that in order to check whether S is a *subspace*, we only have to check TWO of the vector spaces axioms, not all ten of them. The two axioms that we have to check are the closure axioms. In other words, we will show that S is a subspace if and only if it is closed under both vector addition and scalar multiplication. Well, technically there is one other requirement—we have to know that S is *not empty.*

Theorem 3.14. *Let V be a vector space and let S be a* nonempty *subset of V. Then S is a* subspace *if and only if the following two statements hold.*

(a) *If x, $y \in S$ then $x + y \in S$.*

(b) *If $x \in S$ and $c \in \mathbf{R}$ then $cx \in S$.*

Proof. $\Rightarrow$. Suppose that S is subspace. Then S satisfies all ten of the vector space axioms. Statements (a) and (b) are simply the first two of the axioms, so they are satisfied.

$\Leftarrow$. Now suppose that what we know is that S satisfies statements (a) and (b). This tells us that S satisfies the first two of the vector space axioms, and our task is to prove that all of the remaining eight axioms are satisfied as well.

The easy ones to check are the axioms that are worded purely in terms of a "for all" statement. For example, to prove that Axiom (3) in the vector space definition holds, we have to prove that $x + y = y + x$ for all x, $y \in S$. But we already know that the big space V is a vector space, so we know that Axiom (3) holds for V. In other words, we know that $x + y = y + x$ for all x, $y \in V$. Since S is just part of V, something that holds for all vectors in V must hold for vectors in S. Hence $x + y = y + x$ holds for x, $y \in S$ because it holds for all vectors x, $y \in V$.

In fact, this argument takes care of Axioms (3), (4), (7), (8), (9), and (10). The only axiom that isn't worded purely as a "for all" statement are Axiom (5), the existence of a zero vector, and Axiom (6), the existence of additive inverses. So, we need to be a little more careful in how we prove that these axioms hold for S.

Let's look at the zero vector issue first. Since V is a vector space, we know that V has a zero vector. There is a vector, which we call 0, that satisfies $0 + x = x$ for all $x \in V$. Since S is a subset of V, it is true that $0 + x = x$ for all $x \in S$, but this isn't quite what we need. The point of Axiom (5) is that there is a zero vector *in the set* S. We know that there's a zero vector in V, but is it in S?

Well, we know that S is nonempty, so we know that there is *some* vector x that belongs to S. We don't know that the zero vector belongs to S, but we do know that S is closed under vector addition (that's statement (a)), so we know that $cx \in S$ for every scalar c. In particular, by taking $c = 0$ (the number zero), we know that $0x \in S$. And Lemma 3.8 tells us that $0x$ is the zero vector in V. That is $0x = 0$. Since we know that $0x$ belongs to S, it follows that $0 \in S$. Hence S contains the zero vector, so Axiom (5) is satisfied.

The remaining thing we have to check is that Axiom (6) is satisfied. Given a vector $x \in S$, we know that x belongs to V. Since V is a vector space, x has an additive inverse $(-x)$, i.e., there is a vector $(-x) \in V$ that satisfies $x + (-x) = 0$. But the problem is that we have to show that $-x$ belongs to S. We know $-x$ belongs to V, but is it in S? I assign this part to you: Prove that $-x$ has to be in S. Hint: Use the fact that S is closed under scalar multiplication. $\square$

You should ask yourself the following question: If you have a set S that is a subset of a vector space V, and you prove that S is closed under vector addition and scalar addition (i.e., both statements (a) and (b) in the preceding theorem are satisfied), can you conclude that S is a subspace? The answer is NO. There does exist a subset of V that is closed under vector addition and scalar multiplication but is NOT a subspace. What is it?

Exercise 3.15. Let V be a vector space. Prove that the empty set $\emptyset$ is closed under vector addition and scalar multiplication. $\diamondsuit$

3.6.2 Examples

Here are some examples.

Exercise 3.16. Let $C(\mathbf{R})$ be the set of all functions $f\colon \mathbf{R} \to \mathbf{R}$ that are *continuous* at every point. Since $C(\mathbf{R})$ is a set of functions that map real numbers to real numbers, it is a *subset* of the set of all functions that map real numbers to real numbers. In other words,

$$C(\mathbf{R}) \subseteq \mathcal{F}(\mathbf{R}).$$

The symbols "$\subseteq$" mean *subset*, they do not mean *subspace*. Your exercise is to prove that $C(\mathbf{R})$ is a *subspace* of $\mathcal{F}(\mathbf{R})$. Implicitly, we are using the same operations on $C(\mathbf{R})$ that we use on $\mathcal{F}(\mathbf{R})$. That means that vector addition in $C(\mathbf{R})$ is addition of functions, and scalar multiplication is multiplication of a function by a scalar. To prove that $C(\mathbf{R})$ is a subspace, you just have to prove the following three things.

(a) $C(\mathbf{R})$ *is nonempty.* Often (but not always) the easiest way to prove that a set is nonempty is to show that the zero element belongs to it. So you could prove that $C(\mathbf{R})$ is nonempty by proving that the zero function belongs to $C(\mathbf{R})$. So, what is the the zero function and does it belong to $C(\mathbf{R})$, i.e., is it continuous?

(b) $C(\mathbf{R})$ *is closed under vector addition.* To do this, you have to prove that if $f, g \in C(\mathbf{R})$ then $f+g \in C(\mathbf{R})$. In other words, you have to show that if f, g are two continuous functions then $f+g$ is also a continuous function (maybe we already did this?).

(c) $C(\mathbf{R})$ *is closed under scalar multiplication.* Here you must show that if $f \in C(\mathbf{R})$ and $c \in \mathbf{R}$, then $cf \in C(\mathbf{R})$.

Once you proved that statements (a), (b), and (c) above hold, then you know (because of Theorem 3.14) that $C(\mathbf{R})$ is a subspace of $\mathcal{F}(\mathbf{R})$. Hence $C(\mathbf{R})$ is itself a vector space. Consequently, you know that all ten of the vector space axioms are satisfied for $C(\mathbf{R})$ because Theorem 3.14 tells you that they must be, but it's a good idea to ask yourself why they are true. Is it true that $f + g = g + f$ for all $f, g \in C(\mathbf{R})$? Can you explain why this follows from something we know about $\mathcal{F}(\mathbf{R})$? But can you also give a direct proof of this fact? ♢

Exercise 3.17. Show that

$$S = \{f \in C(\mathbf{R}) : f(1) = 0\}$$

is a subspace of $C(\mathbf{R})$. Can you do this by applying Theorem 3.14 using $V = C(\mathbf{R})$? (Yes. Why are you allowed to do this?) ♢

Exercise 3.18. Show that

$$T = \{f \in C(\mathbf{R}) : f(1) = 1\}$$

is NOT a subspace of $C(\mathbf{R})$. There are several ways to do this problem. Here are some.

(a) If you show that T is not closed under vector addition, then it doesn't satisfy one of the vector space axioms, so it can't be a vector space. Also, this would show that T doesn't satisfy the requirements given in Theorem 3.14, so T can't be a subspace.

(b) You could instead show that T is not closed under scalar multiplication. Do you need to prove that T isn't closed under BOTH vector addition and scalar multiplication? (No, why not?).

(c) You could show that the zero function does not belong to T. A subspace is a vector space, so it MUST contain the zero vector. If it doesn't, then it can't be a vector space, so it can't be a subspace. (Note that the converse of this statement isn't true: a subset that contains the zero vector need not be a subspace.) ♢

Exercise 3.19. Determine whether the following subsets of $\mathbf{R}^2$ are subspaces of $\mathbf{R}^2$. (Proof required.)

(a) $X = \{(x, y) \in \mathbf{R}^2 : y = 1\}$.

(b) $Q_1 = \{(x, y) \in \mathbf{R}^2 : x, y > 0\}$.

(c) $Q_2 = \{(x, y) \in \mathbf{R}^2 : x, y \geq 0\}$.

(d) $S = \{(a, 3a) : a \in \mathbf{R}\}$.

(e) $\mathbf{Z}^2 = \{(m, n) : m, n \in \mathbf{Z}\}$. ♢

Exercise 3.20. (a) Give an example of a subset of $\mathbf{R}^2$ that is closed under vector addition but not closed under scalar multiplication.

(b) Give an example of a subset of $\mathbf{R}^2$ that is closed under scalar multiplication but not closed under vector addition.

(c) Give an example of a subset of $\mathcal{F}(\mathbf{R})$ that is closed under vector addition but not closed under scalar multiplication.

(d) Give an example of a subset of $\mathcal{F}(\mathbf{R})$ that is closed under scalar multiplication but not closed under vector addition.

(e) Let V be a vector space. Show that if S is a nonempty subset of V and S is closed under scalar multiplication then $0 \in S$. Must S be a subspace of V? $\diamondsuit$

Exercise 3.21. (a) Let $\mathcal{P}_n$ be the set of all polynomials with degree at most n:

$$\mathcal{P}_n \;=\; \{a_n x^n + \cdots + a_1 x + a_0 : a_0, a_1, \dots, a_n \in \mathbf{R}\}.$$

Prove that $\mathcal{P}_n$ is a subspace of $C(\mathbf{R})$.

(b) Let $\mathcal{P}$ be the set of all polynomials of *any* degree:

$$\mathcal{P} \;=\; \{a_n x^n + \cdots + a_1 x + a_0 : n \geq 0, a_0, a_1, \dots, a_n \in \mathbf{R}\}.$$

Prove that $\mathcal{P}$ is a subspace of $C(\mathbf{R})$.

(c) Prove that $\mathcal{P}$ is the union of $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2, \dots$, i.e.,

$$\mathcal{P} \;=\; \bigcup_{n=0}^{\infty} \mathcal{P}_n \;=\; \mathcal{P}_0 \cup \mathcal{P}_1 \cup \mathcal{P}_2 \cup \cdots .$$

Hence the union of the subspaces $\mathcal{P}_n$ gives us the subspace $\mathcal{P}$.

(d) Is it true that the union of subspaces is always a subspace? (No.) Give an example of two subspaces S and T of $\mathbf{R}^2$ such that $S \cup T$ is NOT a subspace of $\mathbf{R}^2$. $\diamondsuit$

Exercise 3.22. Fix positive integers m and n, and recall from Exercise 3.6 that the set $\mathcal{M}_{m \times n}$ of all $m \times n$ matrices is a vector space (what are the operations?).

(a) Show that

$$S \;=\; \left\{ \begin{bmatrix} 1 & a \\ b & c \end{bmatrix} : a, b, c \in \mathbf{R} \right\}$$

is not a subspace of $\mathcal{M}_{2 \times 2}$.

(c) Let S be the set of all *symmetric* 2×2 matrices. Show that S is a subspace of $\mathcal{M}_{2 \times 2}$. Recall that a 2×2 matrix is symmetric if it equals its own transpose. For example, $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ is symmetric, but $\begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix}$ is not.

(c) Let S be the set of all symmetric 3×3 matrices. Show that S is a subspace of $\mathcal{M}_{3\times 3}$.

(d) Can you generalize part (b) to arbitrary sizes of matrices, i.e., if n is any positive integer and S is the set of all symmetric $n \times n$ matrices, can you show that S is a subspace of $\mathcal{M}_{n\times n}$? (It is a subspace—the issue is figuring out how to write a proof that isn't limited to just one particular size, like 3×3 matrices.) $\diamondsuit$

Exercise 3.23. Let $\mathcal{S}$ be the set of all infinite sequences of real numbers:

$$\mathcal{S} \;=\; \bigl\{x : x = (x_1, x_2, \dots) \text{ where } x_1, x_2, \dots \in \mathbf{R}\bigr\}.$$

For example, $x = (1, 2, 3, 4, \dots)$ is one element of $\mathcal{S}$.

(a) Prove that $\mathcal{S}$ is a vector space. (Sadly, you have to check all ten axioms.)

(b) Let

$$\ell^1 \;=\; \biggl\{x = (x_1, x_2, \dots) \in \mathcal{S} \;:\; \sum_{k=1}^{\infty} |x_k| < \infty\biggr\}.$$

Determine whether the following vectors belong to ℓ^1:

$$\begin{aligned} 0 &= (0, 0, 0, \dots),\\ x &= (1, -1, 1, -1, \dots),\\ y &= (1, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \dots),\\ z &= (1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \dots),\\ w &= (1, \tfrac{1}{4}, \tfrac{1}{9}, \tfrac{1}{16}, \dots). \end{aligned}$$

Prove that ℓ^1 is a subspace of $\mathcal{S}$.

(c) Let

$$c_0 \;=\; \biggl\{x = (x_1, x_2, \dots) \in \mathcal{S} \;:\; \lim_{k\to\infty} x_k = 0\biggr\}.$$

Determine which of the vectors from part (b) belong to c_0, and prove that c_0 is a subspace of $\mathcal{S}$.

(d) Is $\ell^1 \subseteq c_0$? Is $c_0 \subseteq \ell^1$? $\diamondsuit$

3.6.3 Some Properties of Subspaces

The next exercise shows that every vector space contains at least two subspaces.

Exercise 3.24. Let V be a vector space.

(a) Show that $S = \{0\}$ (the set that just contains the zero vector) is a subspace of V.

(b) Show that V is a subspace of V.

(c) Show that $\emptyset$ is not a subspace of V.

(d) Suppose that S is a subspace, and $S \neq \{0\}$. In order for this to happen, there must be at least one nonzero vector in S, i.e., there must be some vector $x \in S$ such that $x \neq 0$. Prove that in this case there must *infinitely many* vectors in S. Consequently, there's only one subspace of V that contains finitely many vectors, and it is the subspace $\{0\}$. ♢

Thus $\{0\}$ and V are always subspaces of a vector space V. We call these the *trivial subspaces* of V.

Note that 0 is *not* a subspace of V. A subspace is a *set* of vectors, but 0 is just a vector, not a set of vectors. Hence the smallest possible subspace of V is the set $\{0\}$. (Even so, it is true that mathematicians are often lazy and write "0 is a subspace" when they really mean that "$\{0\}$ is a subspace." You're not allowed to do this in this class.)

If you did part (d) of Exercise 3.21, then you know that the union of two subspaces might not be a subspace. (If you haven't solved that yet, here's a hint: Let S be the x-axis in $\mathbf{R}^2$ and let T be the y-axis in $\mathbf{R}^2$. Are these subspaces? Is $S \cup T$?)

But what about intersections of subspaces?

Lemma 3.25. *If S and T are subspaces of a vector space V, then $S \cap T$ is a subspace of V.*

Proof. Suppose that S and T are subspaces. We have to show that $S \cap T$ is nonempty, is closed under vector addition, and is closed under scalar multiplication.

The nonempty part is easy, because S and T must both contain the zero vector. That is, $0 \in S$ and $0 \in T$. So, by definition of intersection, we have $0 \in S \cap T$. This shows that $S \cap T$ is nonempty.

To prove closure under vector addition, suppose that x and y are two vectors in $S \cap T$. Then, by definition, we have $x, y \in S$ and $x, y \in T$. Since S is a subspace it follows that $x + y \in S$, and similarly $x + y \in T$ since T is a subspace. Therefore $x + y$ belongs to both S and T, so $x + y \in S \cap T$. Hence $S \cap T$ is closed under vector addition.

Your turn: Prove that $S \cap T$ is closed under scalar multiplication. Once you've done this, it follows that $S \cap T$ is a subspace. ♢

We can generalize Lemma 3.25 to more than one subspace. A first step is to use induction to extend from the intersection of two subspaces to the intersection of any finite number of subspaces.

Exercise 3.26. Show that if $S_1, \dots, S_n$ are subspaces of a vector space V, then $S_1 \cap \cdots \cap S_n$ is a subspace of V. $\diamondsuit$

However, we can do better than this. Using an argument very similar to the proof of Lemma 3.25 but using "for all" instead of "x and y", you can give a direct proof that the intersection of *any collection* of subspaces is a subspace, even if there are infinitely many that we want to intersect. Here's the exercise.

Exercise 3.27. Let V be a vector space. Suppose that I is some set (we call it an *index set*), and for each $i \in I$ we have a subspace S_i of V. Show that

$$\bigcap_{i \in I} S_i \;=\; \bigl\{x : x \in S_i \text{ for every } i \in I\bigr\}$$

is a subspace of V. $\diamondsuit$

The set I in the preceding exercise can be anything. If $I = \{1, \dots, n\}$, then

$$\bigcap_{i \in I} S_i \;=\; \bigcap_{i=1}^{n} S_i \;=\; S_1 \cap S_2 \cap \cdots \cap S_n.$$

If $I = \mathbf{N} = \{1, 2, 3, \dots\}$, then

$$\bigcap_{i \in I} S_i \;=\; \bigcap_{i=1}^{\infty} S_i \;=\; S_1 \cap S_2 \cap S_3 \cap \cdots.$$

If $I = \mathbf{R}$ then we have one subspace S_i for each real number i, and

$$\bigcap_{i \in I} S_i \;=\; \bigcap_{i \in \mathbf{R}} S_i \;=\; \bigl\{x : x \in S_i \text{ for every real number } i\bigr\}.$$

3.6.4 Spans, Part I

Apostol is a bit skimpy on his discussion of spans in Chapter 3, so it would be good for you to look back at Section 1.12 of his text, where he talks about the span of a set of vectors in $\mathbf{R}^n$. We'll be considering spans in abstract vector spaces, not just $\mathbf{R}^n$, but the definitions and ideas are very similar.

The idea of a span is that we have some vectors, and we want to create a subspace that contains those vectors. For example, suppose that we have a single vector x in a vector space V. If x is the zero vector, then the set that contains x is $\{0\}$, which is a subspace of V. However, it's not a very interesting subspace, so let's consider what happens if $x \neq 0$. In this case the set that contains just x (in other words, the set $\{x\}$) isn't a subspace. (Why not? Can you prove that $\{x\}$ isn't a subspace when $x \neq 0$?)

Can we find a nice subspace that contains the vector x? We could take the entire space V—this is a subspace and it contains *everything*, including x. But this is overkill, can we do better? Can we make a subspace S that contains x

but not too much other stuff? It will have to contain something more than just x since $\{x\}$ by itself isn't a subspace, but what else do we really need? We don't really know how to make S yet, but let's think about what vectors we would need. The issue is: *what do subspaces look like*?

The key to this is Theorem 3.14. That theorem told us that in order for a set S to be a subspace, it has to be closed under vector addition and scalar multiplication. Let's think about closure under vector addition first. If S is a subspace and x is in S, then x and x are two vectors in S, so closure under vector addition tells us that $x + x$ has to be in S. This vector $x + x$ is the same as the vector $2x$, so at the very least we have to have $2x \in S$. Applying closure under vector addition again, we must have $3x = 2x + x \in S$. So, every integer multiple of x has to be in S. Is this enough—can we get away with just the integer multiples and nothing else?

Exercise 3.28. Let x be a nonzero vector in a vector space V.

(a) Let S be the set of all positive integer multiples of x, i.e.,

$$S = \{x, 2x, 3x, \dots\} = \{nx : n \in \mathbf{N}\}.$$

Prove that this set S is *not* a subspace of V.

(b) Maybe we need all of the integer multiples, both positive and negative. Let S be the set of all integer multiples of x, i.e.,

$$S = \{\dots, -2x, -x, 0, x, 2x, \dots\} = \{nx : n \in \mathbf{Z}\}.$$

Prove that this set S is *not* a subspace of V. ♢

If we want to make a set S that contains x and is a subspace, we have to have more in S than just the *integer* multiples of x. We determined above that the integer multiples of x do have to be in S in order for closure under vector addition to hold, but that's just not enough by itself to make a subspace. We also have to satisfy closure under scalar multiplication, and that requires that we have *every real multiple* of x in the set. Is this enough? Yes. Well, you need to prove that, hence the next exercise.

Exercise 3.29. Let x be a nonzero vector in a vector space V. Let S be the set of all real multiples of x, i.e.,

$$S = \{cx : c \in \mathbf{R}\}.$$

Prove that S *is a subspace*, and x belongs to S. ♢

What does the subspace S in Exercise 3.29 "look like"?

Example 3.30. (a) Suppose that x is a nonzero vector in $\mathbf{R}^2$, say $x = (x_1, x_2)$. Then the set S given in Exercise 3.29 is

$$S = \{cx : c \in \mathbf{R}\} = \left\{ c \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : c \in \mathbf{R} \right\} = \left\{ \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix} : c \in \mathbf{R} \right\}.$$

We just take the vector x and stretch it by all possible scalars. We get the set of all scalar multiples of x, and this is precisely the line in $\mathbf{R}^2$ that passes through the origin 0 and the vector x.

(b) If x is a nonzero vector in $\mathbf{R}^3$, then $S = \{cx : c \in \mathbf{R}\}$ is again the set of all scalar multiples of x. This is just the line in $\mathbf{R}^3$ that passes through the origin and x. ♢

If $x \in \mathbf{R}^n$ is nonzero, then we can again visualize $S = \{cx : c \in \mathbf{R}\}$ as a line in $\mathbf{R}^n$ that passes through the origin (and through the vector x). We adopt this terminology for use in *any* vector space, as follows.

Notation 3.31. If x is a nonzero vector in a vector space V, then we call

$$S = \{cx : c \in \mathbf{R}\}$$

the *line though* x. (It would be more precise to say that it is *the line through* x *and the origin*, but we usually understand from context that we're talking about subspaces, and any subspace has to contain the origin.)

As we will explain later, we also call S the *span of the set* $\{x\}$, and we will write

$$S = \text{span}\{x\} = \text{span}\{x\} = \{cx : c \in \mathbf{R}\}.$$

When we're dealing with a *single nonzero vector*, the span of $\{x\}$ is the line through the vector x. We'll see that things are more complicated when we want to compute the span of a set that has more than one vector.

Note that if $x = 0$ then span$\{x\}$ isn't a "line," it's just the set that contains the zero vector:

$$x = 0 \implies \text{span}\{0\} = \{c0 : c \in \mathbf{R}\} = \{0\}.$$

This set is still the span of a single vector (the zero vector), but since every scalar multiple of the zero vector is 0, the span is just $\{0\}$. ♢

To illustrate the meaning of a "line" in an abstract vector space, let f be the function whose rule is $f(x) = \sin x$. Forget the fact that f is a function, and think of f as being a dot in the space $C(\mathbf{R})$ of all continuous functions. The zero function 0 is another dot in this space $C(\mathbf{R})$. Think of $2f$ as being the dot that is in the same direction as f except twice as far out, and so forth. Then the set of all scalar multiples of f is a line in $C(\mathbf{R})$. That is, the line in $C(\mathbf{R})$ through the function f is

$$S = \{cf : c \in \mathbf{R}\}.$$

We're *not* saying that the graph of f is a line. We're saying that we start with a single vector f in $C(\mathbf{R})$, and we make a set that consists of all of the scalar multiples of f. This *set of all scalar multiples of* f is a *set of functions*, and it is what we call the *line through* f *in the space* $C(\mathbf{R})$, or *the span of* $\{f\}$.

Exercise 3.32. (a) Let f be the function whose rule is $f(x) = \sin x$, and let g be defined by $g(x) = \cos x$. Does g lie on the line through f? That is, is it true that $g \in \operatorname{span}\{f\}$? Be sure to give a careful proof of your answer.

(b) Same question for $f(x) = x$ and $g(x) = x^2$. Is $g \in \operatorname{span}\{f\}$?

(c) With f and g as in part (b), explicitly describe the line through f, and likewise describe the line through g. That is, tell me exactly what functions are on these two lines. $\diamondsuit$

3.6.5 Spans, Part II

Now suppose that we have two nonzero vectors x, y in a vector space V, and we want to find a subspace that contains both x and y. Again, the entire space V is certainly one subspace that contains x and y, but I'd rather try to find just those vectors that we "really need" in order to make a subspace S that contains x and y.

Let's try to work backwards a little, like we did when we had just one vector to worry about. If S is a subspace and x, y are in S, what other vectors *must* be in S? Since S has to be closed under scalar multiplication, we at least have to have all scalar multiples of x and y in S, as otherwise it couldn't possibly be closed under scalar multiples. Can we get away with just these vectors? In other words, if we take S to be

$$S \;=\; \{cx : c \in \mathbf{R}\} \;\cup\; \{cy : c \in \mathbf{R}\}, \tag{3.3}$$

will S be a subspace? Tell me what you think of the following argument.

> *Argument.* I will try to show that the set S given in equation (3.3) is not a subspace. In order for S to be a subspace, it has to be closed under vector addition as well as scalar multiplication. Therefore, if x and y are in a subspace S then $x + y$ has to belong to S as well. But the vectors in S are cx and cy where $c \in \mathbf{R}$, so $x + y$ isn't one of the vectors in S. Therefore S is not a subspace. $\diamondsuit$

What do you think? This almost seems to work. After all, the elements of the set S given in equation (3.3) have the form cx or cy where c is a scalar. This doesn't *look* like $x + y$—but are you *sure* that $x + y$ couldn't be equal to cx or cy for some c? It's not enough to just say that $x + y$ is written with different symbols than cx. If we really think that $x + y$ is different than cx or cy then we have to *prove* this statement. And there's a big problem here, because *IT'S NOT TRUE* that we have to have $x + y \neq cx$ and $x + y \neq cy$ for any scalar c. It all depends on what the vectors x and y are. Here's an example where $x + y$ does equal cx for some scalar c.

Example 3.33. When you're looking for examples, it's always good to keep things simple. Let's consider the vector space $\mathbf{R}^2$, and let's take $x = (1,0)$ and $y = (2,0)$. Writing vertically, this is

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

For this choice of vectors, we have

$$x + y = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} = 3x.$$

So it is possible that $x+y$ can be a scalar multiple of x.

Exercise: You should explicitly work out what the set S given in equation (3.3) looks like. You should be able to show that, for this choice of x and y, the set

$$S = \{cx : c \in \mathbf{R}\} \cup \{cy : c \in \mathbf{R}\}$$

is precisely the same as the set

$$S = \{dx : d \in \mathbf{R}\}.$$

Hence this set S is simply the line through x, i.e.,

$$S = \operatorname{span}\{x\}. \qquad \diamondsuit$$

Here's a different example where the vector $x+y$ is not equal to cx or cy no matter what scalar c that we choose.

Example 3.34. This time let

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

If c is a scalar, then the vectors $x+y$, cx, and cy are:

$$x + y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad cx = \begin{bmatrix} c \\ 0 \end{bmatrix}, \qquad cy = \begin{bmatrix} 0 \\ c \end{bmatrix}.$$

No matter what scalar c that we choose, we have $x + y \neq cx$ and $x + y \neq cy$. Therefore, for this choice of x and y, the set

$$S = \{cx : c \in \mathbf{R}\} \cup \{cy : c \in \mathbf{R}\}$$

is not closed under vector addition and therefore isn't a subspace of $\mathbf{R}^2$.

Think about exactly what the set S is—it is the union of the x-axis and the y-axis in $\mathbf{R}^2$. This set is closed under scalar multiplication, but it isn't closed under vector addition. $\diamondsuit$

So, when we have two vectors x and y, we run into different possibilities depending on how these vectors are related to each other. Can you see the difference between the two preceding examples? In Example 3.33, the vector y was a scalar multiple of the vector x, while in Example 3.34 the vector y is not a scalar multiple of x. This is related to the issue of whether x and y are *linearly independent*—we will spend a lot of time considering that question later. Linear independence is an easy question when you only have *two* vectors, because it just comes down to the question of whether one vector is a scalar multiple of the other. But linear independence is a much trickier issue when you have more than two vectors! We will study this in detail later.

To summarize where we are, we've seen that if we have two vectors x and y in a vector space V, and we want to create a subspace S that contains both x and y, it's not always enough to take the union of the scalar multiples of x and the scalar multiples of y. This gives you the union of two lines in V, but that won't always be a subspace. It is a subspace if the two lines are identical, but it isn't if the two lines aren't identical. (When are the two lines identical? This does *not* require that x and y be equal—see Example 3.33!)

In order to be *sure* that we have a subspace, we need to combine *both* vector addition *and* scalar multiplication. We need all the scalar multiples of x, all the scalar multiples of y, and we also need all the possible sums of such vectors. Is that enough to make a subspace? Let's check.

Theorem 3.35. *If x and y are any two vectors in a vector space V, then*

$$S \;=\; \bigl\{ax + by : a, b \in \mathbf{R}\bigr\} \tag{3.4}$$

is a subspace of V, and x, y belong to S.

Proof. We have to show that S is nonempty, S is closed under vector addition, S is closed under scalar multiplication, and S contains both x and y.

The fact that S contains x and y is easy, because

$$x \;=\; 1x + 0y \;\in\; S, \qquad y \;=\; 0x + 1y \in S.$$

This also tells us that S is nonempty. We could also prove that by showing that S contains the zero vector. This follows because we have

$$0 \;=\; 0x + 0y \;\in\; S.$$

Now we will show that S is closed under vector addition. To do this, we have to prove that if we choose any two vectors in S, then their sum stays in S. We need to be careful about notation, because the letters x and y are already used for something. So, let's use different letters for the two vectors in S.

Let p and q be any two vectors in S. Because p belongs to S, we must have

$$p \;=\; ax + by \quad \text{for some scalars } a, b \in \mathbf{R}.$$

Similarly, since $q \in S$ we have

$$q = cx + dy \quad \text{for some scalars } c, d \in \mathbf{R}.$$

Since vector addition is both commutative and associative and the distributive rules are satisfied, it follows that

$$\begin{aligned} p + q &= (ax + by) + (cx + dy) \\ &= (ax + cx) + (by + dy) \\ &= (a + c)x + (b + d)y. \end{aligned}$$

You should check that we really had to use commutativity, associativity, and the distributive law to do the computations on the preceding lines! Let's choose some names. Let $r = a + c$ and $s = b + d$. These are just two scalars, so we have

$$p + q = rx + sy \in S.$$

The set S contains every scalar multiple of x plus every scalar multiple of y, and $rx + sy$ is precisely one of these, so it belongs to S. This shows that the set S is closed under vector addition.

Exercise: Prove that S is closed under scalar multiplication. Once that is done, the proof is complete—we've shown that S is a subspace of V. □

Exercise 3.36. In equation (3.4) we took all of the possible scalar multiples of x plus scalar multiples of y, and we were careful to let the scalar that multiplies x be different from the scalar that multiplies y. Could we have gotten away with using the same scalar on both x and y? In other words, if we let S be the set

$$S = \{cx + cy : c \in \mathbf{R}\},$$

will it be true that S must be a subspace? Well, in fact it is—but the problem is that S might not contain either x or y. Here's what you need to do.

(a) Prove that $S = \{cx + cy : c \in \mathbf{R}\}$ is a subspace of V.

(b) Give a specific example of V, x, and y such that the vector $x + y$ does not belong to S. (Keep it simple—try $V = \mathbf{R}^2$.) ♢

Now you need to do some examples.

Exercise 3.37. This exercise takes place in the vector space $V = \mathbf{R}^2$. For the vectors x, y given below, find an explicit description of the set

$$S = \{ax + by : a, b \in \mathbf{R}\}.$$

That is, tell me exactly which vectors are in S. I especially want to know whether S is all of $\mathbf{R}^2$ or is just a part of $\mathbf{R}^2$.

(a) $x = (1, 0)$ and $y = (0, 1)$.

(b) $x = (1, 0)$ and $y = (2, 0)$.

(c) $x = (1, 0)$ and $y = (0, 0)$.

(d) $x = (0, 0)$ and $y = (0, 0)$.

(e) $x = (1, 0)$ and $y = (1, 1)$.

You should be able to prove that the set S that you find in part (a) is the same as the one in part (e), and the one in parts (b) and (c) are likewise equal.

Don't just tell me that S is the set of all vectors $ax + by$ where a, b are in $\mathbf{R}$. I want to be able to explicitly see which vectors are in S. For example, in part (e) you should prove that $S = \mathbf{R}^2$, i.e., *every* vector is in S. I can't see that by looking at the formula $S = \{ax + by : a, b \in \mathbf{R}\}$. For part (e) you have to *prove* that the set S and the set $\mathbf{R}^2$ are equal. How do you prove that two sets are equal? ♢

Now we will introduce some notation for the things that we've done. We keep running into vectors of the form $ax+by$ over and over, so we'll give these kinds of vectors a name.

Definition 3.38. A *linear combination* of two vectors x, $y \in V$ is any vector that has the form

$$ax + by \qquad \text{where } a, b \in \mathbf{R}. \qquad \diamondsuit$$

So, if you have a vector z and you can show that $z = ax + by$ for some scalars a and b, then we say that z is linear combination of x and y. It may or may not be easy to see whether z is a linear combination of x and y. Usually you can't just look at z and tell. If you start with x and y then it's easy to make linear combinations, you just multiply x by some number a, multiply y by b, and add the results together to get $ax + by$. But if I just give you some vector z and ask *is z a linear combination of x and y*, then you've got work to do, because you have to find out if there are some scalars a and b that satisfy $z = ax + by$.

Exercise 3.39. (a) Show that every vector $z \in \mathbf{R}^2$ is a linear combination of the vectors $x = (1, 0)$ and $y = (0, 1)$.

(b) Show that every vector $z \in \mathbf{R}^2$ is a linear combination of the vectors $x = (1, 0)$ and $y = (1, 1)$.

(c) Show that there are vectors in $\mathbf{R}^2$ that are *not* linear combinations of $x = (1, 5)$ and $y = (2, 10)$. ♢

Exercise 3.40. Let f and g be the functions whose rules are

$$f(x) = 1, \qquad x \in \mathbf{R},$$

and

$$g(x) = x, \qquad x \in \mathbf{R}.$$

These functions f and g are vectors in the vector space $\mathcal{F}(\mathbf{R})$. A linear combination of f and g is any function h that can be written as $h = af + bg$ for some scalars a, $b \in \mathbf{R}$.

(a) Let h be the function whose rule is $h(x) = 3x + 2$, $x \in \mathbf{R}$. Prove that h is a linear combination of f and g.

(b) Let k be the function whose rule is $k(x) = \sin x$. Prove (carefully!) that k is not a linear combination of f and g. ♢

Look back at the set $S = \{ax + by : a, b \in \mathbf{R}\}$ that we defined in equation (3.4). Using the terminology of linear combinations, S is simply the set of all possible linear combinations of x and y. We have a special name for this set.

Definition 3.41. Let x and y be two vectors in V. The *span* of the set $\{x, y\}$ is the set of all possible linear combinations of x and y. We write this as follows:

$$\operatorname{span}(\{x, y\}) \;=\; \{ax + by : a, b \in \mathbf{R}\}.$$

Sometimes we just say that $\operatorname{span}(\{x, y\})$ is *the span of x and y*, but a more precise description is that it is the span of the two-element set $\{x, y\}$. To avoid writing lots of parentheses, we often write

$$\operatorname{span}\{x, y\} \qquad \text{instead of} \qquad \operatorname{span}(\{x, y\}).$$

In any case, the span of x and y is the set of *all possible linear combinations* of x and y. ♢

Exercise 3.42. Find an explicit description of $\operatorname{span}\{x, y\}$ for each of the pairs of vectors given in Exercise 3.37.

Hint: You already did the work when you worked out that exercise. This exercise is asking for exactly the same thing—it just formulates the question in a different terminology. ♢

Exercise 3.43. Let f and g be the functions whose rules are $f(x) = 1$ and $g(x) = x$ for every x.

(a) Prove that

$$\operatorname{span}\{f, g\} \;=\; \mathcal{P}_1,$$

where $\mathcal{P}_1$ is the set of all polynomials with degree at most 1 that we introduced in Exercise 3.21.

(b) Let h be defined by $h(x) = x^2$. Find an explicit description of $\operatorname{span}\{f, h\}$ and $\operatorname{span}\{g, h\}$. Does either of these sets equal $\mathcal{P}_2$? ♢

Exercise 3.44. Challenge problem: Prove that *no matter which two functions f and g that we choose*, we can NEVER have

$$\operatorname{span}\{f, g\} \;=\; \mathcal{F}(\mathbf{R}). \qquad ♢$$

The preceding exercise says that no matter what two vectors f and g that we choose, $\mathcal{F}(\mathbf{R})$ is not equal to the span of f and g. In contrast, another vector space that we often work with, namely $\mathbf{R}^2$ can be written as the span of two vectors. Does $\mathbf{R}^3$ equal the span of two vectors? That is, do there exist vectors $x,\, y \in \mathbf{R}^3$ such that

$$\mathbf{R}^3 \;=\; \operatorname{span}\{x, y\}?$$

Can you prove your answer?

3.6.6 Spans, Part III

You can probably guess what's coming next: We want to form linear combinations of three, four, or more vectors. Here's what we mean by a linear combination of n vectors.

Definition 3.45 (Linear Combination). Let $x_1, \dots, x_n$ be n vectors in a vector space V. A *linear combination* of $x_1, \dots, x_n$ is any vector of the form

$$c_1 x_1 + \cdots + c_n x_n \;=\; \sum_{k=1}^{n} c_k x_k$$

where $c_1, \dots, c_n \in \mathbf{R}$. $\diamondsuit$

Given vectors $x_1, \dots, x_n$ it's easy to make a linear combination—you just multiply each x_k by some scalar c_k and add the results together. In general, the converse question is much harder, i.e., if I just give you some vector z and ask you whether it is a linear combination of $x_1, \dots, x_n$ then you have to do some work to figure out whether there are scalars $c_1, \dots, c_n$ such that $z = c_1 x_1 + \cdots + c_n x_n$. Back when you took MATH 1502, you learned how to do this for vectors that are in the vector space $\mathbf{R}^n$. You have to set up a system of linear equations and see if it has a solution. Since this was taught in 1502, I'm going to assume that you remember how to do this!

We define the *span* of a set of vectors to be the set of all possible linear combinations of those vectors. Here's the precise definition.

Definition 3.46 (Span). Let $x_1, \dots, x_n$ be n vectors in a vector space V. The *span* of the set $\{x_1, \dots, x_n\}$ (or simply the *span of* $x_1, \dots, x_n$ for short), is the set of all possible linear combinations of these vectors. We call this set $\operatorname{span}(\{x_1, \dots, x_n\})$ or simply $\operatorname{span}\{x_1, \dots, x_n\}$. It is given by the following formula:

$$\begin{aligned} \operatorname{span}\{x_1, \dots, x_n\} \;&=\; \bigl\{c_1 x_1 + \cdots c_n x_n : c_1, \dots, c_n \in \mathbf{R}\bigr\} \\ &=\; \biggl\{\sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R}\biggr\}. \qquad \diamondsuit \end{aligned}$$

The case where we have $n = 1$ in Definition 3.46 can be a little confusing. In this case the sums in the definition only have one term. That is, Definition 3.46 says that the span of a single vector x_1 is

$$\text{span}\{x_1\} \;=\; \bigl\{c_1x_1 : c_1 \in \mathbf{R}\bigr\}.$$

When we only have a single vector, we usually dispense with subscripts and just write the vector as x instead of x_1, and similarly we let c denote a generic scalar instead of using c_1. Using the choices of letters, the span of the single vector x is

$$\text{span}\{x\} \;=\; \bigl\{cx : c \in \mathbf{R}\bigr\}.$$

If x is nonzero then this is precisely the line through x. On the other hand, *if x is zero* then we have

$$\text{span}\{0\} \;=\; \bigl\{c0 : c \in \mathbf{R}\bigr\} \;=\; \{0\}.$$

We proved in Theorem 3.35 that the span of two vectors is a subspace. The proof that the span of n vectors is a subspace is very similar, so I assign it to you to do as the next exercise.

Exercise 3.47. Given finitely many vectors $x_1, \dots, x_n$ in a vector space V, prove that $\text{span}\{x_1, \dots, x_n\}$ is a subspace of V. $\diamondsuit$

One thing we are very interested in is whether the span of our vectors is the entire space V or not. Here is some terminology for that.

Definition 3.48. Let $x_1, \dots, x_n$ be n vectors in a vector space V. We say that $x_1, \dots, x_n$ *span the space V* if we have

$$\text{span}\{x_1, \dots, x_n\} \;=\; V.$$

If this doesn't happen then $x_1, \dots, x_n$ do not span V. $\diamondsuit$

Here are some exercises about vectors in $\mathbf{R}^n$. We need to be a little careful, because we're tempted to use the same letter n for the number of vectors as for the dimension of $\mathbf{R}^n$, but there's no reason that these need to be the same. So I'll use the letter d for the dimension instead, i.e., we will be working in the vector space $\mathbf{R}^d$.

Exercise 3.49. (a) Find two vectors x_1, x_2 in $\mathbf{R}^2$ that span $\mathbf{R}^2$. Find two different vectors y_1, y_2 in $\mathbf{R}^2$ that also span $\mathbf{R}^2$. Find three vectors z_1, z_2, z_3 that span $\mathbf{R}^2$. Show that there is no single vector that spans $\mathbf{R}^2$, i.e., no matter what vector $x \in \mathbf{R}^2$ that we choose, we never have

$$\mathbf{R}^2 \;=\; \text{span}\{x\}.$$

Be sure to *prove* this carefully.

(b) Find (explicitly) d vectors $x_1, \dots, x_d$ that span $\mathbf{R}^d$. Given an explicit example of d vectors $y_1, \dots, y_d$ that do not span $\mathbf{R}^d$. Can you find vectors $z_1, \dots, z_d$ whose span is a line in $\mathbf{R}^d$?

(c) Let

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad x_d = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Do the vectors $x_1, \dots, x_d$ span $\mathbf{R}^d$?

(d) Challenge (hard!): Show that if $x_1, \dots, x_{d-1}$ are *any* choice of $d-1$ vectors in $\mathbf{R}^d$, then

$$\mathbf{R}^d \neq \operatorname{span}\{x_1, \dots, x_{d-1}\}. \qquad \diamondsuit$$

Here is an exercise about functions. Recall from Exercise 3.21 that $\mathcal{P}_n$ stands for the set of all polynomials that have degree at most n, while $\mathcal{P}$ stands for the set of all polynomials of any degree. Also remember that a polynomial is a function that has the form

$$f(x) = \sum_{k=0}^{n} c_k x^k \tag{3.5}$$

for some finite integer n and some scalars $c_0, c_1, \dots, c_n$. If $c_n \neq 0$ then we say that f *has degree* n. By definition, the series in equation (3.5) is a *finite sum*, i.e., there can only be finitely many terms. *Polynomials* are not defined using infinite sums.

Exercise 3.50. Let $p_0, p_1, p_2, \dots$ be the functions whose rules are

$$p_0(x) = 1, \qquad p_1(x) = x, \qquad p_2(x) = x^2, \qquad \dots.$$

That is, p_k is the function whose rule is $p_k(x) = x^k$, $x \in \mathbf{R}$.

(a) Prove that

$$\operatorname{span}\{p_0, p_1, \dots, p_n\} = \mathcal{P}_n.$$

(b) Find three functions q_0, q_1, q_2 different from p_0, p_1, p_2 but such that

$$\operatorname{span}\{q_0, q_1, q_2\} = \mathcal{P}_2.$$

Note that according to part (a) we also have $\operatorname{span}\{p_0, p_1, p_2\} = \mathcal{P}_2$.

(c) Find four functions r_1, r_2, r_3, r_4 that span $\mathcal{P}_2$.

(d) Prove that if $f_1, \dots, f_m$ are finitely many polynomials, then

$$\operatorname{span}\{f_1, \dots, f_m\} \neq \mathcal{P}.$$

This part is tricky because you don't know what $f_1, \dots, f_m$ are. In particular, you don't know that f_k has degree k. The only thing that you know is that each function f_k is a polynomial. The polynomial f_k has some degree, but you don't know what it is. Also, m can be any finite number. It can only be *one particular* finite number, but you don't know what it is. In particular, you can't say "let $m = \infty$", because that's not a number. ♢

3.6.7 Spans, Part IV

You may have thought that we've covered all the possibilities for a span—we've done the span of one vector, two vectors, and finally n vectors. But there are more situations to consider.

First, what if we have infinitely many vectors? Does it make sense to define the span of a set of infinitely many vectors? We must be *very careful* here, because whatever definition we make has to apply to generic vector spaces, not just to $\mathbf{R}^d$.

You might be tempted to think that if we have infinitely many vectors $x_1, x_2, \dots$, then we can define a linear combination of these vectors to be a vector that has the form

$$c_1 x_1 + c_2 x_2 + \cdots = \sum_{k=1}^{\infty} c_k x_k.$$

However, there are a number of problems with this. The first problem is what we mean by an infinite series when we're dealing with vectors in a generic vector space. We won't get into this too much, but you should remember from MATH 1502 that if we're talking about real numbers, then an infinite series really means a *limit of the partial sums*. That is, if $a_k \in \mathbf{R}$ then we say that

$$\sum_{k=1}^{\infty} a_k = L \qquad \text{(the infinite series converges and equals } L\text{)}$$

if and only if the partial sums of the series converge to L, i.e.,

$$\lim_{N\to\infty} \sum_{k=1}^{N} a_k = \lim_{N\to\infty} (a_1 + \cdots + a_N) = L.$$

But in order to define what a limit is, we have to be able to talk about the *distance* between numbers. The partial sums have a limit if they get "closer and closer" to L. However, distance and "closer and closer" are statements about *numbers*—there's no part of the definition of a vector space that says we have to be able to define a distance between vectors. We might be able to, but often we can't. If distance has no meaning in our particular vector space then we can't talk about limits and therefore infinite sums have no meaning. As long as we're talking about *generic* vector spaces, we have to exclude any discussion of infinite sums, because they don't always make sense. In *some* vector spaces it does make sense, and in fact those are the types of vector spaces that I usually work in myself. We'll come back to those types of spaces later (we call them *normed vector spaces*), but for now we must restrict our discussion to finite sums.

So, even if we have infinitely many vectors, when we form a linear combination we're only allowed to use finitely many of them at a time.

Exercise 3.51. Suppose that $x_1, x_2, \dots$ are infinitely many vectors in a vector space V. Suppose that we choose just finitely many of these, say

$$x_{n_1},\, x_{n_2},\, \dots,\, x_{n_k}.$$

A linear combination of these vectors is any vector of the form

$$y \;=\; c_{n_1}x_{n_1} + c_{n_2}x_{n_2} + \cdots + c_{n_k}x_{n_k},$$

where $c_{n_1}, \dots, c_{n_k}$ are scalars. Prove that we can find some integer N and some scalars c_k such that

$$y \;=\; \sum_{k=1}^{N} c_k x_k \;=\; c_1x_1 + \cdots + c_Nx_N.$$

Hint: Start with an example. What would you do if you had $k = 3$ and $n_1 = 5$, $n_2 = 11$, $n_3 = 97$? $\diamondsuit$

We define the span of infinitely many vectors to be the set of all possible linear combinations of finitely many of those vectors at a time. If our vectors are $x_1, x_2, \dots$, then the preceding exercise shows us that any linear combination of finitely many of these vectors can be written as $c_1x_1 + \cdots + c_Nx_N$ for some N and scalars $c_1, \dots, c_N$ (note that some of the c_k might be zero!). Taking all of these possible linear combinations gives us the span of $x_1, x_2, \dots$.

Definition 3.52. If $x_1, x_2, \dots$ are vectors in a vector space V, then the *span* of these vectors is the set of all possible linear combinations of finitely many of them. We write this as

$$\operatorname{span}\{x_1, x_2, \dots\} \;=\; \left\{\sum_{k=1}^{N} c_k x_k : N \in \mathbf{N},\; c_1, \dots, c_N \in \mathbf{R}\right\}.$$

Note that there's no limit on how big the number N can be, although for each *particular* linear combination the series $\sum_{k=1}^{N} c_k x_k$ has only finitely many terms. We can have $N = 1$, $N = 100$, $N = 10^{100}$ and so forth, but it is finite. There are no infinite sums in our definition of span. $\diamondsuit$

Exercise 3.53. Let $p_0, p_1, p_2, \dots$ be the polynomials defined by $p_k(x) = x^k$ for $x \in \mathbf{R}$. Prove that

$$\operatorname{span}\{p_0, p_1, \dots\} \;=\; \mathcal{P}.$$

What is

$$\operatorname{span}\{p_1, p_2, \dots\}? \qquad \diamondsuit$$

There are infinite sets that *cannot* be listed. However, the span of such a set is defined in the same way, it is the set of all possible linear combinations of finitely many of the vectors in the set. We won't have to worry too much about this type of situation so I won't get into more detail about it now.

On the other hand, there is another extreme case that we do have to worry about. What if we have a set of NO vectors, i.e., we have the empty set. What is the span of the empty set? This is one of those questions that doesn't have a completely satisfying answer. You might think that $\operatorname{span}(\emptyset)$ should be the empty set, because it's the set of all linear combinations of vectors in $\emptyset$, and there aren't any vectors in $\emptyset$ so there are any linear combinations. This is a perfectly valid point of view, but for reasons that we'll see later it turns to not be the best definition of $\operatorname{span}(\emptyset)$. So for now you'll just have to believe me, the following definition is better.

Definition 3.54 (Span of the Empty Set). We *declare* that the span of the empty set is

$$\operatorname{span}(\emptyset) \;=\; \{0\}.$$

That is, the span of the empty set is the set that contains the zero vector but no other vectors. $\diamondsuit$

Here's one reason why this is better than saying that $\operatorname{span}(\emptyset)$ is the empty set. In every case that we did before, the span turned out to be a subspace of V. In some sense, the span of $x_1, \dots, x_n$ is the "smallest" subspace that we can make that contains the vectors $x_1, \dots, x_n$. If we decided to make $\operatorname{span}(\emptyset)$ be the empty set, then it wouldn't be a subspace. On the other hand, if we set $\operatorname{span}(\emptyset) = \{0\}$, then it is a subspace, and in fact it is the very smallest subspace that we can make that contains the empty set. So if we think of a span in terms of "smallest subspace" instead of linear combinations, then Definition 3.54 may make sense.

Here's an exercise that may make this idea of "smallest subspace" a little more palatable. I'll state this for the case of finitely many vectors, but the same idea works for the span of infinitely many vectors.

Exercise 3.55. Let $x_1, \dots, x_n$ be finitely many vectors in a vector space V. Let S be the span of $x_1, \dots, x_n$, i.e.,

$$S \;=\; \operatorname{span}\{x_1, \dots, x_n\} \;=\; \left\{ \sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R} \right\}.$$

According to an earlier exercise, the set S is a subspace, and it contains each of the vectors $x_1, \dots, x_n$ (if you didn't prove that before, you should do it now).

There are other subspaces that contain the vectors $x_1, \dots, x_n$. For example, the entire space V is a subspace that contains $x_1, \dots, x_n$. Our goal in this exercise is to show that the space S is somehow the "smallest" among all of these subspaces. To do this, let T be any one of these subspaces. That is, suppose that:

(a) T is a subspace of V AND

(b) $x_1, \dots, x_n \in T$.

Prove that S is smaller than T in the sense that we must have

$$S \subseteq T.$$

Thus, no matter which subspace T that we choose, if T contains the vectors $x_1, \dots, x_n$ then the subspace S is sitting inside T. ♢

3.7 Dependent and Independent Sets

In the last section, we defined the *span* of a set of vectors to be the set of all possible linear combinations of these vectors. In particular, the span of a fixed set of finitely many vectors $S = \{x_1, \dots, x_n\}$ is

$$\operatorname{span}(S) = \operatorname{span}\{x_1, \dots, x_n\} = \left\{ \sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R} \right\}.$$

If we have a set S that contains infinitely many vectors, then the span of S is still the set of all linear combinations of elements of S, but we have to remember that any particular linear combination is a sum of *finitely many* of the elements of S. In particular, if we can write $S = \{x_1, x_2, \dots\}$ then the span of S is

$$\operatorname{span}\{x_1, x_2, \dots\} = \left\{ \sum_{k=1}^{n} c_k x_k : n > 0, c_1, \dots, c_n \in \mathbf{R} \right\}.$$

When we form the span of a set of infinitely many vectors, we include linear combinations of $x_1, \dots, x_n$ *for every possible value of* n. We might form a linear combination of a hundred, a million, or a billion vectors, but each individual linear combination is a sum of finitely many of the infinitely many vectors in S. However we write it, the span of a set of vectors S is *the set of all linear combinations of elements of* S.

Before proceeding, I should mention that Apostol uses a somewhat uncommon notation to denote a span. Instead of writing $\operatorname{span}(S)$, he prefers to write $L(S)$. The letter L stands for "linear" and reminds us that this is the *linear span*, which is the set of all *linear combination* of vectors from S. I prefer to write $\operatorname{span}(S)$, but please keep in mind when you read the text that these notations mean exactly the same thing:

$$L(S) \;=\; \operatorname{span}(S).$$

3.7.1 The Definition of Independence

One of the things we were interested in Section 3.6 was whether the span of a set of vectors was the entire vector space. We said that a set S *spans* the space V if the span of S is all of V. That is,

$$S \text{ spans } V \qquad \Longleftrightarrow \qquad \operatorname{span}(S) \;=\; V.$$

What we're interested in now is whether there are any duplicates in these linear combinations. By definition, the span of $S = \{x_1, \dots, x_n\}$ is

$$\operatorname{span}(S) \;=\; \operatorname{span}\{x_1, \dots, x_n\} \;=\; \biggl\{ \sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R} \biggr\},$$

but are all of those linear combinations different, or could some of them be the same? Before addressing this question, we need the following terminology.

Notation 3.56. Let $x_1, \dots, x_n$ be finitely many vectors in V. We said before that a *linear combination* of these vectors is any vector that can be written as

$$\sum_{k=1}^{n} c_k x_k \;=\; c_1 x_1 + \cdots + c_n x_n \qquad \text{for some scalars } c_1, \dots, c_n \in \mathbf{R}.$$

Now we introduce a new term. We will say that $\sum_{k=1}^{n} c_k x_k$ is a *nontrivial linear combination* if $c_1, \dots, c_n$ are *not all zero.* $\diamondsuit$

For example, $x_1 + 2x_2 - 5x_3$ is a nontrivial linear combination of x_1, x_2, x_3, but $0x_1 + 0x_2 + 0x_3$ is a trivial linear combination of x_1, x_2, x_3. Naturally, the trivial linear combination of $x_1, \dots, x_n$ equals the zero vector, but you should ask yourself whether it is possible for a *nontrivial* linear combination of $x_1, \dots, x_n$ to equal the zero vector. (Yes, it's possible—give an example!)

We want to know if all of the linear combinations of $x_1, \dots, x_n$ are different, or if there could be some duplications. Because we are working with *linear* combinations, the question of whether there are duplicates can be reformulated as a question about the nontrivial linear combinations of the vectors, as follows.

Lemma 3.57. *Let $x_1, \dots, x_n$ be finitely many vectors in V. Then*

$$\sum_{k=1}^{n} c_k x_k \;=\; \sum_{k=1}^{n} d_k x_k \quad\Longleftrightarrow\quad \sum_{k=1}^{n} (c_k - d_k) x_k \;=\; 0. \tag{3.6}$$

Therefore, the following two statements are equivalent.

(a) *Two different linear combinations of $x_1, \dots, x_n$ are equal.*

(b) *Some nontrivial linear combination of $x_1, \dots, x_n$ equals* 0.

Proof. Equation (3.6) is easy to prove (you should do it—be sure that you prove the *if and only if* statement in that equation!). Therefore, let's concentrate on proving that statement (a) is true if and only if statement (b) is true.

(a) $\Rightarrow$ (b). Suppose that two different linear combinations of $x_1, \dots, x_n$ are equal. This means that we have

$$\sum_{k=1}^{n} c_k x_k \;=\; \sum_{k=1}^{n} d_k x_k$$

for some scalars c_k and d_k, and furthermore these are not the same linear combination. Not being the same linear combination does NOT mean that $c_k \neq d_k$ for every k! For example,

$$2x_1 - 3x_2 + 5x_3 \qquad \text{and} \qquad 2x_1 - 4x_2 + 5x_3$$

are two different linear combinations of x_1, x_2, x_3 yet we do not have $c_k \neq d_k$ for every k. Instead, the assumption that $\sum_{k=1}^{n} c_k x_k$ and $\sum_{k=1}^{n} d_k x_k$ are different linear combinations just means that there is *at least one index j* such that $c_j \neq d_j$.

However, even though $\sum_{k=1}^{n} c_k x_k$ and $\sum_{k=1}^{n} d_k x_k$ are different linear combinations, our hypothesis is that they give us the *same vector*. That is, we are assuming that $\sum_{k=1}^{n} c_k x_k = \sum_{k=1}^{n} d_k x_k$. Consequently (why?), we have

$$\sum_{k=1}^{n} (c_k - d_k) x_k \;=\; 0.$$

We *might* have $c_k = d_k$ for some indices k, but we know that for the particular index j we have $c_j \neq d_j$. Therefore, for the index j we have $c_j - d_j \neq 0$. Hence $\sum_{k=1}^{n} (c_k - d_k) x_k$ is a *nontrivial linear combination* of $x_1, \dots, x_n$. Yet we know that $\sum_{k=1}^{n} (c_k - d_k) x_k = 0$, so we have a nontrivial linear combination that equals 0.

(b) $\Rightarrow$ (a). Suppose that some nontrivial linear combination of $x_1, \dots, x_n$ equals the zero vector. That is, there are some scalars $c_1, \dots, c_n$ *not all zero* such that $\sum_{k=1}^{n} c_k x_k = 0$. Then we have

$$\sum_{k=1}^{n} c_k x_k \;=\; 0 \;=\; \sum_{k=1}^{n} 0 x_k.$$

Since it's not true that $c_k = 0$ for every k, the two linear combinations $\sum_{k=1}^{n} c_k x_k$ and $\sum_{k=1}^{n} 0x_k$ are different linear combinations. Hence we have shown that there are two different linear combinations of $x_1, \dots, x_n$ that are equal. □

In summary, there are no duplicates in the set of linear combinations if and only if there is no nontrivial linear combination that equals the zero vector. We have a special name for this situation.

Definition 3.58 (Linear Independence). Let $x_1, \dots, x_n$ be finitely many vectors in V.

(a) We say that $\{x_1, \dots, x_n\}$ is a *linearly independent set of vectors* if there is no nontrivial linear combination that equals the zero vector.

Writing this definition in contrapositive form, $\{x_1, \dots, x_n\}$ is linearly independent if the only linear combination that equals the zero vector is the trivial combination. This is usually the best form to work with. In symbols, $\{x_1, \dots, x_n\}$ *is linearly independent if*

$$\sum_{k=1}^{n} c_k x_k \;=\; 0 \quad\Longrightarrow\quad c_1 \;=\; \cdots \;=\; c_n \;=\; 0.$$

We use several different abbreviations for independence. For example, we might say that $x_1, \dots, x_n$ *are linearly independent* or we might just write that $x_1, \dots, x_n$ *are independent.*

(b) We say that $\{x_1, \dots, x_n\}$ is a *linearly dependent set of vectors* if it is not linearly independent. Hence $x_1, \dots, x_n$ are linearly dependent if and only if there is at least one nontrivial linear combination that equals the zero vector. In other words, $x_1, \dots, x_n$ are dependent if

$$\sum_{k=1}^{n} c_k x_k \;=\; 0 \quad \text{for some } c_k \text{ that are NOT ALL ZERO.} \qquad \diamondsuit$$

Thus, $x_1, \dots, x_n$ are linearly independent if and only if each linear combination of $x_1, \dots, x_n$ gives us a unique vector. Another way to say this is that if $x_1, \dots, x_n$ are linearly independent, then when we look at their span, which is

$$\operatorname{span}\{x_1, \dots, x_n\} \;=\; \biggl\{\sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R}\biggr\},$$

there are no duplicates in the linear combinations that make this span. The span is the set of all linear combinations, and our vectors are independent if there are no duplications among this linear combinations.

Note that we *always* have $0x_1 + \cdots + 0x_n = 0$. This does NOT say that $x_1, \dots, x_n$ are dependent. In order to be dependent there must be a *nontrivial* linear combination that equals the zero vector.

If you want to prove that $x_1, \dots, x_n$ are independent, then what you need to prove is this implication:

$$\sum_{k=1}^{n} c_k x_k = 0 \implies c_1 = \cdots = c_n = 0.$$

To do this, you suppose that you have some scalars c_k such that $\sum_{k=1}^{n} c_k x_k = 0$, and then you must show that each c_k must be zero.

Here is a good exercise to start with.

Exercise 3.59. Let f and g be the functions whose rules are $f(x) = \sin x$ and $g(x) = \cos x$. These are vectors in the space $\mathcal{F}(\mathbf{R})$. Determine whether $\{f, g\}$ is an independent or a dependent set of vectors.

To solve this problem, you have to determine whether there is some nontrivial linear combination of f and g that equals the zero vector. If you believe that f and g are independent, then you have to show that there are no scalars a, $b \in \mathbf{R}$ with a, b not both zero such that $af + bg = 0$ (the zero function). If you believe that $\{f, g\}$ is dependent, then you have to prove that there are scalars a, b that are not both zero but such that $af + bg = 0$.

Let's suppose that you believe that f, g are independent. Then you begin your proof with "Suppose that $af + bg = 0$ for some scalars a, b". Then you use this information. By definition, the assumption $af + bg$ equals the zero function means that these two functions have the same rule, i.e., $(af + bg)(x)$ and $0(x)$ are equal for every x. Plugging in the definitions of f, g, and the zero function, it follows that *for every* $x \in \mathbf{R}$ *we have*

$$\begin{aligned} a \sin x + b \cos x &= af(x) + bg(x) \\ &= (af + bg)(x) \\ &= 0(x) \\ &= 0. \end{aligned}$$

Now you have to prove, somehow, that a and b must be zero. You could try choosing some specific values of x; you can do this because you know that $a \sin x + b \cos x = 0$ for every x, so if you look at any particular value of x then you will have $a \sin x + b \cos x = 0$. For example, you could try a point like $x = 0$. Or you can use any other tools that apply to functions—you could try to differentiate, or take limits, or do something else. Each problem is different, you have to find the tool that will work for your problem. $\diamondsuit$

Here are some exercises about important special cases of vectors that are independent or dependent.

Exercise 3.60. (a) Suppose that $S = \{x_1, \dots, x_n\}$ includes the zero vector, i.e., there is at least one j such that $x_j = 0$. Show that S is *dependent*.

(b) In this part we consider sets that contain only a single vector. Let x be a vector in V. Show that if $x = 0$ then $S = \{x\} = \{0\}$ is a linearly dependent set, Show that if $x \neq 0$ then $S = \{x\}$ is a linearly independent set.

(c) Now we consider sets of two vectors. Let x and y be two vectors in V (they can be any two vectors, they might even be equal). Show that $S = \{x, y\}$ is a linearly dependent set if and only if one of these vectors is a scalar multiple of the other. That is, you must prove that

$$\{x, y\} \text{ is dependent} \qquad \Longleftrightarrow \qquad x = cy \text{ or } y = cx \text{ for some scalar } c \in \mathbf{R}.$$

Is it true that $\{x, y\}$ is dependent if and only if $x = cy$ for some scalar c? (No, it's not true. Why not? Always worry about special cases!)

(d) In part (a) you showed that a set that contains the zero vector is linearly dependent. Prove that the converse of part (a) is FALSE. That is, give an example of a set of vectors that is dependent even though *every vector in the set is nonzero*.

(e) Give a concrete example of three nonzero vectors x, y, z such that $\{x, y, z\}$ is dependent and no two of x, y, z are equal. (Choose your favorite vector space V for this example.)

(f) In part (c) you showed that a set of two vectors is dependent if and only if one of the vectors is a scalar multiple of the others. Show that the analogous statement for three vectors is FALSE. More specifically, prove that one direction of the implication is still valid for three vectors but the other direction is not. Even more specifically, this means that you should show that if any one of x, y, z is a scalar multiple of one of the other two, then $\{x, y, z\}$ is dependent, BUT there exist examples of dependent sets $\{x, y, z\}$ where none of the three vectors is a scalar multiple of the others. $\diamondsuit$

Here are some exercises related to independence and span of subsets and supersets.

Exercise 3.61. (a) Suppose $x_1, \dots, x_n$ are linearly independent vectors in V. Show that if $1 \leq k \leq n$, then $x_1, \dots, x_k$ are independent.

(b) Is it true that if $x_1, \dots, x_n$ are linearly independent vectors in V and we choose more vectors $x_{n+1}, \dots, x_k$ then $x_1, \dots, x_n, x_{n+1}, \dots, x_k$ must be linearly independent? Either prove that this is true or give a counterexample.

(c) Suppose that vectors $x_1, \dots, x_n$ span V. Show that if we choose more vectors $x_{n+1}, \dots, x_k \in V$ then $x_1, \dots, x_n, x_{n+1}, \dots, x_k$ will span V.

(d) Is it true that if $x_1, \dots, x_n$ span V and we fix $1 \leq k \leq n$, then $x_1, \dots, x_k$ will span V? Either prove that this is true or give a counterexample. $\diamondsuit$

We have been concentrating on independence of finite sets of vectors. There are other cases to consider. First, what do we do if we have a set of no vectors? Since there is no nontrivial linear combination of vectors in the empty set that equals 0, it makes sense to say that the empty set is independent. We make this a definition.

Definition 3.62. The empty set is declared to be a linearly independent subset of V. $\diamondsuit$

The set $\{0\}$ that contains the zero vector is not the empty set. That is, $\{0\} \neq \emptyset$. The empty set $\emptyset$ is an *independent* set of vectors, while $\{0\}$ is a *dependent* set of vectors (why?).

The other extreme is infinite sets of vectors. As we discussed before, we are not allowed to use infinite series when we form linear combinations. Therefore, we define an infinite set of vectors to be independent if any finite subset is independent. Here is the definition.

Definition 3.63. Let S be any subset of a vector space V. Then we say that S is a *linearly independent set of vectors* if whenever we choose n different vectors $x_1, \dots, x_n$ from S, the set $\{x_1, \dots, x_n\}$ is independent. $\diamondsuit$

If it is possible to write our infinite set as a list, i.e., we can write $S = \{x_1, x_2, \dots\}$, then S is independent if and only if $\{x_1, \dots, x_n\}$ is independent for *every* value of n. We'll come back to this, but here is an example that you should think about. For each integer $k = 0, 1, 2, \dots$, let p_k be the function whose rule is $p_k(x) = x^k$. This is a polynomial, so it belongs to the set $\mathcal{P}$ of all polynomials. Let $S = \{p_0, p_1, p_2, \dots\}$. Is S a linearly independent set of vectors?

3.7.2 Span, Independent, Injectivity, and Surjectivity

Now I want to relate the issues of spanning and independence back to things we learned earlier, namely injectivity and surjectivity. We'll see that the question of whether a set of vectors spans V is just another way of asking whether a certain function is surjective, and likewise independence is really a question about injectivity of some appropriate function.

To see where this function comes from, remember that both spanning and independence are questions about linear combinations. A linear combination of vectors $x_1, \dots, x_n$ is a vector of the form

$$\sum_{k=1}^{n} c_k x_k \;=\; c_1 x_1 + \cdots + c_n x_n.$$

Each choice of scalars $c_1, \dots, c_n$ gives you linear combination. Some of these choices might give you the same vector—we said that if that happens, then we call $x_1, \dots, x_n$ a dependent set, while if every choice of scalars $c_1, \dots, c_n$ gives you a different vector then $x_1, \dots, x_n$ are independent. So independence is connected to how the linear combination $c_1x_1 + \cdots + c_nx_n$ relates to the scalars $c_1, \dots, c_n$. We can think of this in terms of inputs and outputs. You input scalars $c_1, \dots, c_n$ and you output a linear combination $c_1x_1+\cdots+c_nx_n$. If every input gives you a different output then you have linear independence. Of course, there are n scalars $c_1, \dots, c_n$ and not just one input, but you can think of this as just being an input of one vector $c = (c_1, \dots, c_n)$ from $\mathbf{R}^n$ instead of an input of n scalars from $\mathbf{R}$. Each vector $c = (c_1, \dots, c_n)$ in $\mathbf{R}^n$ gives you one choice of scalars $c_1, \dots, c_n$, and we can associate that with input vector c with the output $c_1x_1 + \cdots + c_nx_n$. Inputs go to outputs—we have a function. Here's the theorem.

Theorem 3.64. *Let $x_1, \dots, x_n$ be finitely many vectors in V. Define a function $T\colon \mathbf{R}^n \to V$ by the rule*

$$T(c) \;=\; \sum_{k=1}^{n} c_k x_k, \qquad \text{for } c = (c_1, \dots, c_n) \in \mathbf{R}^n. \tag{3.7}$$

Then

$$T \text{ is injective} \quad \Longleftrightarrow \quad x_1, \dots, x_n \text{ are independent.}$$

Proof. $\Rightarrow$. Suppose that T is injective. We must show that $x_1, \dots, x_n$ are independent. To do this, we must show that the only linear combination that equals the zero vector is the trivial linear combination. So, suppose that we have a linear combination that equals the zero vector, i.e., suppose there are some scalars $c_1, \dots, c_n$ such that

$$\sum_{k=1}^{n} c_k x_k \;=\; 0.$$

Somehow we must show that each scalar c_k is zero.

Let $c = (c_1, \dots, c_n)$. Then c is a vector in $\mathbf{R}^n$, which is the domain of T, and by the definition of T we have

$$T(c) \;=\; \sum_{k=1}^{n} c_k x_k \;=\; 0.$$

(Note that the 0 on the line above is the zero vector in V.)

Now, the zero vector is a vector in $\mathbf{R}^n$, so T must map it somewhere. What is the image of the zero vector under T? The zero vector in $\mathbf{R}^n$ is the vector $0 = (0, \dots, 0)$. Applying T to this vector, we have by definition of T that

$$T(0) \;=\; \sum_{k=1}^{n} 0 x_k \;=\; 0.$$

(Note that there are three different types of 0 on the preceding line! One is the zero vector in $\mathbf{R}^n$, one is the number 0, and one is the zero vector in V.)

Thus, we have shown that

$$T(c) = 0 = T(0).$$

But our function T is injective, so this implies that $c = 0$. That is, the vector $c = (c_1, \dots, c_n)$ in $\mathbf{R}^n$ equals the zero vector $0 = (0, \dots, 0)$ in $\mathbf{R}^n$. By definition, this means that $c_1 = 0$, $c_2 = 0$, ..., $c_n = 0$. Hence the linear combination $\sum_{k=1}^{n} c_k x_k$ is indeed the trivial linear combination.

In summary, we've shown that the only linear combination of $x_1, \dots, x_n$ that equals the zero vector is the trivial linear combination, so we conclude that $x_1, \dots, x_n$ are linearly independent vectors.

$\Leftarrow$. Now it's your turn. Suppose that $x_1, \dots, x_n$ are linearly independent, and prove that T is injective. To do this you must show that *if* $T(c) = T(d)$ *for some vectors* c, $d \in \mathbf{R}^n$, *then* $c = d$. So, you start by supposing that you have $T(c) = T(d)$ for some c, $d \in \mathbf{R}^n$. Then you start working and keep going until you have shown that $c = d$. $\square$

Spanning is also an issue about the function T. By definition, vectors $x_1, \dots, x_n$ span V if the set of all linear combinations of $x_1, \dots, x_n$ equals the entire space V. Each linear combination is an output of the function T, so spanning is related to whether the set of outputs of T equals the set V. Here is what you have to prove.

Exercise 3.65. Let $x_1, \dots, x_n$ be finitely many vectors in V. Let $T\colon \mathbf{R}^n \to V$ be defined just as in Theorem 3.64.

(a) Prove that

$$\operatorname{range}(T) = \operatorname{span}\{x_1, \dots, x_n\}.$$

Remember that to prove that two sets are equal, you must show that each element of the first set belongs to the second, and vice versa.

(b) Prove that

$$T \text{ is surjective} \iff x_1, \dots, x_n \text{ span } V. \qquad \diamondsuit$$

Here is some additional terminology, and then some further exercises about the function T. We will assume in this definition and exercises that $x_1, \dots, x_n$ are some finitely many vectors from V, and the function$T$ is defined by the rule given in equation (3.7).

Definition 3.66. The *kernel* or *nullspace* of T is

$$\ker(T) = \{c \in \mathbf{R}^n : T(c) = 0\}. \qquad \diamondsuit$$

Exercise 3.67. Prove that T is a *linear function*. This means that you have to prove that

$$\forall\, x, y \in V, \quad \forall\, a, b, \in \mathbf{R}, \quad T(ax + by) \;=\; aT(x) + bT(y). \qquad \diamondsuit$$

Exercise 3.68. Prove that the following three statements are equivalent.

(a) $x_1, \dots, x_n$ are linearly independent.

(b) T is injective.

(c) $\ker(T) = \{0\}$. $\diamondsuit$

When we say that that the three statements (a), (b), (c) are equivalent, we mean that if any one of them is true then the other two are also true. Essentially, there are *six* implications that need to be proved:

$$\text{(a)} \Rightarrow \text{(b)}, \quad \text{(b)} \Rightarrow \text{(a)}, \quad \text{(a)} \Rightarrow \text{(c)}, \quad \text{(c)} \Rightarrow \text{(a)}, \quad \text{(b)} \Rightarrow \text{(c)}, \quad \text{(c)} \Rightarrow \text{(b)}.$$

However, you don't actually have to write six proofs. For example, if you proved the *three* implications

$$(a) \Rightarrow (b), \quad (b) \Rightarrow (c), \quad (c) \Rightarrow (a),$$

then you would be done, because by combining these three implications you can get the missing ones. For example, if we have proved (a) $\Rightarrow$ (b) and (b) $\Rightarrow$ (c), then by combining these two implications we get (a) $\Rightarrow$ (c) for free. Just make sure that you prove enough implications that all of the six possible implications follow.

Here's another exercise that also asks you to prove that three statements are equivalent.

Exercise 3.69. Prove that the following three statements are equivalent.

(a) $x_1, \dots, x_n$ span V.

(b) T is surjective.

(c) $\mathrm{range}(T) = V$. $\diamondsuit$

You should also formulate and prove a theorem that gives statements that are equivalent to "T is a bijection".

3.7.3 Examples and Exercises

Here are some exercises for you to work.

Exercise 3.70. Let $n \in \mathbf{N}$ be a fixed positive integer. Let $\mathcal{M}_{n\times n}$ be the vector space of all $n \times n$ matrices.

(a) Find a set of vectors that spans $\mathcal{M}_{n\times n}$.
Hint: Try the particular case $n = 2$ first.

(b) Find a set of vectors that spans $\mathcal{M}_{n\times n}$ and is linearly independent.
Remark: A set that both spans and is linearly independent is called a *basis* for the space, and the number of vectors in the basis is the *dimension* of the space. What is the dimension of $\mathcal{M}_{n\times n}$? ◇

Exercise 3.71. Define

$$V = \left\{ x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \mathbf{R}^4 : x_1 + x_2 + x_3 + x_4 = 0 \right\}.$$

Find three vectors v_1, v_2, $v_3 \in \mathbf{R}^4$ such that $V = \operatorname{span}\{v_1, v_2, v_3\}$. Determine whether your vectors are linearly independent or not. ◇

Exercise 3.72. Let $\mathcal{S}$ be the vector space of all infinite sequences of real numbers that was introduced in Exercise 3.23. For each integer $n = 1, 2, \dots,$ let

$$e_n = (0, \dots, 0, 1, 0, 0, \dots),$$

where the 1 is located in the nth component. Let $\mathcal{E} = \{e_1, e_2, \dots\}$. Either show that $\operatorname{span}(\mathcal{E}) = \mathcal{S}$ or find a vector in $\mathcal{S}$ that is not in $\operatorname{span}(\mathcal{E})$. If $\operatorname{span}(\mathcal{E})$ is not equal to all of $\mathcal{S}$, then find an explicit description of $\operatorname{span}(\mathcal{E})$, i.e., tell me exactly which vectors x are in $\operatorname{span}(\mathcal{E})$ without appealing to spans or linear combinations. ◇

Exercise 3.73. Let $\mathcal{M}_{n\times n}$ be the vector space of all $n\times n$ matrices. An $n\times n$ matrix A is said to be *skew-symmetric* if the transpose of A equals the matrix $-A$, i.e., if $A^{\mathrm{T}} = -A$.

(a) Give an example of each of the following, or explain why no such matrix exists: i. a diagonal 3×3 skew-symmetric matrix; ii. a non-diagonal 3×3 skew-symmetric matrix; iii. a 3×3 skew-symmetric matrix with all nonzero entries.

Remark: A diagonal matrix is a matrix such that all the entries off of the diagonal are zero. In particular, the zero matrix is a diagonal matrix.

(b) Let

$$S = \{A \in \mathcal{M}_{n\times n} : A \text{ is skew-symmetric}\}.$$

Is S a subspace of $\mathcal{M}_{n\times n}$? Hint: $(A + B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$.

(c) If S is a subspace, then find a set of vectors that spans S.
Hint: Try to do the specific cases $n = 2$ or $n = 3$ first.

(d) Find a basis for S, i.e., a set of vectors that both spans S and is linearly independent. The dimension of S is the number of vectors in a basis for S. What is the dimension of S? ♢

Exercise 3.74. For each integer $n \in \mathbf{N}$, let f_n be the function whose rule is

$$f_n(x) \;=\; e^{nx}, \qquad x \in \mathbf{R}.$$

(a) Given any finite integer $n \in \mathbf{N}$, show that $\{f_1, \dots, f_n\}$ is a linearly independent set of vectors in $\mathcal{F}(\mathbf{R})$.

(b) Use part (a) to prove that $\mathcal{F}(\mathbf{R})$ is not finite-dimensional. ♢

Exercise 3.75. Let $n \in \mathbf{N}$ be a fixed positive integer. Let $\mathcal{P}$ be the vector space of all polynomials, and $\mathcal{P}_n$ be the vector space of all polynomials whose degree is at most n. Let

$$S \;=\; \bigl\{p \in \mathcal{P}_n : p(0) = 0\bigr\}.$$

(a) Prove that S is a subspace of $\mathcal{P}_n$.

(b) Find a set of vectors that spans S. Hint: Try to do the specific cases $n = 1$ or $n = 2$ first.

(c) Find a basis for S, i.e., a set of vectors that both spans S and is linearly independent. The dimension of S is the number of vectors in a basis for S. What is the dimension of S? ♢

3.7.4 Too many vectors are dependent

One of the things that we will study in the next section is *bases*, which are sets of vectors that both span the vector space *and* are linearly independent. But first we need to prove a very fundamental theorem. Basically this theorem says that if we choose too many vectors in a space, then that set of vectors will be dependent. More precisely, if we know that our space is spanned by a set of k vectors, then any set of more than k vectors in that space will be dependent.

The proof of this theorem will take more work than any proof that we've done so far. Make sure that you can understand how the proof progresses line by line. At this point, there's not much chance that you could think of such a proof on your own, but you can *understand* the proof. Work through, line by line, and make sure that you believe that each step is a logical consequence of what has been done before.

Here's a more detailed introduction to the theorem. Suppose that we have k vectors, say $x_1, \dots, x_k$. The span of these vectors is a subspace. We want to know how many independent vectors can we choose from $\operatorname{span}\{x_1, \dots, x_k\}$. The theorem will show that as soon as we choose $k+1$ vectors $y_1, \dots, y_k, y_{k+1}$

from this span then we must have a *dependent* set. The y_i vectors might be different from the x_j vectors, the only thing that we require is that the y_i vectors lie in the span of $x_1, \dots, x_k$. By iterating the theorem, it will follow as a corollary that if we choose more than k vectors from $\text{span}\{x_1, \dots, x_k\}$ then we must have an dependent set.

Theorem 3.76. *Let $S = \{x_1, \dots, x_k\}$ be a set of finitely many vectors in a vector space V. Then every set of $k+1$ vectors in* $\text{span}(S)$ *is dependent. That is,*

$$y_1, \dots, y_k, y_{k+1} \in S \quad \Longrightarrow \quad \{y_1, \dots, y_k, y_{k+1}\} \text{ is dependent.}$$

Proof. The proof will be by induction on the index k.

Base Step $k = 1$. Often (but not always!), the base step of a proof by induction is easy, and this is certainly true for this proof. If $k = 1$, then our set consists of a single vector. Let's just call this vector x instead of x_1, so our set S is simply $S = \{x\}$. Since we have only a single vector, its span is just the set of all scalar multiples of x:

$$\text{span}(S) = \text{span}\{x\} = \{cx : c \in \mathbf{R}\}.$$

Our task is to show that if we choose any *two* vectors from this span, say $y_1, y_2 \in \text{span}\{x\}$, then we will have a dependent set. Since y_1 and y_2 must each be multiples of x, we have $y_1 = ax$ and $y_2 = bx$ for some real numbers a and b. If you haven't done this already, you should prove now that a set that consists of scalar multiples of a single vector is dependent. Hence $\{y_1, y_2\}$ is a dependent set of vectors, and therefore the conclusion of the theorem is valid when we have $k = 1$.

Inductive Step. Now we will show that *if* the theorem is true for some value of k, *then* it is also true for the value $k+1$. Note that we're not directly showing that the theorem is true for k. We're only showing that *if it is* true for some k then it is true for the next integer $k+1$. To make things a little notationally easier, I'm actually going to show that we can go from $k-1$ to k.

So, suppose that the conclusion of the theorem is valid for the value $k-1$. That is, we suppose that the following *inductive hypothesis* is true:

any choice of k vectors in the span of $k-1$ vectors must be dependent.

Writing this in symbols, and choosing letters different from y_i and x_j to avoid confusion, we are assuming that the following statement is true:

$$z_1, \dots, z_k \in \text{span}\{w_1, \dots, w_{k-1}\} \quad \Longrightarrow \quad z_1, \dots, z_k \text{ are dependent.} \tag{3.8}$$

The previous line is our inductive hypothesis. Assuming that the inductive hypothesis is true, we must prove that a similar statement holds for the next integer. That is, we must prove that if $y_1, \dots, y_{k+1} \in \text{span}\{x_1, \dots, x_k\}$, then $y_1, \dots, y_{k+1}$ are dependent. This is an if-then statement, so we suppose that we have $k+1$ vectors

$$y_1, \ldots, y_k, y_{k+1} \in \text{span}\{x_1, \ldots, x_k\},$$

and our goal is to prove that $y_1, \ldots, y_{k+1}$ are dependent.

It will get a little notationally ugly because we have so many vectors to work with. Let's look at the vector y_1. We know that y_1 belongs to span$\{x_1, \ldots, x_k\}$. By definition, this means that y_1 is a linear combination of $x_1, \ldots, x_k$. Therefore, there are some scalars $c_1, \ldots, c_k$ such that

$$y_1 = c_1x_1 + c_2x_2 + \cdots + c_kx_k.$$

Unfortunately, we are going to want to write a similar equation for each of the vectors $y_2, \ldots, y_{k+1}$. Each one of these vectors will also be a linear combination of $x_1, \ldots, x_k$, but the scalars will be different. So we need some way to tell which scalars belong to which vector. We'll solve this by using double subscripts. That is, we will write

$$y_1 = a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,k}x_k,$$

and we will have a similar equation for each the vectors $y_2, \ldots, y_k, y_{k+1}$. Writing all this out, there are scalars $a_{i,j}$ such that

$$\begin{aligned} y_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,k}x_k, \\ y_2 &= a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,k}x_k, \\ &\vdots \\ y_{k+1} &= a_{k+1,1}x_1 + a_{k+1,2}x_2 + \cdots + a_{k+1,k}x_k. \end{aligned}$$

Note that the scalar $a_{i,j}$ is located in the ith row of the preceding set of equations, and furthermore it is in the jth column on the right of the equals sign.

Now we will split into cases. Remember that our goal is to prove that $y_1, \ldots, y_{k+1}$ is a dependent set of vectors.

Case 1: First column is all zeros.

This is the easy case. If the first columns of scalars is all zero (meaning $a_{1,1}, a_{2,1}, \ldots, a_{k+1,1}$ are all zero), then our system of equations from above reduces to the following system:

$$\begin{aligned} y_1 &= a_{1,2}x_2 + \cdots + a_{1,k}x_k, \\ y_2 &= a_{2,2}x_2 + \cdots + a_{2,k}x_k, \\ &\vdots \\ y_{k+1} &= a_{k+1,2}x_2 + \cdots + a_{k+1,k}x_k. \end{aligned}$$

This tells us that each vector y_i is a linear combination of the vectors $x_2, \ldots, x_k$. In other words, we have

$$y_1, y_2, \ldots, y_{k+1} \in \operatorname{span}\{x_2, \ldots, x_k\}.$$

Thus we have $k+1$ vectors in the span of $k-1$ vectors. Our induction hypothesis tells us that any set of k vectors in $\operatorname{span}\{x_2, \ldots, x_k\}$ must be dependent. Therefore $y_1, y_2, \ldots, y_k$ is a set of dependent vectors. Since these vectors are dependent, any larger collection of vectors is dependent as well, i.e., $y_1, y_2, \ldots, y_k, y_{k+1}$ is dependent (WHY? Prove it!). Therefore we're done with the proof in this case.

Case 2: The first column is not all zeros.

This is the harder case. Somehow we have to find a way to use our inductive hypothesis, which involves $k-1$ vectors instead of k vectors. In Case 1 this was easy because all the coefficients in front of x_1 were zero, so in effect that vector just dropped out and left us with the $k-1$ vectors $x_2, \ldots, x_k$. Now we're not so lucky.

Remember our system of equations:

$$\begin{aligned} y_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,k}x_k, \\ y_2 &= a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,k}x_k, \\ &\vdots \\ y_{k+1} &= a_{k+1,1}x_1 + a_{k+1,2}x_2 + \cdots + a_{k+1,k}x_k. \end{aligned} \tag{3.9}$$

Our assumption now is that $a_{1,1}, \ldots, a_{k+1,1}$ are not *all* zero. So at least one of these scalars must be nonzero. If it is $a_{i,1}$, then we can make our lives simpler by interchanging the first and ith rows. In other words, just switch the names of y_1 and y_i, and corresponding switch the names of the scalars in those rows. This does not change anything except what we call the vectors, and it makes things simpler because now we have the first scalar in the first row nonzero. So our first row is

$$y_1 = a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,k}x_k, \tag{3.10}$$

and the scalar $a_{1,1}$ is nonzero. Since $a_{1,1} \neq 0$, we can multiply both sides of equation (3.10) by the scalar $a_{2,1}/a_{1,1}$. This gives us a rather ugly equation:

$$\frac{a_{2,1}}{a_{1,1}}y_1 = \frac{a_{2,1}}{a_{1,1}}\,a_{1,1}x_1 + \frac{a_{2,1}}{a_{1,1}}\,a_{1,2}x_2 + \cdots + \frac{a_{2,1}}{a_{1,1}}\,a_{1,k}x_k.$$

The good news is that the scalar is front of x_1 is not so bad, it's just $a_{2,1}$. The other scalars are not very nice, but we don't really care what they are, so let's just rename them as follows:

$$d_2y_1 = a_{2,1}x_1 + b_{2,2}x_2 + \cdots + b_{2,k}x_k.$$

Similarly, if we multiply both sides of equation (3.10) by $a_{3,1}/a_{1,1}$ and simplify, then we get an equation of the form

$$d_3 y_1 = a_{3,1}x_1 + b_{3,2}x_2 + \cdots + b_{3,k}x_k.$$

Doing this over and over, we get the following system of equations:

$$\begin{aligned} y_1 &= a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,k}x_k, \\ d_2 y_1 &= a_{2,1}x_1 + b_{2,2}x_2 + \cdots + b_{2,k}x_k, \\ &\vdots \\ d_{k+1} y_1 &= a_{k+1,1}x_1 + b_{k+1,2}x_2 + \cdots + b_{k+1,k}x_k. \end{aligned} \tag{3.11}$$

Note that in the system above, we have a multiple of y_1 on the left side of the equals sign *in every row.* In contrast, the system we had before had $y_1, y_2, \ldots, y_{k+1}$ on the left of the equals sign.

So now we have two similar, but not identical, systems of equations. We subtract the second system, given in equation (3.11), from the first, given in (3.9). Each system has identical first rows, so the first row just drops out. However, they are different from the second row onwards, so they don't entirely cancel. But the *first column* on the right of the equals sign is the same in both systems, so that first column does cancel out when we subtract them. Here's what we are left with:

$$\begin{aligned} y_2 - d_2 y_1 &= (a_{2,2} - b_{2,2})x_2 + \cdots + (a_{2,k} - b_{2,k})x_k, \\ &\vdots \\ y_{k+1} - d_{k+1} y_1 &= (a_{k+1,2} - b_{k+1,2})x_2 + \cdots + (a_{k+1,k} - b_{k+1,k})x_k. \end{aligned}$$

There are k equations, each with a vector on the left-hand side and a linear combination on the right-hand side. It doesn't really matter what the scalars in this system actually are. Instead, what is important is that each of the vectors $y_2 - d_2 y_1, \ \ldots, \ y_{k+1} - d_{k+1} y_1$ is a linear combination of $x_2, \ldots, x_k$. That is,

$$y_2 - d_2 y_1, \ \ldots, \ y_{k+1} - d_{k+1} y_1 \ \in \ \operatorname{span}\{x_2, \ldots, x_k\}.$$

We have k vectors in the span of $k-1$ vectors. Our inductive hypothesis tells us that these k vectors must be dependent. Hence there is some nontrivial linear combination that equals the zero vector, i.e., there are some scalars $t_2, \ldots, t_{k+1}$ *not all zero* such that

$$t_2(y_2 - d_2 y_1) \ + \ \cdots \ + \ t_{k+1}(y_{k+1} - d_{k+1} y_1) \ = \ 0.$$

If we rearrange this equation, we get

$$(-t_2 d_2 - \cdots - t_{k+1} d_{k+1}) y_1 \ + \ t_2 y_2 \ + \ \cdots \ + \ t_{k+1} y_{k+1} \ = \ 0.$$

That is, there is a nontrivial linear combination of $y_1, y_2, \ldots, y_{k+1}$ that equals the zero vector (why is it nontrivial?). Hence $y_1, y_2, \ldots, y_{k+1}$ is a linearly dependent set of vectors. This is exactly what we needed to prove to complete the induction. □

Your task now is to extend this result to *more than k vectors* in the span instead of just $k+1$ vectors in the span.

Exercise 3.77. Assume that $x_1, \dots, x_k$ are finitely many vectors in a vector space V. Show that if $m > k$, then

$$y_1, \dots, y_m \in \operatorname{span}\{x_1, \dots, x_k\} \implies y_1, \dots, y_m \text{ are dependent.}$$

Hint: Don't make this too hard! You don't have to try to redo the proof of Theorem 3.76 to cover m instead of $k+1$ vectors. It's actually very easy, because we have already proved Theorem 3.76, and you can use that theorem instead of trying to reprove it. ♢

3.8 Bases and Dimension

So far, we have looked at sets of vectors that span, and sets of vectors that are independent. A set that has *both* of these properties will be called a basis for the space.

Definition 3.78. Let $\mathcal{B}$ be a set of vectors in a vector space V, i.e., $\mathcal{B} \subseteq V$. We say that $\mathcal{B}$ is a *basis* for V if the following two requirements are both satisfied:

(a) $\mathcal{B}$ spans V, i.e., $\operatorname{span}(\mathcal{B}) = V$, and

(b) $\mathcal{B}$ is linearly independent. ♢

Here are some examples. You've probably already done most of the work needed to prove that these are indeed bases for the given spaces.

Exercise 3.79. (a) Let d be a fixed positive integer. For $k = 1, \dots, d$, let e_k denote the vector in $\mathbf{R}^d$ that has a 1 in the kth component and zeros elsewhere:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad e_d = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Show that $\mathcal{B} = \{e_1, \dots, e_d\}$ is a basis for $\mathbf{R}^d$ (we call this the *standard basis* for $\mathbf{R}^d$).

(b) Let $\mathcal{P}$ be the vector space of all polynomials. Show that

$$\mathcal{B} = \{1, x, x^2, \dots\}$$

is a basis for $\mathcal{P}$ (we call this the *standard basis* for $\mathcal{P}$).

(c) For each integer $n = 1, 2, \dots$, let

$$e_n = (0, \dots, 0, 1, 0, 0, \dots),$$

where the 1 is located in the nth component. Let c_{00} be the space of all sequences that have only finitely many nonzero components:

$$c_{00} \;=\; \bigl\{x = (x_1, \dots, x_n, 0, 0, \dots) : n \in \mathbf{N}, x_1, \dots, x_n \in \mathbf{R}\bigr\}.$$

Prove that $\mathcal{E} = \{e_1, e_2, \dots\}$ is a basis for c_{00} (compare this to the result you obtained in Exercise 3.72).

(d) There are always many different bases for a given vector space. For each of parts (a), (b), (c) above, find a different set that is a basis for the given space. $\diamondsuit$

3.8.1 Finite Bases

Suppose that we have a vector space V, and there is a *finite* set of vectors $\mathcal{B} = \{x_1, \dots, x_n\}$ that is a basis for V. What does this tell us about the vectors in V? First, since we know that $\mathcal{B}$ spans V, we know that V equals the set of all linear combinations of $x_1, \dots, x_n$:

$$V \;=\; \mathrm{span}(B) \;=\; \mathrm{span}\{x_1, \dots, x_n\} \;=\; \biggl\{\sum_{k=1}^{n} c_k x_k : c_1, \dots, c_n \in \mathbf{R}\biggr\}.$$

Hence every vector in V can be written as $x = c_1 x_1 + \cdots + c_n x_n$ for some choice of scalars $c_1, \dots, c_n \in \mathbf{R}$. However, we know more: Because the vectors $x_1, \dots, x_n$ are independent, there is *one and only one* choice of scalars $c_1, \dots, c_n$ such that $x = c_1 x_1 + \cdots + c_n x_n$ (why is there a unique choice of scalars—prove this). This proves the following theorem.

Theorem 3.80. *If there is a finite set of vectors $\mathcal{B} = \{x_1, \dots, x_n\}$ that is a basis for a vector space V, then every vector $x \in V$ can be written*

$$x \;=\; \sum_{k=1}^{n} c_k x_k \quad \textit{for unique scalars } c_1, \dots, c_n \in \mathbf{R}. \qquad \diamondsuit$$

Thus, when we have a basis for a vector space, every vector can be written as a *unique linear combination* of the vectors in the basis.

A given vector space always has many different bases. However, we will show that if a vector space has a finite basis (a basis with finitely many vectors), then *every basis contains exactly the same number of vectors*. We will prove this, and then we will define the *dimension* of the space to be the number of vectors in a basis.

Theorem 3.81. *Suppose that there is a finite set of vectors $\mathcal{B} = \{x_1, \dots, x_n\}$ that is a basis for a vector space V. Then every basis for V must consist of precisely n vectors.*

Proof. Suppose that $\mathcal{E} = \{y_1, \dots, y_m\}$ is also a basis for V. Our goal is to show that $m = n$. We will do this by showing that the cases $m > n$ and $m < n$ are impossible.

Case $m > n$. Suppose that m was strictly larger than n. Since $\mathcal{B}$ is a basis for V, we know that it spans V. Then we have

$$y_1, \dots, y_m \;\in\; V \;=\; \operatorname{span}(\mathcal{B}) \;=\; \operatorname{span}\{x_1, \dots, x_n\}.$$

That is, $y_1, \dots, y_m$ are more than n vectors that are contained in the span of the n vectors $x_1, \dots, x_n$. It therefore follows from Exercise 3.76 that $y_1, \dots, y_m$ are dependent, which contradicts the fact that $\mathcal{E} = \{y_1, \dots, y_m\}$ is a basis. Therefore this case cannot happen.

Case $m < n$. Now suppose that m was strictly less than n. Then we have

$$x_1, \dots, x_n \;\in\; V \;=\; \operatorname{span}(\mathcal{E}) \;=\; \operatorname{span}\{y_1, \dots, y_m\}.$$

But then $x_1, \dots, x_n$ must be dependent (why?), which contradicts the fact that $\mathcal{B} = \{x_1, \dots, x_n\}$ is a basis. Therefore this case can't happen either.

Since we've ruled out the possibilities that $m > n$ or $m < n$, the only possibility left is that $m = n$. In summary, if we have a basis that has finitely many elements, then there must be exactly n vectors in this basis.

This does leave one more possibility, however—could there be an *infinite* set that is a basis for V? We will have to rule this out. Suppose that there was some infinite set S that was a basis for V. That is, $\operatorname{span}(S) = V$ and S is independent, but there are infinitely many vectors in S We have to show that this leads to a contradiction. Here's one way. Since S contains infinitely many vectors, we can choose infinitely many distinct vectors $y_1, y_2, \dots \in S$ (this may or may not be a list of all the vectors in S, but it doesn't matter if we don't get all of them, just that we have infinitely many vectors from S). Remember that $\mathcal{B} = \{x_1, \dots, x_n\}$ is a basis for V. Let m be any number strictly greater than n (for example, you could take $m = n + 1$). Then, just like one of the cases from above, we would have

$$y_1, \dots, y_m \;\in\; V \;=\; \operatorname{span}(\mathcal{B}) \;=\; \operatorname{span}\{x_1, \dots, x_n\}.$$

But then it follows from Exercise 3.77 that $y_1, \dots, y_m$ are dependent, which contradicts the fact that every finite set of vectors in S is independent. Hence this case can't happen either, i.e., there's no infinite set of vectors that is both independent and spans V. □

Now that we know that every basis for V has the same number of vectors, we call that number the dimension of the space.

Definition 3.82. Let V be a vector space.

(a) If there is a finite set of vectors $\mathcal{B}$ that is a basis for V, then we say that V is *finite-dimensional*, and the number of vectors in $\mathcal{B}$ is called the *dimension* of V. We denote this number by $\dim(V)$.

(b) If there is no finite set of vectors that is a basis for V, then we say that V is *infinite-dimensional*. ◊

3.8.2 Exercises on Finite-Dimensional Vector Spaces

Here are some exercises for you.

Exercise 3.83. (a) Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices. There are many bases for $\mathcal{M}_{m\times n}$, but what do you think is the "standard basis" for $\mathcal{M}_{m\times n}$? What is the dimension of $\mathcal{M}_{m\times n}$?

(b) An $n\times n$ matrix $A = [a_{ij}]_{i,j=1,\dots,n}$ is said to be *upper triangular* if every entry that is strictly below the diagonal of A is zero, i.e., $a_{ij} = 0$ whenever $i > j$. Let S be the set of all $n \times n$ upper triangular matrices. Find a basis for S and find the dimension of S.

(c) An $n \times n$ matrix A is *symmetric* if it equals its own transpose. That is, a symmetric matrix satisfies $A = A^{\mathrm{T}}$. In terms of entries, this means that $a_{ij} = a_{ji}$ for all i, j. Let S be the set of all symmetric $n \times n$ matrices. Find a basis for S and find the dimension of S. ◊

Exercise 3.84. Let V be a vector space.

(a) Suppose that $x_1, \dots, x_n$ are independent vectors in V, and let $S = \operatorname{span}\{x_1, \dots, x_n\}$. Find the dimension of S.

(b) Suppose that $y_1, \dots, y_m$ are any vectors in V (not necessarily independent), and let $S = \operatorname{span}\{y_1, \dots, y_m\}$. Show that $\dim(S) \le m$.

(c) Give an example that shows that it is possible to have $\dim(S) < m$ in part (b).

(d) Assume $x \in V$ is a nonzero vector, and let L be the line through x (see Notation 3.31). Find $\dim(L)$. ◊

Exercise 3.85. Let V be a vector space.

(a) Suppose that $x_1, \dots, x_k$ are independent vectors in V. Prove that $\dim(V) \ge k$.

(b) Give an example that shows that $\dim(V) > k$ is possible in part (a). Is $\dim(V) = \infty$ possible? ◊

Exercise 3.86. (a) Let V be a finite-dimensional vector space. Prove that any subspace of V is finite-dimensional. If W is a vector space and $W \supseteq V$, must W be finite-dimensional?

(b) Let V be an infinite-dimensional vector space. Prove that any vector subspace that contains V must be infinite-dimensional. Must every subspace of V be infinite-dimensional? □

Exercise 3.87. Let V be a vector space. Prove that the following two statement are equivalent.

(a) V is infinite-dimensional (according to Definition 3.82, this means that there is no finite set of vectors that is a basis for V).

(b) There is no finite set of vectors that spans V. ◇

3.8.3 A Basis for the Zero Space

Before going further, we need to consider the pesky "zero space" $\{0\}$. There are only two subsets of this space, namely the empty set and $\{0\}$. You should show that:

- $\emptyset$ is independent but does not span $\{0\}$;
- $\{0\}$ spans $\{0\}$ but is not independent.

Consequently, there is no subset of $\{0\}$ that both spans this space and is independent. Therefore, $\{0\}$ does not have a basis and hence has no dimension. However, this fact is rather inconvenient, and if we insist on saying that $\{0\}$ has no basis then we constantly will have to make exceptions in the statements of our theorems. It will be a lot easier to simply declare that the dimension of the space $\{0\}$ is zero. We make this into a definition.

Definition 3.88. We declare that the empty set $\emptyset$ is a basis for the vector space $\{0\}$, and we declare that the dimension of this space is zero, i.e.,

$$\dim(\{0\}) = 0. \qquad \diamondsuit$$

3.8.4 More Facts about Finite-Dimensional Vector Spaces

Suppose that V is a finite-dimensional vector space whose dimension is n, and we have some independent vectors $x_1, \dots, x_k \in V$, but $k < n$. Then $x_1, \dots, x_k$ cannot span V (why, exactly—give a proof!), and therefore this set of vectors is not a basis for V. However, we will prove that we can find some vectors to add onto this collection in order to give us a basis for V. A key ingredient to the proof is given in the following exercise.

Exercise 3.89. Suppose that $x_1, \dots, x_k$ are independent vectors in V, but they do not form a basis for V.

(a) Prove that $\operatorname{span}\{x_1, \dots, x_k\}$ is a proper subspace of V, i.e.,

$$\operatorname{span}\{x_1, \dots, x_k\} \neq V.$$

(b) Show that if y is any vector in V that is not in $\operatorname{span}\{x_1, \dots, x_k\}$, then $\{x_1, \dots, x_k, y\}$ is an independent set of vectors in V. ♢

In other words, the preceding exercise says that if we have some independent vectors $x_1, \dots, x_k$ that don't span the entire space V, then we can create a larger independent set of vectors $x_1, \dots, x_k, y$ just by adding on any vector y that is not in the span of $x_1, \dots, x_k$. The idea of the proof of the next theorem is that if this new set still doesn't span V, then we can add on yet another vector to make an even larger independent set, and keep doing that until we get to a set that does span V.

Theorem 3.90. *Let V be a finite-dimensional vector space. If $x_1, \dots, x_k$ are independent vectors in V that do not span V, then there exist vectors $x_{k+1}, \dots, x_n$ such that $\mathcal{B} = \{x_1, \dots, x_k, x_{k+1}, \dots, x_n\}$ is a basis for V.*

Proof. Let

$$W_k = \operatorname{span}\{x_1, \dots, x_k\}.$$

Since the vectors $x_1, \dots, x_k$ are independent and span W_k, the collection $\{x_1, \dots, x_k\}$ is a basis for W_k. Now, we can't have $W_k = V$. (Why not? Explain this!) Hence W_k is a proper subset of V, i.e., there are vectors in V that are not in W_k. Choose any vector that is in V but not in W_k, and call this vector x_{k+1}. That is, we choose a vector

$$x_{k+1} \in V \setminus W_k.$$

Exercise 3.89 tells us that $x_1, \dots, x_k, x_{k+1}$ is independent. Let

$$W_{k+1} = \operatorname{span}\{x_1, \dots, x_k, x_{k+1}\}.$$

Note that $\{x_1, \dots, x_k, x_{k+1}\}$ is a basis for W_{k+1} (why?), and therefore $\dim(W_{k+1}) = k + 1$.

There are two possibilities. The first possibility is that $W_{k+1} = V$. In this case $\{x_1, \dots, x_k, x_{k+1}\}$ is a basis for V. Since V has dimension n, this implies that $n = k + 1$. Therefore we are done, we've found a basis for V by adding vectors to our original set of independent vectors $x_1, \dots, x_k$.

The second possibility is that $W_{k+1} \neq V$. Then we have $k+1$ independent vectors $x_1, \dots, x_k, x_{k+1}$ that do not span V. Remember that we can never have more than n independent vectors in V (this is because of Theorem 3.76), so the only way that this possibility can happen is if $k + 1 \leq n$ (and in fact we'll see that we must actually have $k + 1 < n$). Since $x_1, \dots, x_k, x_{k+1}$ are independent but don't span V, we can argue just as we did above—we choose a vector

$$x_{k+2} \in V \setminus W_{k+1}.$$

Exercise 3.89 tells us that $x_1, \dots, x_k, x_{k+1}, x_{k+2}$ is an independent set of vectors (therefore, because of Theorem 3.76, we must actually have $k+2 \leq n$, and hence $k+1 < n$ just like we said). We let

$$W_{k+2} = \operatorname{span}\{x_1, \dots, x_k, x_{k+2}\}.$$

Again there are two possibilities: Either $W_{k+2} = V$ and we are done with the proof, or $W_{k+2} \neq V$ and we repeat the process to find yet another vector x_{k+3}.

Now, it's very important to observe that this process can't go on forever, because each time we do it we get one additional vector, yet we know that we can never have a set of more than n independent vectors in V. Therefore, for some integer ℓ we must get to the case $W_{k+\ell} = V$. After ℓ steps we will have $k + \ell = n$, and we will have found a set of independent vectors $x_1, \dots, x_k, x_{k+1}, \dots, x_n$ that spans V. □

The next exercise is a "two out of three" result. It says that if you know two of three possible things about a set, then you automatically know the third thing. In order to prove this kind of statement you need to write three proofs, one for each of the following implications.

$$(a) + (b) \Rightarrow (c), \qquad (a) + (c) \Rightarrow (b), \qquad (b) + (c) \Rightarrow (a).$$

Theorem 3.90 should be useful in proving this exercise.

Exercise 3.91. Let V be a finite-dimensional vector space, and let $n = \dim(V)$. Let $\mathcal{B} = \{x_1, \dots, x_k\}$ be a finite set of vectors from V. Prove that if any two of the following statements are true, then the third statement is true as well.

(a) $\mathcal{B}$ is independent.

(b) $\mathcal{B}$ spans V.

(c) $k = n$. ◇

The preceding exercise gives you several ways to prove that a set $\mathcal{B}$ is a basis for a finite-dimensional vector space. First, you could simply apply the definition and prove that $\mathcal{B}$ both spans and is independent. If you do this then you proving that statements (a) and (b) in the exercise hold, and you automatically get statement (c) for free (the basis has n vectors).

A second possibility is that you could prove that statements (a) and (c) hold. That is, you prove that $\mathcal{B}$ has n vectors, and your vectors are independent. You then automatically get statement (c) for free—you get to conclude that your set spans V without having to prove it. The catch is that you have to know what n is, i.e., you have to know the dimension of V ahead of time. If you know what $n = \dim(V)$ is, then Exercise 3.91 tells you that it's enough to find n independent vectors, because it automatically follows that those vectors must span V.

The third choice is to prove statements (b) and (c). That is, you prove that you have n vectors and these vectors span V. Again, you have to know ahead of time what $\dim(V)$ is—if you don't know this, then there's no way that you can know that you have $n = \dim(V)$ vectors.

In summary, if you know the dimension of V, then Exercise 3.91 gives you some new ways to prove that a given set is a basis. However, if you don't know the dimension of V, then the exercise isn't helpful. In that case, in order to prove that a set is a basis your only choice is to prove that the set is both independent and spans V.

Here are some related exercises.

Exercise 3.92. Let V be a finite-dimensional vector space and let $n = \dim(V)$.

(a) Show that if S is a subspace of V and $\dim(S) = n$, then $S = V$. In other words, there is only one n-dimensional subspace of an n-dimensional vector space.

(b) Let S be a subspace of V. Show that every basis for S is part of some basis for V.

(c) Let S be a subspace of V. Is it true that every basis for V contains a subset that is a basis for S? If true, then you should prove it; if false, give a counterexample. ♢

3.8.5 Maximal Spanning Sets, Minimal Independent Sets

We would like to find some additional equivalent ways to characterize bases. The following theorem says that a set is a basis if and only if it is a "maximal" independent set, i.e., it is independent, and if you add on any more vectors to the set then it becomes dependent.

Theorem 3.93. *Let $\mathcal{B} = \{x_1, \dots, x_n\}$ be a set of finitely many independent vectors in a vector space V. If no larger set of vectors that contains $x_1, \dots, x_n$ is independent, then $\mathcal{B}$ is a basis for V.*

Proof. We are told that $\{x_1, \dots, x_n\}$ is a set of independent vectors, but as soon as we add more vectors it becomes dependent. We want to prove that $\{x_1, \dots, x_n\}$ is a basis. Since we know these vectors are independent, we just have to show that they span all of V. We proceed by contradiction: Suppose that they don't span V. This means that $\text{span}\{x_1, \dots, x_n\}$ is a proper subset of V, so there exists some vector $y \in V$ that is not in $\text{span}\{x_1, \dots, x_n\}$. But then Exercise 3.89 implies that $\{x_1, \dots, x_n, y\}$ is an independent set of vectors, which contradicts the fact that no larger set of vectors is independent. □

Here is a similar exercise formulated in terms of spans.

Exercise 3.94. Show that a minimal spanning set is a basis. That is, suppose that $\{x_1, \dots, x_n\}$ spans V, but no proper subset of $\{x_1, \dots, x_n\}$ spans V, and prove that $\{x_1, \dots, x_n\}$ is a basis for V. ♢

You can use the preceding exercise to solve the following exercise.

Exercise 3.95. Show that every spanning set contains a basis. That is, suppose that $\{x_1, \dots, x_n\}$ spans V, and prove that there is some subset of $\{x_1, \dots, x_n\}$ that is a basis for V. ♢

3.8.6 Bases for Infinite-Dimensional Vector Spaces

By definition, a basis is a set that both spans and is independent. Just keep in mind that even if our basis contains infinitely many vectors, when we form linear combinations we only sum finitely many of these vectors at a time.

For example, let $\mathcal{P}$ be the set of all polynomials, and let $\mathcal{E} = \{1, x, x^2, \dots\}$. Even though we have infinitely many monomials $1, x, x^2, \dots$ to choose from, when we form linear combinations we only select finitely many of these at a time. We can choose any finite number of them, but only finitely many to use to form a given linear combination. A typical linear combination has the form

$$a_0 \cdot 1 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

for some integer $n \geq 0$ and some scalars $a_0, a_1, \dots, a_n \in \mathbf{R}$. Taking every such linear combination gives us the span of $1, x, x^2, \dots$. That is,

$$\begin{aligned} \operatorname{span}(\mathcal{E}) &= \operatorname{span}\{1, x, x^2, \dots\} \\ &= \bigl\{a_0 + a_1 x + \cdots + a_n x^n : n \geq 0, a_0, a_1, \dots, a_n \in \mathbf{R}\bigr\} \\ &= \mathcal{P}. \end{aligned}$$

The span of the set of monomials $\mathcal{E} = \{1, x, x^2, \dots\}$ is $\mathcal{P}$. Hence $\mathcal{E}$ spans $\mathcal{P}$.

Similarly, when we test for linear independence, we deal with finite linear combinations. We proved earlier that $\mathcal{E} = \{1, x, x^2, \dots\}$ is linearly independent, because there is no finite nontrivial linear combination that equals that zero polynomial. So $\mathcal{E}$ is both independent and spans $\mathcal{P}$, so it is a basis for $\mathcal{P}$.

Here are some exercises on infinite bases.

Exercise 3.96. Suppose that $\mathcal{B} = \{x_1, x_2, \dots\}$ is a basis for a vector space V. Given $x \in V$, prove that there exists a unique choice of integer $n \geq 1$ and scalars $c_1, \dots, c_n$ such that

$$x = \sum_{k=1}^{n} c_k x_k \qquad \text{and} \qquad c_n \neq 0.$$

Explain why we have to include the condition $c_n \neq 0$ in order to have a unique representation of x. ♢

Exercise 3.97. Let $e_n = (0, \dots, 0, 1, 0, 0, \dots)$ be the infinite sequence that has a zero in each component except for the nth component, where it has a 1. Let $\mathcal{E} = \{e_1, e_2, e_3, \dots\}$, and let

$$V = \operatorname{span}(\mathcal{E}) = \operatorname{span}\{e_1, e_2, e_3, \dots\}.$$

Prove that $\mathcal{E}$ is a basis for V. Give an explicit description of this set V, without reference to spans or linear combinations. ♢

Exercise 3.98. Suppose that V is an infinite-dimensional vector space. Show that there is no finite set of vectors $\{x_1, \dots, x_n\}$ that spans V. ♢

3.9 Components

If $\mathcal{B} = \{x_1, \dots, x_n\}$ is a basis for a vector space V, then every vector $x \in V$ can be written in the form

$$x = \sum_{k=1}^{n} c_k x_k = c_1 x_1 + \cdots + c_n x_n$$

for a *unique* choice of scalars $c_1, \dots, c_n$. Hence the vector x is completely determined by the scalars $c_1, \dots, c_n$ and vice versa. We give the following name to these scalars.

Definition 3.99. Assume $\mathcal{B} = \{x_1, \dots, x_n\}$ is a basis for a vector space V. The unique numbers $c_1, \dots, c_n$ that satisfy

$$x = \sum_{k=1}^{n} c_k x_k = c_1 x_1 + \cdots + c_n x_n$$

are called the *components* of x with respect to the basis $\mathcal{B}$.

Since there are n components, $(c_1, \dots, c_n)$ is a vector in $\mathbf{R}^n$. We call $(c_1, \dots, c_n)$ the *component vector* of x with respect to the basis $\mathcal{B}$, and we write

$$[x]_{\mathcal{B}} = (c_1, \dots, c_n). \qquad \diamondsuit$$

In summary, x is a vector in V, but $[x]_{\mathcal{B}}$ is the vector in $\mathbf{R}^n$ that tells us how to write x in terms of the basis $\mathcal{B}$. The components of x will depend on what basis we choose.

Example 3.100. Consider the standard basis $\mathcal{E} = \{1, x, \dots, x^n\}$ for the space $\mathcal{P}_n$ of all polynomials of degree at most n. The unique way to write a polynomial in terms of *this* basis is

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

The components of p with respect to the standard basis are $(a_0, a_1, \dots, a_n)$. This is a vector in $\mathbf{R}^{n+1}$, we write

$$[p]_{\mathcal{E}} \;=\; (a_0, a_1, \dots, a_n) \;\in\; \mathbf{R}^{n+1}.$$

Although p is a *function*, its component vector is a vector in $\mathbf{R}^{n+1}$. $\diamondsuit$

In particular, consider the polynomial $p(x) = 1 + x + x^2$. This is a polynomial in $\mathcal{P}_2$. The standard basis for $\mathcal{P}_2$ is $\mathcal{E} = \{1, x, x^2\}$. The component vector for p with respect to the standard basis is

$$[p]_{\mathcal{E}} \;=\; (1, 1, 1).$$

However, $\mathcal{B} = \{1 + x^2, x, x + x^2\}$ is also a basis for $\mathcal{P}_2$ (prove this!). Since

$$p(x) \;=\; 1 + x + x^2 \;=\; 1(1 + x^2) + 1x + 0(x + x^2),$$

the components of p with respect to the basis $\mathcal{B}$ are

$$[p]_{\mathcal{B}} \;=\; (1, 1, 0).$$

Exercise 3.101. Let

$$v_1 \;=\; \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad v_2 \;=\; \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad v_3 \;=\; \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

(a) Show that $\mathcal{B} = \{v_1, v_2, v_3\}$ is a basis for $\mathbf{R}^3$.

(b) Given an arbitrary vector $x = (a, b, c) \in \mathbf{R}^3$, find $[x]_{\mathcal{B}}$. $\diamondsuit$

3.10 Exercises

Section 3.10 in Apostol's text is another section of exercises. Some additional practice problems are given below.

3.6. Suppose that S and T are each subspaces of a vector space V.

(a) Define

$$S + T \;=\; \{x + y : x \in S, y \in T\}.$$

Prove that $S + T$ is a subspace of V.

(b) Here's a specific example to illustrate part (a). Suppose that S is the set of vectors in $\mathbf{R}^2$ that lie on the x-axis, and T is the set of vectors that lie on the y-axis. Can you write S and T in set form? Find *explicitly* what $S + T$ is.

(c) Let c be a real number, and let

$$cS \;=\; \{cx : x \in S\}.$$

Prove that cS is a subspace of V. Can you be more explicit—what is cS if $c \neq 0$, and what is cS if $c = 0$?

3.7. Let $C(\mathbf{R})$ denote the vector space of all continuous functions that map real numbers to real numbers. Let $f_1, f_2, f_3 \in C(\mathbf{R})$ be the continuous functions whose rules are

$$f_1(x) = 2x - 1, \qquad f_2(x) = x^2 - 1, \qquad f_3(x) = e^x + 1.$$

Define a function $T\colon C(\mathbf{R}) \to \mathbf{R}$ by the rule

$$T(f) \;=\; \int_0^1 f(x)\,dx, \qquad f \in C(\mathbf{R}).$$

(a) Compute $T(f_1)$, $T(f_2)$, and $T(f_3)$. Is T injective?

(b) Is $f_1 \in \operatorname{span}\{f_1, f_2, f_3\}$? Is $\{f_1, f_2, f_3\}$ is linearly independent?

(c) Let $V = \{f \in \mathbf{C}(\mathbf{R}) : T(f) = 0\}$. Prove that V is a subspace of $\mathbf{C}(\mathbf{R})$. Is $\operatorname{span}\{f_1, f_2, f_3\} \subseteq V$? Is $V \subseteq \operatorname{span}\{f_1, f_2, f_3\}$?

3.8. (a) Show that any set of independent vectors is a basis for its span.

(b) Must a set of nonzero vectors be a basis for its span?

3.9. Let $n \in \mathbf{N}$ be a fixed positive integer. Let $\mathcal{P}$ be the vector space of all polynomials, and $\mathcal{P}_n$ be the vector space of all polynomials whose degree is at most n. Let

$$S_n \;=\; \bigl\{p \in \mathcal{P}_n : p(1) = 0\bigr\}.$$

(a) Prove that S_n is a subspace of $\mathcal{P}_n$.

(b) Find a set of vectors that spans S_n. Hint: Try to do the specific cases $n = 1$ or $n = 2$ first.

(c) Find a basis for S_n. What is the dimension of S_n?

(d) Let $n = 5$ and set $q(x) = (x-1)^2\,(x^2 + x + 1)$. Show that both q and q' are in S_5.

(e) Let $\mathcal{E} = \{1, x, x^2, x^3, x^4, x^5\}$ be the standard basis for $\mathcal{P}_5$. Compute $[q]_{\mathcal{E}}$ and $[q']_{\mathcal{E}}$, the coordinates of q and q' with respect to the standard basis. Also compute $[q]_{\mathcal{B}}$ and $[q']_{\mathcal{B}}$, the coordinates of q and q' with respect to the basis $\mathcal{B}$ for S_5 that you found in part (c).

3.10. Prove that $\{1, 1+x, 1+x^2, 1+x^3, \dots\}$ is a basis for the set of all polynomials $\mathcal{P}$.

3.11. For each $n \in \mathbf{N}$, let f_n be function whose rule is $f_n(x) = \sin nx$ for $x \in \mathbf{R}$. Prove that $\{f_1, \dots, f_n\}$ is linearly independent.

Hint: You can assume that the following calculus fact is true:

$$m \neq n \implies \int_0^{2\pi} \sin mx \, \sin nx \, dx = 0.$$

3.12. Let V be a finite-dimensional vector space, and let $n = \dim(V)$.

(a) Show that V is *isomorphic* to $\mathbf{R}^n$, i.e., show that there exists a bijection $f\colon \mathbf{R}^n \to V$ such that

$$\forall\, x, y \in \mathbf{R}^n, \quad \forall\, a, b \in \mathbf{R}, \quad f(ax + by) \;=\; af(x) + bf(y).$$

Hint: Let $\mathcal{B} = \{x_1, \dots, x_n\}$ be a basis for V. Then every vector $v \in V$ can be written $v = c_1x_1 + \cdots + c_nx_n$ for a unique choice of scalars $c_1, \dots, c_n \in \mathbf{R}$.

(b) Show that the inverse function $f^{-1}\colon V \to \mathbf{R}^n$ is also an isomorphism.
Hint: We already know that f^{-1} is a bijection, so what you have to prove is that $f^{-1}(au + bv) = af^{-1}(u) + bf^{-1}(v)$ for all u, $v \in V$ and a, $b \in \mathbf{R}$.

3.11 Inner Products and Norms

We've mentioned before that we can't define infinite sums of vectors if we don't have some way to define a limit. Further, defining a limit requires us to have some way of saying how close vectors are to each other—in other words, we need to have a way to define the distance between vectors. A *norm* on a vector space lets us do that.

Definition 3.102 (Seminorms and Norms). Let V be a vector space. We call $\|\cdot\|$ a *norm* on V if for each vector $x \in V$ we can define a real number $\|x\|$ in such a way that the following hold for all all vectors x, $y \in V$ and all scalars $c \in \mathbf{R}$:

(a) $0 \le \|x\| < \infty$,

(b) $\|x\| = 0$ if and only if $x = 0$.

(b) $\|cx\| = |c|\,\|x\|$, and

(c) $\|x + y\| \le \|x\| + \|y\|$ (this is called the *Triangle Inequality*). $\diamondsuit$

We call $\|x\|$ the *length* of the vector x, and we say that

$$\|x - y\|$$

is the *distance* between the vectors x and y. A vector x that has length 1 is called a *unit vector*, or is said to be *normalized*.

3.11.1 Norms on $\mathbf{R}^n$

Here is the norm that we think of first when our vector space is $\mathbf{R}^n$.

Exercise 3.103 (Euclidean Norm). We usually define the length of a vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ by the formula

$$\|x\| \;=\; \bigl(x_1^2 + \cdots + x_n^2\bigr)^{1/2}.$$

Your exercise is to check that this formula does satisfy the definition of a norm. We call this the *Euclidean norm* on $\mathbf{R}^n$, and we call $\|x - y\|$ the *Euclidean distance* between vectors. □

However, the Euclidean norm is not the only norm on $\mathbf{R}^n$. There are infinitely many other norms on $\mathbf{R}^n$. The next exercise will give two norms on $\mathbf{R}^n$ that are commonly encountered.

Exercise 3.104. For each vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, define

$$\|x\|_1 \;=\; |x_1| + \cdots + |x_n| \qquad \text{and} \qquad \|x\|_\infty \;=\; \max\{|x_1|, \dots, |x_n|\}.$$

We call these the ℓ^1*-norm* and the ℓ^∞*-norm* of the vector x.

(a) Prove that $\|\cdot\|_1$ is a norm on $\mathbf{R}^n$.

(b) Prove that $\|\cdot\|_\infty$ is a norm on $\mathbf{R}^n$.

(c) Find positive constants A and B such that

$$A\,\|x\|_1 \;\le\; \|x\|_\infty \;\le\; B\,\|x\|_1 \qquad \text{for all } x \in \mathbf{R}^n. \tag{3.12}$$

(d) Find the best possible values for A and B. In other words, find the largest possible value of A and the smallest possible value of B so that equation (3.12) is simultaneously valid *for every vector* $x \in \mathbf{R}^n$. ♢

Here are some additional norms on $\mathbf{R}^n$. We'll state the formula for these norms, but we won't prove that they actually are norms. It's not that hard to do, but it is usually proved in a graduate real analysis course.

Example 3.105. Let p be any real number that lies in the range $1 \le p < \infty$. For each vector $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, define

$$\|x\|_p \;=\; \bigl(|x_1|^p + \cdots + |x_n|^p\bigr)^{1/p}.$$

It's easy to see that properties (a), (b), and (c) in Definition 3.102 are satisfied. However, it quite a bit trickier to prove that property (d), the Triangle Inequality, is satisfied when p is not 1, 2, or ∞ (try it!). It can be shown that the Triangle Inequality does hold, and once this is proved then we know that $\|\cdot\|_p$ is a norm on $\mathbf{R}^n$. We call this the ℓ^p*-norm* on $\mathbf{R}^n$. You'll note that for $p = 1$, this is the same as the ℓ^1-norm introduced in Exercise 3.104, and for $p = 2$ it is simply the Euclidean norm. For this reason, we will often denote the Euclidean norm using the symbols $\|\cdot\|_2$. That is, we write

$$\|x\|_2 \;=\; \bigl(x_1^2 + \cdots + x_n^2\bigr)^{1/2}. \qquad ♢$$

The preceding example discussed the ℓ^p-norms for finite p. We defined the ℓ^∞-norm in Exercise 3.104; it is given by the formula

$$\|x\|_\infty = \max_k |x_k|.$$

The next exercise shows that $\|\cdot\|_\infty$ is in some sense a limit of the norms $\|\cdot\|_p$ as $p \to \infty$.

Exercise 3.106. Show that for every vector $x \in \mathbf{R}^n$ we have

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p.$$

Hint: First prove this for vectors y that satisfy $\|y\|_\infty = 1$. Then let x be an arbitrary nonzero vector, and set $y = x/\|x\|_\infty$. $\diamondsuit$

The following exercise may give some insight into these norms.

Exercise 3.107. We will consider vectors in $\mathbf{R}^2$ for this exercise. Plot the following sets ("C" is for "circle" and B is for "ball").

(a) $C = \{x \in \mathbf{R}^2 : \|x\|_1 = 1\}$ and $B = \{x \in \mathbf{R}^2 : \|x\|_1 < 1\}$.

(b) $C = \{x \in \mathbf{R}^2 : \|x\|_2 = 1\}$ and $B = \{x \in \mathbf{R}^2 : \|x\|_2 < 1\}$.

(c) $C = \{x \in \mathbf{R}^2 : \|x\|_\infty = 1\}$ and $B = \{x \in \mathbf{R}^2 : \|x\|_\infty < 1\}$. $\diamondsuit$

3.11.2 Norms on ℓ^1, ℓ^2, and ℓ^∞

Back in Exercise 3.23, we introduced some vector spaces whose elements were infinite sequences of real numbers. First, we looked at the set of all possible infinite sequences:

$$\mathcal{S} = \big\{x : x = (x_1, x_2, \dots) \text{ where } x_1, x_2, \dots \in \mathbf{R}\big\}.$$

Then we took the subset of $\mathcal{S}$ that contains the "summable" sequences:

$$\ell^1 = \left\{x = (x_1, x_2, \dots) \in \mathcal{S} \;:\; \sum_{k=1}^\infty |x_k| < \infty\right\}.$$

For example, the sequence $x = (1, \frac{1}{4}, \frac{1}{9}, \dots)$ belongs to ℓ^1, but $x = (1, \frac{1}{2}, \frac{1}{3}, \dots)$ does not. We saw that ℓ^1 is a subspace of $\mathcal{S}$. The next exercise asks you to show that we can define a norm on ℓ^1.

Exercise 3.108. Show that

$$\|x\|_1 = \sum_{k=1}^\infty |x_k|$$

is a norm on the space ℓ^1. $\diamondsuit$

We will also consider the spaces ℓ^2 and ℓ^∞ that are defined by

$$\ell^2 = \left\{ x = (x_1, x_2, \dots) \in \mathcal{S} \ : \ \sum_{k=1}^{\infty} |x_k|^2 < \infty \right\}$$

and

$$\ell^\infty = \left\{ x = (x_1, x_2, \dots) \in \mathcal{S} \ : \ \sup_k |x_k| < \infty \right\}.$$

We have to use a sup and not a max in the definition of ℓ^∞ because a bounded sequence doesn't have to have a maximum element. For example, the sequence

$$x = \left(\tfrac{1}{2}, \tfrac{3}{4}, \tfrac{4}{5}, \dots\right)$$

belongs to ℓ^∞, but there is no maximum component. Although we won't do it, you can similarly define spaces ℓ^p for each index p in the range $1 \le p < \infty$.

Exercise 3.109. (a) Prove that

$$\|x\|_2 = \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2}$$

is a norm on ℓ^2. What is the ℓ^2-norm of the vector $x = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$?

(b) Prove that

$$\|x\|_\infty = \sup_k |x_k|$$

is a norm on ℓ^∞.

(c) Prove that $\ell^1 \subseteq \ell^2 \subseteq \ell^\infty$.

(d) Prove that $\ell^1 \subsetneq \ell^2 \subsetneq \ell^\infty$.

Hint: Since you already did part (c), all you need to do is show that there exists a vector x that belongs to ℓ^2 but doesn't belong to ℓ^1, and similarly there is a vector y that belongs to ℓ^∞ but not ℓ^2.

(e) Prove that for every vector $x \in \ell^1$ we have

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1.$$

However, prove that there is no finite real number C such that

$$\|x\|_1 \le C\|x\|_\infty \text{ for all } x \in \ell^1.$$

That is, although the inequality on the line above can hold for some particular vectors in ℓ^1, there's no way that it can hold simultaneously for *all* vectors in ℓ^1.

(f) Prove that $\|\cdot\|_2$ is a norm on the space ℓ^1, and likewise prove that $\|\cdot\|_\infty$ is a norm on ℓ^1.

(g) Prove that $\|\cdot\|_1$ is not a norm on either ℓ^2 or ℓ^∞. $\diamondsuit$

3.11.3 Norms on $C[0,1]$

We can also define norms on spaces of functions. We will define most of these in terms of integrals, so for simplicity we will restrict our attention to spaces of continuous functions, because we know from MATH 1502 that we can integrate continuous function. In other words, we know that the *Riemann integral* of any continuous function exists. The Riemann integral of *some* discontinuous functions does exist, but there are many discontinuous functions whose Riemann integral does not exist.

In graduate real analysis you will learn how to integrate functions that are not continuous. This is not as easy as it may sound—it takes a considerable amount of work to define the integral of more general functions and to derive the properties of the integral—but the result is a very powerful and useful theory of integration. This *Lebesgue integral* is the foundation of almost all of modern analysis.

However, we're not ready for the Lebesgue integral, so we'll stick to the Riemann integral, and we'll consider the vector space $C[0,1]$ that consists of all the continuous functions whose domain is the interval $[0,1]$. The next exercise gives some norms on this space.

Exercise 3.110. (a) The L^1-*norm* of a function $f \in C[0,1]$ is defined by

$$\|f\|_1 \;=\; \int_0^1 |f(x)|\,dx.$$

Prove that $\|\cdot\|_1$ is a norm on $C[0,1]$.

(b) The L^2-*norm* of a function $f \in C[0,1]$ is defined by

$$\|f\|_2 \;=\; \left(\int_0^1 |f(x)|^2\,dx\right)^{1/2} \;=\; \left(\int_0^1 f(x)^2\,dx\right)^{1/2}.$$

Prove that $\|\cdot\|_2$ is a norm on $C[0,1]$.

(c) The L^∞-*norm* of a function $f \in C[0,1]$ is defined by

$$\|f\|_\infty \;=\; \sup_{0\le x\le 1} |f(x)|.$$

Prove that $\|\cdot\|_\infty$ is a norm on $C[0,1]$.

(d) Prove that for each function $f \in C[0,1]$ we have

$$\|f\|_1 \;\le\; \|f\|_\infty.$$

Challenge: Prove that we actually have

$$\|f\|_1 \;\le\; \|f\|_2 \;\le\; \|f\|_\infty.$$

Compare this to the inequality you obtained in part (e) of Exercise 3.109!

(e) Prove that there is no finite real number C such that

$$\|f\|_\infty \;\leq\; C\,\|f\|_1 \text{ for every } f \in C[0,1].$$

Again, compare this to the part (e) of Exercise 3.109. $\diamondsuit$

Can you guess how to define the L^p-norm of a function $f \in C[0,1]$?

Exercise 3.111. (a) Find the L^1, L^2, and L^∞ norms of each of the vectors $1, x, x^2, x^3, \dots$ in $C[0,1]$.

(b) Find the L^2-distance between x^m and x^n in $C[0,1]$. That is, given integers m, $n \geq 0$, find

$$\|x^m - x^n\|_2 \;=\; \left(\int_0^1 (x^m - x^n)^2\,dx\right)^{1/2}. \qquad \diamondsuit$$

Sometimes we use domains other than $[0,1]$. In this case we just change the limits of integration appropriately. Here's an example.

Exercise 3.112. Let $f(x) = \sin x$ and $g(x) = \cos x$. These vectors belong to the space $C[0,2\pi]$. Find the Euclidean lengths of f and g in this space, i.e., find

$$\|f\|_2 \;=\; \left(\int_0^{2\pi} |f(x)|^2\,dx\right)^{1/2} \;=\; \left(\int_0^{2\pi} \sin^2 x\,dx\right)^{1/2}$$

and

$$\|g\|_2 \;=\; \left(\int_0^{2\pi} |g(x)|^2\,dx\right)^{1/2} \;=\; \left(\int_0^{2\pi} \cos^2 x\,dx\right)^{1/2}$$

Hint: First find the value of

$$\|f\|_2^2 + \|g\|_2^2.$$

Then show that $\|f\|_2^2 = \|g\|_2^2$ (make the change of variable $y = x - \frac{\pi}{2}$). $\diamondsuit$

3.11.4 Balls

You should think about what a "ball" in $C[0,1]$ looks like. Here's an exercise that might help with that.

Exercise 3.113. For this exercise we'll consider the L^∞-norm on $C[0,1]$. The *distance* between two functions f and g in this norm is $\|f-g\|_\infty$. In particular, the distance between f and the zero function 0 is the number $\|f - 0\|_\infty = \|f\|_\infty$. Let B be the set of all functions that are within distance 1 from the zero vector:

$$B \;=\; \bigl\{f \in C[0,1] : \|f\|_\infty < 1\bigr\}.$$

The letter B is for "ball," because the set B is really the open unit ball centered at the origin in the space $C[0,1]$.

(a) Try to give an "explicit" description of the functions in B. That is, how can I tell just by looking at a function f whether it belongs to the ball B?

(b) Now let g be some function in $C[0,1]$. Let B be the set of all functions that are within a distance 1 from g:

$$B \;=\; \bigl\{f \in C[0,1] : \|f-g\|_\infty < 1\bigr\}.$$

This set is the open ball of radius 1 centered at g. Give an "explicit" description of the functions in B. That is, how can I tell just by looking at a function f whether it belongs to the ball B? $\diamondsuit$

Here's the formal definition of an open ball in a normed vector space.

Definition 3.114. Let $\|\cdot\|$ be a norm on a vector space V. Given a vector $x \in V$ and given a positive real number $r > 0$, the *open ball centered at x with radius r* is the set

$$B_r(x) \;=\; \bigl\{y \in V : \|x-y\| < r\bigr\}.$$

That is, the open ball $B_r(x)$ consists of all vectors y that lie within a distance r from x. $\diamondsuit$

Once we have open balls, we can define open sets.

Definition 3.115. Let $\|\cdot\|$ be a norm on a vector space V. We say that a set $U \subseteq V$ is *open* if

$$\forall\, x \in U, \quad \exists\, r > 0 \text{ such that } B_r(x) \subseteq U.$$

That is, U is open if each point in U can be surrounded by an open ball that is entirely contained in U. $\diamondsuit$

Note that although we called an open ball "open," we haven't yet proved that it actually satisfies the requirements of Definition 3.115. That's your next exercise.

Exercise 3.116. Let $\|\cdot\|$ be a norm on a vector space V. Choose any vector $x \in V$ and any $r > 0$. Show that $B_r(x)$ is an open set.

Hint: You must show that if $y \in B_r(x)$, then there is some $s > 0$ such that $B_s(y) \subseteq B_r(x)$. To find s, pretend that you were working with vectors in $\mathbf{R}^2$, draw a picture, and try to figure out what s has to be (it will be determined both by the value of r and the value $\|x-y\|$). Because we have the Triangle Inequality, this is precisely the value of s that you will need, no matter what the vector space actually is. You still have to prove that this value of s works, i.e., you have to prove that $B_s(y)$ is a subset of $B_r(x)$. So, once you've gotten your guess for s, you must choose an arbitrary vector $z \in B_s(y)$ and prove that $z \in B_r(x)$. $\diamondsuit$

Here are some more exercises about open sets.

Exercise 3.117. Let $\|\cdot\|$ be a norm on a vector space V. Give a precise definition of what it means for a set to not be open. That is, complete the following: A set $E \subseteq V$ is not open if _________. ♢

Exercise 3.118. Let $\|\cdot\|$ be a norm on a vector space V.

(a) Suppose that I is any set, and for each $i \in I$ we have an open set $U_i \subseteq V$. Prove that the union of these open sets is open. That is, prove that

$$\bigcup_{i \in I} U_i$$

is open. Thus, *the union of any collection of open sets is open.*

(b) Prove that *the intersection of finitely many open sets* is open. That is, suppose that $U_1, \dots, U_n$ are open, and prove that

$$U_1 \cap \cdots \cap U_n \text{ is open.}$$

(c) Prove by example that the intersection of infinitely many open sets need not be open. That is, give an example of a normed vector space V and open sets $U_1, U_2, \dots$ such that

$$W = \bigcap_{k=1}^{\infty} U_k$$

is not open. Be careful—be sure to *prove* that your set W is not open, don't just say that it doesn't look like it's not open.

(d) Prove that $\emptyset$ and V are both open sets.

(e) Suppose that U is a nonempty open set. Prove that U is the union of some collection of open balls.

Hint: One open ball for each element of U. ♢

3.11.5 Inner Products

Some (but not all!) norms are associated with an *inner product*. The definition of an inner product is inspired by the properties that the dot product of vectors in $\mathbf{R}^n$ possesses. An inner product is a function of *two* vectors—for each pair of vectors it gives a real number, and this number must satisfy certain properties. Here is the definition.

Definition 3.119 (Inner Product). Let V be a vector space. We call $\langle \cdot, \cdot \rangle$ an *inner product* on V if for each pair of vectors x, $y \in V$ we can define a real number $\langle x, y \rangle$ in such a way that the following hold for all all vectors x, y, $z \in V$ and all scalars $c \in \mathbf{R}$:

(a) $0 \le \langle x, x\rangle < \infty$,
(b) $\langle x, x\rangle = 0$ if and only if $x = 0$.
(c) $\langle x, y\rangle = \langle y, x\rangle$, and
(d) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
(e) $\langle cx, y\rangle = c\langle x, y\rangle$.

A vector space that has an inner product is called an *inner product space* or a *Euclidean space* (Apostol prefers the term Euclidean space). ♢

Sometimes other symbols are used to denote an inner product. For example, Apostol likes to write (x, y) instead of $\langle x, y\rangle$. It doesn't matter what symbols are used, as long as the five properties given in the definition of an inner product are satisfied (Apostol lists the properties in a slightly different way, but you can check that they are the same properties that I give).

Some changes are need if we want to work with complex scalars instead of real scalars. In that case, the symmetry condition given in property (c) has to be changed to $\langle x, y\rangle = \overline{\langle y, x\rangle}$, i.e., a complex conjugate is introduced when the ordering is interchanged in the inner product. We will only consider real scalars.

Eventually, we will prove that every inner product gives us a norm. This associated norm will be given by the rule

$$\|x\| \;=\; \langle x, x\rangle^{1/2} \;=\; \sqrt{\langle x, x\rangle}, \qquad x \in V.$$

We will call this the *induced norm*, but you should note that we haven't yet proved that it actually is a norm!

The dot product is the prototypical example of an inner product, and we can easily see that the norm induced from the dot product really is a norm.

Exercise 3.120. (a) The dot product of two vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbf{R}^n$ is

$$x \cdot y \;=\; x_1 y_1 + \cdots + x_n y_n \;=\; \sum_{k=1}^{n} x_k y_k.$$

Show that $\langle x, y\rangle = x \cdot y$ defines an inner product on $\mathbf{R}^n$. Note that the norm induced from the dot product is

$$\|x\|_2 \;=\; (x \cdot x)^{1/2} \;=\; \bigl(x_1^2 + \cdots + x_n^2\bigr)^{1/2},$$

which is precisely the Euclidean norm on $\mathbf{R}^n$. We therefore say that *the Euclidean norm on* $\mathbf{R}^n$ *is induced from the dot product.*

(b) This part will show that the dot product is not the only inner product on $\mathbf{R}^n$. We will modify the dot product by introducing some "weights." Let $w_1, \dots, w_n > 0$ be fixed positive scalars. Show that

$$\langle x, y \rangle \;=\; w_1 x_1 y_1 + \cdots + w_n x_n y_n, \qquad x, y \in \mathbf{R}^n,$$

defines an inner product on $\mathbf{R}^n$.

(c) This part requires a little knowledge about matrices. We say that an $n \times n$ matrix A is *positive definite* if $Ax \cdot x > 0$ for each nonzero vector $x \in \mathbf{R}^n$. Prove that if A is a positive definite matrix, then

$$\langle x, y \rangle \;=\; Ax \cdot y, \qquad x, y \in \mathbf{R}^n,$$

defines an inner product on $\mathbf{R}^n$. Show that if A is a diagonal matrix that has all positive diagonal entries, then A is a positive definite matrix, and the inner product that we just defined coincides with the type of inner product given in part (b).

(d) This part is more challenging. Let $\langle \cdot, \cdot \rangle$ be an *arbitrary* inner product on $\mathbf{R}^n$. Show that there is some positive definite matrix A such that

$$\langle x, y \rangle \;=\; Ax \cdot y, \qquad \text{for all } x, y \in \mathbf{R}^n. \qquad \diamondsuit$$

You should prove that every inner product has the following properties.

Exercise 3.121. Let $\langle \cdot, \cdot \rangle$ be an inner product on a vector space V. Prove that the following statements hold.

(a) $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ for all $x,\ y,\ z \in V$ and $a,\ b \in \mathbf{R}$.

(b) $\langle x, ay + bz \rangle = a\langle x, y \rangle + b\langle x, z \rangle$ for all $x,\ y,\ z \in V$ and $a,\ b \in \mathbf{R}$.

(c) $\langle x, 0 \rangle = 0 = \langle 0, x \rangle$ for all $x \in V$.

(d) For all $x,\ y,\ z,\ w \in V$ and $a,\ b,\ c,\ d \in \mathbf{R}$ we have

$$\langle ax + by, cz + dw \rangle \;=\; ac\, \langle x, z \rangle + ad\, \langle x, w \rangle + bc\, \langle y, z \rangle + bd\, \langle y, w \rangle.$$

This is the *Distributive Law* for the inner product. $\diamondsuit$

Here is something that every induced norm must satisfy (we're doing this a bit out of order since we haven't yet proved that the induced norm actually is a norm, but that doesn't matter for this exercise).

Exercise 3.122. Let $\langle \cdot, \cdot \rangle$ be an inner product on a vector space V, and let $\|x\| = \langle x, x \rangle^{1/2}$ be the induced norm. Prove that the following equality must hold for all vectors $x,\ y \in V$:

$$\|x+y\|^2 + \|x-y\|^2 \;=\; 2\big(\|x\|^2 + \|y\|^2\big).$$

This is called the *Parallelogram Law.* $\diamondsuit$

Thus, whenever we have an inner product, the induced norm *must* satisfy the Parallelogram Law. Consequently, if we're given a norm that doesn't satisfy the Parallelogram Law, then it can't be induced from some inner product. Here's an exercise that makes this precise.

Exercise 3.123. (a) Suppose that $\|\cdot\|$ is a norm on a vector space V. Suppose that this norm does not satisfy the Parallelogram Law, i.e., there exist two vectors w, $z \in V$ such that

$$\|w+z\|^2 + \|w-z\|^2 \neq 2\big(\|w\|^2 + \|z\|^2\big).$$

Show that there is no inner product $\langle \cdot, \cdot \rangle$ on V that induces this norm. That is, there is no inner product on V that satisfies

$$\langle x, x \rangle^{1/2} = \|x\|, \qquad \text{for all } x \in V.$$

(b) Show that the ℓ^1-norm on $\mathbf{R}^n$ is not induced from an inner product. That is, show that there is no inner product on $\mathbf{R}^n$ that satisfies

$$\langle x, x \rangle^{1/2} = \|x\|_1 = |x_1| + \cdots + |x_n|, \qquad x \in \mathbf{R}^n.$$

Hint: Use part (a). Choose two "easy" vectors w, z and test whether the ℓ^1-norm satisfies the Parallelogram Law for those particular vectors.

(c) Show that the ℓ^∞-norm on $\mathbf{R}^n$ is not induced from any inner product on $\mathbf{R}^n$.

(d) Show that the ℓ^2-norm on $\mathbf{R}^n$ is induced from an inner product (it's not enough to show that the Parallelogram Law holds!).

Hint: We've already done it. $\diamondsuit$

Here is a similar exercise for the vector space ℓ^2.

Exercise 3.124. (a) Given any two infinite sequences $x = (x_1, x_2, \dots)$ and $y = (y_1, y_2, \dots)$ in ℓ^2, define

$$\langle x, y \rangle = \sum_{k=1}^{\infty} x_k y_k.$$

Show that this defines an inner product on ℓ^2. What is the induced norm—is it the ℓ^2-norm or is it something else?

Challenge: How do you know that the series defining $\langle x, y \rangle$ actually converges? We'll take this for granted for now, but it is a consequence of the *Cauchy–Schwarz Inequality*, which we will prove soon.

(b) Show that the inner product defined in part (a) is an inner product on the space ℓ^1, as well as being an inner product on ℓ^2. What is the induced norm? (Note that it's not the ℓ^1-norm.)

Hint: This is easy, because an earlier exercise showed that ℓ^1 is a subspace of ℓ^2.

(c) Show that there is no inner product on ℓ^1 whose induced norm is the ℓ^1-norm. $\diamondsuit$

Here's one more exercise, for the space $C[0,1]$.

Exercise 3.125. (a) Given functions $f,\, g \in C[0,1]$, show that

$$\langle f, g\rangle \;=\; \int_0^1 f(x)\, g(x)\, dx \tag{3.13}$$

is an inner product on $C[0,1]$. Show that the induced norm is the L^2-norm.

(b) Show that there is no inner product on $C[0,1]$ whose induced norm is the L^1-norm.

Hint: Choose two "easy" functions, but not *too* easy—don't choose constant functions, for example.

(c) Show that there is no inner product on $C[0,1]$ whose induced norm is the L^∞-norm. $\diamondsuit$

In summary, induced norms are special. Not every norm is induced from an inner product. If you have an inner product, then you'll have a norm (the induced norm), but if you're given a norm, there may or may not be an inner product associated with it.

3.11.6 Cauchy–Schwarz and the Triangle Inequality

Before we can prove that the induced norm really is a norm, there's an inequality that we have to prove first. This inequality is called the *Schwarz Inequality*, the *Cauchy–Schwarz Inequality*, or the *Cauchy–Bunyakovski–Schwarz Inequality*. The most accurate name is the longest one, though Apostol prefers to call it Cauchy–Schwarz. (Please note that there is no "t" in Schwarz's name.)

Theorem 3.126 (Cauchy–Schwarz). *If $\langle\cdot,\cdot\rangle$ is an inner product on a vector space V, then*

$$|\langle x, y\rangle| \;\le\; \|x\|\,\|y\| \;=\; \langle x, x\rangle^{1/2}\,\langle y, y\rangle^{1/2}, \qquad x, y \in V.$$

This is called the Cauchy–Schwarz Inequality*.*

Proof. Choose any particular vectors $x,\, y \in V$. If $y = 0$ then we have both $\langle x, y\rangle = 0$ and $\|y\| = 0$, so we are done in this case. Therefore, we can concentrate on the case $y \ne 0$.

Using the distributive law, for each real number c we have

$$\begin{aligned} 0 \;&\le\; \|x - cy\|^2 \\ &=\; \langle x - cy, x - cy\rangle \end{aligned}$$

$$\begin{aligned} &= \langle x, x\rangle - c\langle x, y\rangle - c\langle x, y\rangle + c^2\langle y, y\rangle \\ &= \|x\|^2 - 2c\langle x, y\rangle + c^2\,\|y\|^2. \end{aligned}$$

This is true for every real number c. In particular, it is true for the real number

$$c \;=\; \frac{\langle x, y\rangle}{\|y\|^2}.$$

Substituting this value for c, we obtain

$$\begin{aligned} 0 \;&\leq\; \|x\|^2 - 2c\langle x, y\rangle + c^2\,\|y\|^2 \\ &=\; \|x\|^2 - 2\,\frac{\langle x, y\rangle}{\|y\|^2}\,\langle x, y\rangle + \frac{\langle x, y\rangle^2}{\|y\|^4}\,\|y\|^2 \\ &=\; \|x\|^2 - \frac{\langle x, y\rangle^2}{\|y\|^2}. \end{aligned}$$

Rearranging this, we magically get

$$\langle x, y\rangle^2 \;\leq\; \|x\|^2\,\|y\|^2,$$

so the result follows by taking square roots. □

Let's see what Cauchy–Schwarz tells us about some particular vector spaces.

Example 3.127. Let $x = (x_1, x_2, \dots)$ and $y = (y_1, y_2, \dots)$ be vectors in ℓ^2. We saw earlier that

$$\langle x, y\rangle \;=\; \sum_{k=1}^{\infty} x_k y_k$$

defines an inner product on ℓ^2. The induced norm is the ℓ^2-norm:

$$\|x\|_2 \;=\; \langle x, x\rangle^{1/2} \;=\; \biggl(\sum_{k=1}^{\infty} x_k^2\biggr)^{1/2}.$$

The Cauchy–Schwarz Inequality tells us that

$$|\langle x, y\rangle| \leq \|x\|_2\,\|y\|_2.$$

Writing this out in terms of components, this is

$$\biggl|\sum_{k=1}^{\infty} x_k y_k\biggr| \;\leq\; \biggl(\sum_{k=1}^{\infty} x_k^2\biggr)^{1/2} \biggl(\sum_{k=1}^{\infty} y_k^2\biggr)^{1/2}.$$

You can prove this directly, although now that we know that Cauchy–Schwarz holds, we don't have to.

Challenge: Can you explain why this inequality implies that if x and y are in ℓ^2, then the sum defining $\langle x, y\rangle$ converges? ◊

Example 3.128. Let f and g be two functions in $C[0,1]$. We saw earlier that

$$\langle f, g \rangle \;=\; \int_0^1 f(x)\, g(x)\, dx$$

defines an inner product on $C[0,1]$, and the induced norm is the L^2-norm, which is

$$\|f\|_2 \;=\; \biggl(\int_0^1 f(x)^2\, dx\biggr)^{1/2}.$$

The Cauchy–Schwarz Inequality tells us that

$$|\langle f, g\rangle| \le \|f\|_2\, \|g\|_2.$$

Writing the preceding line out in terms of function values, it turns into the following inequality:

$$\biggl|\int_0^1 f(x)\, g(x)\, dx\biggr| \;\le\; \biggl(\int_0^1 f(x)^2\, dx\biggr)^{1/2} \biggl(\int_0^1 g(x)^2\, dx\biggr)^{1/2}. \qquad \diamondsuit$$

Now that we have the Cauchy–Schwarz Inequality, we will use it to prove that the induced norm really is a norm.

Theorem 3.129. *Let $\langle \cdot, \cdot \rangle$ be an inner product on a vector space V. Then the induced norm,*

$$\|x\| \;=\; \langle x, x\rangle^{1/2}, \qquad x \in V,$$

is a norm on V.

Proof. We have to verify that each of the four properties of a norm that are given in Definition 3.102 are satisfied.

Property (a). By definition of an inner product, we have $0 \le \langle x, x\rangle < \infty$ for each $x \in V$. Taking square roots, it follows that $0 \le \|x\| < \infty$ for each vector $x \in V$. This establishes that statement (a) of Definition 3.102 holds.

Property (b). Suppose that $\|x\| = 0$. Squaring both sides, it follows that

$$\langle x, x\rangle \;=\; \|x\|^2 \;=\; 0.$$

Therefore, by definition of an inner product, we have $x = 0$. This establishes that statement (b) of Definition 3.102 holds.

Property (c). Choose any vector $x \in V$ and any scalar $c \in \mathbf{R}$. Then we have

$$\|cx\| \;=\; \langle cx, cx\rangle^{1/2} \;=\; \bigl(c^2\, \langle x, x\rangle\bigr)^{1/2} \;=\; \bigl(c^2\, \|x\|^2\bigr)^{1/2} \;=\; |c|\, \|x\|.$$

Hence statement (c) of Definition 3.102 holds.

Property (d). Choose any two vectors x, $y \in V$. Then we have

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle && \text{Distributive Law} \\
&= \|x\|^2 + 2\langle x,y\rangle + +\|y\|^2 && \text{Definition} \\
&\le \|x\|^2 + 2\|x\|\,\|y\| + +\|y\|^2 && \text{Cauchy–Schwarz} \\
&= \bigl(\|x\| + \|y\|\bigr)^2.
\end{aligned}$$

Taking square-roots, it follows that $\|x+y\| \le \|x\|+\|y\|$. Therefore the Triangle Inequality holds, and this is statement (d) of Definition 3.102. □

3.12 Orthogonality

Two vectors in $\mathbf{R}^n$ are perpendicular if their dot product is zero. We extend this notion to any vector space that has an inner product.

Definition 3.130. Let $\langle\cdot,\cdot\rangle$ be an inner product on a vector space V.

(a) We say that two vectors x and y are *orthogonal* or *perpendicular* if $\langle x,y\rangle = 0$. In this case we write $x \perp y$. That is,

$$x \perp y \quad\Longleftrightarrow\quad \langle x,y\rangle = 0.$$

(b) We say that two vectors x and y are *orthonormal* if $\langle x,y\rangle = 0$ and $\|x\| = \|y\| = 1$. That is, orthonormal vectors are unit vectors that are perpendicular. ♢

Using this definition, the zero vector is orthogonal to every other vector, because $\langle x,0\rangle = 0$ for every $x \in V$. We might not like this, but this is what Definition 3.130 says, so we have to accept it. If we want to talk about perpendicular vectors but exclude the zero vector, we have to say something like "Let x, y be nonzero orthogonal vectors." On the other hand, *orthonormal* vectors must be nonzero, because they are unit vectors.

Here are some simple, but useful, facts.

Exercise 3.131. Let $\langle\cdot,\cdot\rangle$ be an inner product on a vector space V.

(a) Let x be a vector in V. Show that

$$x \perp x \quad\Longleftrightarrow\quad x = 0.$$

That is, the only vector that is orthogonal to itself is the zero vector.

(b) Show that the only vector that is orthogonal to every vector in V is the zero vector. That is, prove that

$$x \perp y \text{ for every } y \in V \quad\Longleftrightarrow\quad x = 0. \qquad ♢$$

We often have more than two vectors, so we extend the definition of orthogonality and orthonormality to larger sets. For simplicity, we will state the definition for countable (listable) sets, but similar definitions hold for arbitrary infinite sets.

Definition 3.132. Let $\langle\cdot,\cdot\rangle$ be an inner product on a vector space V.

(a) We say that $\{x_1, x_2, \dots\}$ is an *orthogonal set* if $\langle x_m, x_n\rangle = 0$ whenever $m \ne n$.

(b) We say that $\{x_1, x_2, \dots\}$ is an *orthonormal set* if it is an orthogonal set, and furthermore we have $\|x_n\| = 1$ for every n. ♢

Here's a convenient way to reword the definition of orthonormality.

Exercise 3.133. (a) Show that $\{x_1, x_2, \dots\}$ is an orthonormal set if and only if

$$\forall\, m,n \in \mathbf{N}, \qquad \langle x_m, x_n\rangle \;=\; \begin{cases} 1, & m = n,\\ 0, & m \ne n. \end{cases}$$

(b) Suppose that $\{x_1, x_2, \dots\}$ is an orthogonal set. Show that if $x_n \ne 0$ for every n, then

$$\left\{\frac{x_1}{\|x_1\|}, \frac{x_2}{\|x_2\|}, \dots\right\}$$

is an orthonormal set. We say that we obtain this orthonormal sequence by *normalizing* the sequence $\{x_1, x_2, \dots\}$ (i.e., normalizing is simply dividing a vector by its length to obtain a unit vector). ♢

Here are some examples.

Exercise 3.134. (a) Consider the vector space $C[0, 2\pi]$, where the inner product is

$$\langle f, g\rangle \;=\; \int_0^{2\pi} f(x)\, g(x)\, dx.$$

Show that $\{1, \sin x, \cos x\}$ is an orthogonal family. Is it orthonormal? If not, use Exercise 3.133 to construct a related orthonormal set.

(b) Now consider the space ℓ^2, whose elements are infinite sequences. Exhibit an infinite orthonormal set $\{x_1, x_2, \dots\}$ in ℓ^2.

Hint: What are the "simplest" vectors in ℓ^2?

(c) Is the set of monomials $\{1, x, x^2, \dots\}$ an orthogonal set in $C[0, 1]$? If we define the angle between x^m and x^n to be the number θ that satisfies

$$\cos\theta \;=\; \frac{\langle x^m, x^n\rangle}{\|x^m\|_2\, \|x^n\|_2},$$

then what is the angle between x^m and x^n? Are these two functions "close" to being orthogonal when m and n are large, or are they "far" from being orthogonal?

(d) Repeat part (c), but use the domain $[-1,1]$ instead of $[0,1]$. In particular, show that 1 and x^2 are orthogonal to x, but x^2 is not orthogonal to 1. How would you picture this in a 3-dimensional diagram? Can you find a first-degree polynomial p such that $\{1, p, x^2\}$ is an orthogonal set?

(e) Challenge: Use trig identities to extend part (a). Show that

$$\{1, \sin x, \cos x, \sin 2x, \cos 2x, \sin 3x, \cos 3x, \dots\}$$

is an orthogonal set in $C[0, 2\pi]$.

Hint: This is much easier to do if you consider the complex-valued functions $e^{inx} = \cos nx + i \sin x$. Compute the inner products $\langle e^{imx}, e^{inx} \rangle$ by using a u-substitution ($u = i(m-n)x$). This shows that $\{e^{inx}\}_{n \in \mathbf{Z}}$ is an orthogonal set. Then consider the real and imaginary parts to show that the set of sines and cosines is orthogonal. ♢

The Pythagorean Theorem tells us something about orthogonal vectors in $\mathbf{R}^n$. Because inner products have the same properties that the dot product has, we have a Pythagorean Theorem in any inner product space.

Exercise 3.135 (Pythagorean Theorem). Let $\langle \cdot, \cdot \rangle$ be an inner product on a vector space V.

(a) Given vectors $x, y \in V$, prove that

$$x \perp y \qquad \Longleftrightarrow \qquad \|x+y\|^2 \;=\; \|x\|^2 + \|y\|^2.$$

Illustrate via a diagram why this is the "Pythagorean Theorem."

(b) Use induction to extend part (a) as follows: Show that if $x_1, \dots, x_n$ are orthogonal vectors, then

$$\left\| \sum_{k=1}^{n} x_k \right\|^2 \;=\; \sum_{k=1}^{n} \|x_k\|^2.$$

How does this simplify if $x_1, \dots, x_n$ are orthonormal? ♢

3.12.1 Orthogonality (almost) implies Independence

Any set that includes the zero vector is dependent. Since the zero vector is orthogonal to every vector, an orthogonal set can be dependent. However, we will prove that a set of nonzero orthogonal vectors must be independent.

Theorem 3.136. *If $\{x_1, \dots, x_n\}$ is a set of nonzero orthogonal vectors in an inner product space V, then $\{x_1, \dots, x_n\}$ is independent.*

Proof. Suppose that $x_1, \dots, x_n$ are nonzero, orthogonal vectors. Suppose that $c_1x_1 + \cdots + c_nx_n = 0$ for some scalars $c_1, \dots, c_n$. We must show that each c_k is zero.

We begin with c_1. Since the inner product of the zero vector with any other vector is zero, we have $\langle 0, x_1\rangle = 0$. Substituting $c_1x_1 + \cdots + c_nx_n = 0$, we get

$$\begin{aligned} 0 &= \langle 0, x_1\rangle \\ &= \bigl\langle c_1x_1 + c_2x_2 + \cdots + c_nx_n,\, x_1\bigr\rangle \\ &= c_1\langle x_1, x_1\rangle + c_2\langle x_2, x_1\rangle + \cdots + c_n\langle x_n, x_1\rangle \\ &= c_1 \cdot 1 + c_2 \cdot 0 + \cdots + c_n \cdot 0 \\ &= c_1. \end{aligned}$$

Hence $c_1 = 0$. We repeat the process, using the vectors $x_2, \dots, x_n$ in turn to get $c_2 = \cdots = c_n = 0$. Hence $\{x_1, \dots, x_n\}$ is independent. □

The proof we gave is not the only one. Try the following: Suppose that $c_1x_1 + \cdots + c_nx_n = 0$, and then compute

$$\bigl\langle c_1x_1 + \cdots + c_nx_n,\, c_1x_1 + \cdots + c_nx_n\bigr\rangle.$$

You should be able to show from this that every c_k is zero. Yet another approach is to apply the extended form of the Pythagorean Theorem derived in part (b) of Exercise 3.135 to show that $c_k = 0$ for every k. Try doing both of these!

Here's an extension to infinitely many vectors.

Exercise 3.137. Suppose that $\{x_1, x_2, \dots\}$ is an infinite set of nonzero orthogonal vectors in an inner product space V. Show that $\{x_1, x_2, \dots\}$ is independent. ♢

3.12.2 Orthonormal Bases

The best situation of all is when we have an orthonormal set of vectors that is also a basis for V.

Definition 3.138. If $\mathcal{B} = \{x_1, \dots, x_n\}$ is an orthonormal set of vectors that is a basis for an inner product space V, then we call $\mathcal{B}$ an *orthonormal basis* (ONB) for V. ♢

Exercise 3.139. (a) Show that the standard basis is an orthonormal basis for $\mathbf{R}^n$.

(b) Find an orthonormal basis for $\mathbf{R}^3$ that is different from the standard basis. ♢

If you have a basis, then you know that every vector x can be written as some unique linear combination of the basis vectors. The nice thing about an *orthonormal basis* is that you know exactly what the scalars in the linear combination have to be. This is part of the next exercise.

Exercise 3.140. Suppose that $\{x_1, \dots, x_n\}$ is an orthonormal basis for an inner product space V.

(a) Prove the *Plancherel Equality*:

$$\|x\|^2 \;=\; \sum_{k=1}^{n} |\langle x, x_k\rangle|^2, \qquad x \in V.$$

(b) Prove the *Parseval Equality*:

$$\langle x, y\rangle \;=\; \sum_{k=1}^{n} \langle x, x_n\rangle\, \langle x_n, y\rangle.$$

(c) Prove that the *unique* way to write a vector $x \in V$ as a linear combination of $x_1, \dots, x_n$ is

$$x \;=\; \sum_{k=1}^{n} \langle x, x_k\rangle\, x_k. \qquad \diamondsuit$$

The following exercise pushes this a bit further.

Exercise 3.141. Suppose that $\{x_1, \dots, x_n\}$ is an *orthonormal set* of vectors in an inner product space V (but you don't know whether they are a basis for V). Prove that the following statements are equivalent (each implies the others).

(a) $\{x_1, \dots, x_n\}$ is an ONB for V.

(b) $n = \dim(V)$.

(c) For each vector $x \in V$ there are scalars $c_1, \dots, c_n$ such that

$$x \;=\; \sum_{k=1}^{n} c_k x_k.$$

(d) For each vector $x \in V$ we have

$$x \;=\; \sum_{k=1}^{n} \langle x, x_k\rangle\, x_k.$$

(e) For each vector $x \in V$ we have

$$\|x\|^2 = \sum_{k=1}^{n} |\langle x, x_k\rangle|^2.$$

(f) For each pair of vectors $x, y \in V$ we have

$$\langle x, y\rangle = \sum_{k=1}^{n} \langle x, x_n\rangle \langle x_n, y\rangle. \qquad \diamondsuit$$

3.13 Exercises

Section 3.13 in Apostol contains some exercises for you to work. Here are some additional exercises.

3.13. Parts (a) and (b) of this problem were appeared in an earlier problem, but we repeat them here, and give some additional parts. Let V be a finite-dimensional vector space, and let $n = \dim(V)$.

(a) Show that V is *isomorphic* to $\mathbf{R}^n$, i.e., show that there exists a bijection $f\colon \mathbf{R}^n \to V$ such that

$$\forall\, x, y \in \mathbf{R}^n, \quad \forall\, a, b \in \mathbf{R}, \quad f(ax + by) = af(x) + bf(y).$$

Hint: Let $\mathcal{B} = \{x_1, \dots, x_n\}$ be a basis for V. Then every vector $v \in V$ can be written $v = c_1x_1 + \cdots + c_nx_n$ for a unique choice of scalars $c_1, \dots, c_n \in \mathbf{R}$.

(b) Show that the inverse function $f^{-1}\colon V \to \mathbf{R}^n$ is also an isomorphism.
Hint: We already know that f^{-1} is a bijection, so what you have to prove is that $f^{-1}(au + bv) = af^{-1}(u) + bf^{-1}(v)$ for all $u, v \in V$ and $a, b \in \mathbf{R}$.

(c) Define a function $\langle\cdot,\cdot\rangle$ on $V \times V$ by the rule

$$\langle u, v\rangle = f^{-1}(u) \cdot f^{-1}(v), \qquad \text{for } u, v \in V,$$

where the right-hand side of the equation above is the dot product on $\mathbf{R}^n$. Show that $\langle\cdot,\cdot\rangle$ is an inner product on V.

(d) Now take the particular case where $V = \mathcal{P}_{n-1}$, the space of polynomials of degree at most $n-1$. We know that $(f, g) = \int_{-1}^{1} f(x)\, g(x)\, dx$ defines one inner product on $\mathcal{P}_{n-1}$. Is this the same inner product as the one defined in part (c)?

3.14. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be an orthonormal basis for $\mathbf{R}^n$ (not necessarily the standard basis!). Given $x \in \mathbf{R}^n$, find $[x]_{\mathcal{B}}$ (the coordinate vector for x with respect to the basis $\mathcal{B}$). Your answer should be in terms of x and $v_1, \dots, v_n$.
Note: The inner product on $\mathbf{R}^n$ is the usual dot product.

3.15. Let $\{x_1, x_2, \dots\}$ be an orthonormal set of vectors in an inner product space V. Find the distance between x_m and x_n, i.e., find $\|x_m - x_n\|$ when $m \neq n$.

3.16. Let $x_1, x_2, \dots$ be vectors in a normed space X. We say that the infinite series $\sum_{k=1}^{\infty} x_k$ *converges* if there is a vector $x \in V$ such that

$$\lim_{N\to\infty} \left\| x - \sum_{k=1}^{N} x_k \right\| = 0.$$

(a) Suppose that the two series $\sum_{k=1}^{\infty} x_k$ and $\sum_{k=1}^{\infty} y_k$ both converge. Show that $\sum_{k=1}^{\infty}(x_k + y_k)$ converges.

(b) Suppose that $\sum_{k=1}^{\infty} x_k$ converges. Show that the sum of this series is unique, i.e., if x and y both satisfy

$$\lim_{N\to\infty} \left\| x - \sum_{k=1}^{N} x_k \right\| = 0$$

and

$$\lim_{N\to\infty} \left\| y - \sum_{k=1}^{N} x_k \right\| = 0,$$

then we must have $x = y$.