2 Vector Spaces

In this course we want to study "high-dimensional spaces" and "vectors". That's not very specific, though, until we explain exactly what we mean by those things.

An important idea of this course is that it is helpful to study the same things from more than one perspective; sometimes a question that is difficult from one perspective is easy from another, so the ability to have multiple viewpoints and translate between them is extremely useful.

In this course we will take three different perspectives, which I am calling "geometric", "algebraic", and "formal". The first involves spatial reasoning and pictures; the second involves arithmetic and algebraic computations; the third involves formal definitions and properties.

A common definition of a vector is "something that has size and direction." This is a *geometric* viewpoint, since it calls to mind a picture. We can also view it from an *algebraic* point of view by giving it a set of coordinates. For instance, we can specify a two-dimensional vector by giving a pair of real numbers (x, y), which tells us where the vector points from the origin at $(0, 0)$.

The formal perspective is the most abstract and sometimes the most confusing, but often the most fruitful. This is the approach we took in section 1.1 when we defined a field: there, we took the properties the real numbers satisfy, and looked for other types of numbers that work the same way. Here we're going to start with the "ordinary" types of vectors we see in physics or in multivariable calculus, and abstract out their properties.

In the table below I have several concepts, and ways of thinking about them in each perspective. It's fine if you don't know what some of these things mean, especially in the "formal" column; if you knew all of this already you wouldn't need to take this course.

Geometric	Algebraic	Formal
size and direction	n-tuples	vectors
consecutive motion	pointwise addition	vector addition
stretching, rotations, reflections	matrices	linear functions
number of independent directions	number of coordinates	dimension
plane	system of linear equations	subspace
angle	dot product	inner product
Length	magnitude	norm

2.1 Motivation: Geometric Vectors

You should be familiar with the *Cartesian plane* from high school geometry. (It is named after the French mathematician René Descartes, who is credited with inventing the idea of putting numbered cordinates on the plane.)

As probably looks familiar from high school geometry, given two points A and B in the plane, we can write $\overrightarrow{AB}$ for the vector with *initial point* A and *terminal point* B.

Since a vector is just a length and a direction, the vector is "the same" if both the initial and terminal points are shifted by the same amount. If we fix an *origin* point O, then any point A gives us a vector $\overrightarrow{OA}$. Any vector can be shifted until its initial point is O, so each vector corresponds to exactly one point. We call this *standard position.*

We represent points algebraically with pairs of real numbers, since points in the plane are determined by two coordinates. We use $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ to denote the set of all ordered pairs of real numbers; thus $\mathbb{R}^2$ is an algebraic description of the Cartesian plane. (We use $\mathbb{R}$ to denote the set of real numbers, and the superscript ^{2} tells us that we need two of them). We define the origin O to be the "zero" point $(0, 0)$.

Definition 2.1. If $A = (x, y)$ is a point in $\mathbb{R}^2$, then we denote the vector $\overrightarrow{OA}$ by $\begin{bmatrix} x \\ y \end{bmatrix}$.

We can do something very similar with threespace.

Definition 2.2. We define *Euclidean threespace* to be the three-dimensional space described by three real coordinates. We notate it $\mathbb{R}^3$. The point $(0, 0, 0)$ is called the *origin* and often notated O.

If $A = (x, y, z)$ is a point, then the vector $\overrightarrow{OA}$ is denoted $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

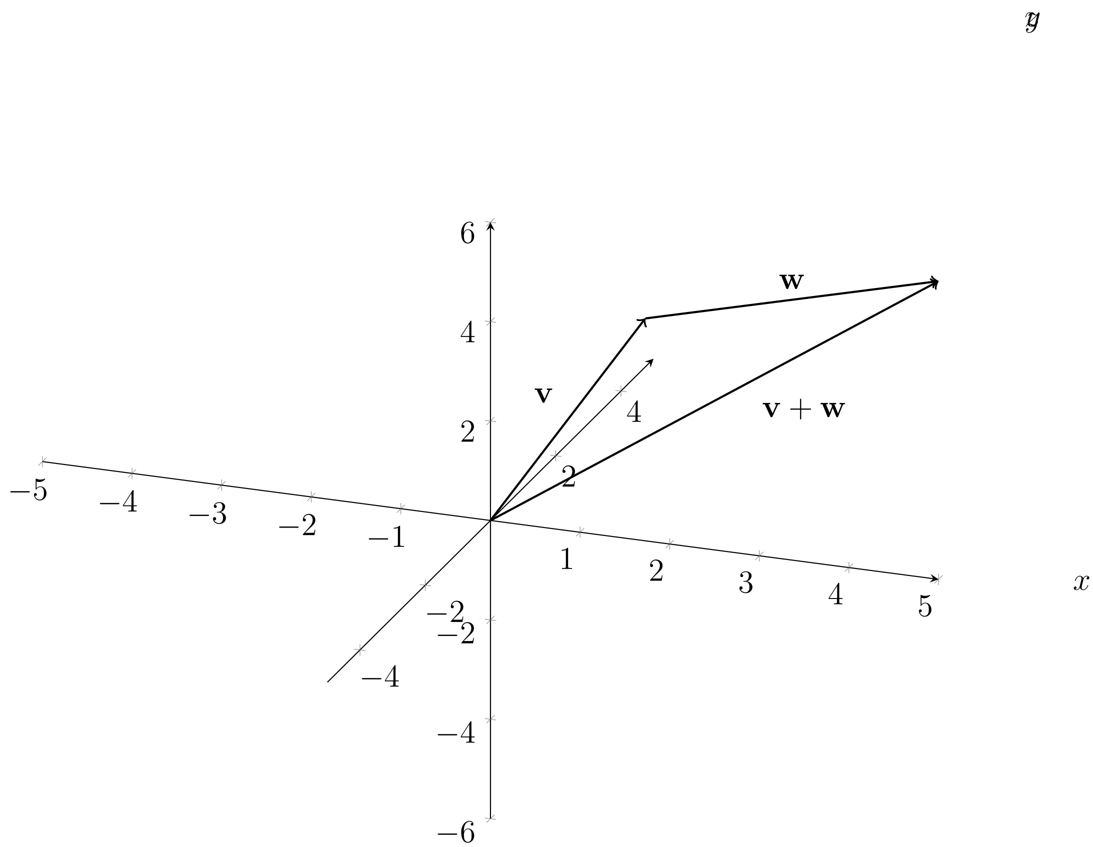
There are two operations we can do on these vectors:

1. We can *add* two vectors together. Geometrically, this corresponds to following one vector and then the other; you can picture this as laying them tip-to-tail. Algebraically, we just add the coordinates.

2. We can *multiply* a vector by a *scalar*. Geometrically corresponds to stretching a vector by some factor. Algebraically we just multiply each coordinate by the scalar.

Example 2.3. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $w = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}$. Then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 5 \\ 0 \\ 6 \end{bmatrix}, \quad 3 \cdot \mathbf{v} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}, \text{ and } \quad (-2) \cdot \mathbf{w} = \begin{bmatrix} -8 \\ 4 \\ -6 \end{bmatrix}.$$



2.2 An Algebraic Generalization

There are two straightforward ways we can generalize our Cartesian space $\mathbb{R}^3$. The most obvious is just to replace the 3 with a 4, or a 5, or a 6. If $\mathbb{R}^2$ is ordered pairs of real numbers, and $\mathbb{R}^3$ is ordered triples, then $\mathbb{R}^n$ is ordered n-tuples.

Definition 2.4. We define *real n-dimensional space* to be the set of n-tuples of real numbers, $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}\}$.

By "abuse of notation" we will also use $\mathbb{R}^n$ to refer to the set of vectors in $\mathbb{R}^n$. We define scalar multiplication and vector addition by

$$r \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} rx_1 \\ rx_2 \\ \vdots \\ rx_n \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Example 2.5. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 5 \\ -1 \\ 2 \\ 8 \end{bmatrix}$ be vectors in $\mathbb{R}^4$. Then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ -1 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 4 \\ 12 \end{bmatrix}, \quad -3 \cdot \mathbf{v} = \begin{bmatrix} -3 \\ -9 \\ -6 \\ -12 \end{bmatrix}.$$

The other way we can generalize this is to not work over the real numbers. The real numbers are a good model for every-day geometry, so we started there. But *algebraically* we could do all of these same operations with any other field.

Definition 2.6. Let $\mathbb{F}$ be any field. Then $\mathbb{F}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{F}\}$ is the set of ordered n-ples over $\mathbb{F}$. We then define scalar multiplication and vector addition by

$$r \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} rx_1 \\ rx_2 \\ \vdots \\ rx_n \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

Notice that definition 2.6is *exactly* the same as definition 2.4, except we don't specify what the field is.

Example 2.7. Let $\mathbf{v} = (3 + i, 1, 2i)$ and $\mathbf{w} = (2, 5i, 4 - 2i)$ be vectors in $\mathbb{C}^3$. Then

$$\mathbf{v} + \mathbf{w} = (5 + i, 1 + 5i, 4)$$
$$(2 - i)\mathbf{v} = (7 - i, 2 - i, 2 + 4i).$$

Notice that the scalar is a complex number, because we're working over $\mathbb{C}$.

Example 2.8. Let $\mathbf{v} = (1,2,3)$ and $\mathbf{w} = (3,4,1)$ be vectors in $\mathbb{F}_5$. Then

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= (4,1,4) \\ 2\mathbf{v} &= (2,4,1). \end{aligned}$$

Our scalar is indeed an element of $\mathbb{F}_5$, and all the arithmetic is being done mod 5.

2.3 Defining Vector Spaces

We want to figure out what properties we're actually using to work with these sets of vectors. Obviously, we have a set of vectors, and a set of scalars; and we have two operations, addition and scalar multiplication. These operations also behave "nicely", following all of the rules in this long and tedious definition:

Definition 2.9. Let $\mathbb{F}$ be a field, and V be a set, together with two operations:

- A *vector addition* which allows you to add two elements of V and get a new element of V. If $\mathbf{v}, \mathbf{w} \in V$ then the sum is denoted $\mathbf{v} + \mathbf{w}$ and must also be an element of V.
- A *scalar multiplication* which allows you to multiply an element of V by a "scalar" element of $\mathbb{F}$ and get a new element of V. If $a \in \mathbb{F}$ and $\mathbf{v} \in V$ then the scalar multiple is denoted $a \cdot \mathbf{v}$ and must also be an element of V.

Further, suppose the following axioms hold for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and any $a, b \in \mathbb{F}$:

1. (Closure under addition) $\mathbf{u} + \mathbf{v} \in V$
2. (Closure under scalar multiplication) $a\mathbf{u} \in V$
3. (Additive commutativity) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
4. (Additive associativity) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
5. (Additive identity) There is an element $\mathbf{0} \in V$ called the "zero vector", such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for every $\mathbf{u}$.
6. (Additive inverses) For each $\mathbf{u} \in V$ there is another element $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
7. (Distributivity) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
8. (Distributivity) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$

9. (Multiplicative associativity) $a(b\mathbf{u}) = (ab)\mathbf{u}$

10. (Multiplicative Identity) $1\mathbf{u} = \mathbf{u}$.

Then we say V is a *Vector Space* over $\mathbb{F}$, and we call its elements *vectors*.

Remark 2.10. Technically, those first two axioms are superfluous; if you can add two elements, you can add two elements and also get something. But they still need to be true: if adding two vectors doesn't give you another vector, you don't have a vector space. And we have to check them to make sure our vector space definition makes sense.

Example 2.11. $\mathbb{F}^n$ is a vector space, with the previously defined vector addition and scalar multiplication. We check:

Let $\mathbf{u} = (u_1, \dots, u_n).\mathbf{v} = (v_1, \dots, v_n), \mathbf{w} = (w_1, \dots, w_n) \in \mathbb{F}^n, \quad r, s \in \mathbb{F}$. Then, knowing the usual rules of commutativity and associativity of basic arithmetic, we can compute:

1. $\mathbf{u} + \mathbf{v} = (u_1, \dots, u_n) + (v_1, \dots, v_n) = (u_1 + v_1, \dots, u_n + v_n) \in \mathbb{F}^n$.

2. $$r\mathbf{u} = r(u_1, \dots, u_n) = (ru_1, \dots, ru_n) \in \mathbb{F}.$$

3. $$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1, \dots, u_n) + (v_1, \dots, v_n) = (u_1 + v_1, \dots, u_n + v_n) \\ &= (v_1 + u_1, \dots, v_n + u_n) = (v_1, \dots, v_n) + (u_1, \dots, u_n) = \mathbf{v} + \mathbf{u} \end{aligned}$$

4. $$\begin{aligned} (\mathbf{u} + \mathbf{v}) + \mathbf{w} &= (u_1 + v_1, \dots, u_n + v_n) + (w_1, \dots, w_n) = (v_1 + u_1 + w_1, \dots, v_n + u_n + w_n) \\ &= (v_1, \dots, v_n) + (u_1 + w_1, \dots, u_n + w_n) = \mathbf{v} + (\mathbf{u} + \mathbf{w}) \end{aligned}$$

5. We have $\mathbf{0} = (0, \dots, 0)$. Then
$$\mathbf{0} + \mathbf{v} = (0 + v_1, \dots, 0 + v_n) = (v_1, \dots, v_n) = \mathbf{v}.$$

6. Set $-\mathbf{u} = (-u_1, \dots, -u_n)$. Then
$$\mathbf{u} + (-\mathbf{u}) = (u_1 + (-u_1), \dots, u_n + (-u_n)) = (0, \dots, 0) = \mathbf{0}.$$

7.

$$\begin{aligned} r(\mathbf{u}+\mathbf{v}) &= r(u_1+v_1,\dots,u_n+v_n) = (r(u_1+v_1),\dots,r(u_n+v_n)) \\ &= (ru_1+rv_1,\dots,ru_n+rv_n) = (ru_1,\dots,ru_n)+(rv_1,\dots,rv_n) = r\mathbf{u}+r\mathbf{v}. \end{aligned}$$

8.

$$\begin{aligned} (r+s)\mathbf{u} &= (r+s)(u_1,\dots,u_n) = ((r+s)u_1,\dots,(r+s)u_n) \\ &= (ru_1+su_1,\dots,ru_n+su_n) = (ru_1,\dots,ru_n)+(su_1,\dots,su_n) = r\mathbf{u}+s\mathbf{u}. \end{aligned}$$

9.

$$r(s\mathbf{u}) = r(su_1,\dots,su_n) = (rsu_1,\dots,rsu_n) = rs(u_1,\dots,u_n).$$

10.

$$1\mathbf{u} = 1(u_1,\dots,u_n) = (1\cdot u_1,\dots,1\cdot u_n) = (u_1,\dots,u_n) = \mathbf{u}.$$

So what else is a vector space and "looks like $\mathbb{R}^n$"? The most important example in this course will be *matrices*.

Definition 2.12. A *matrix over a field* $\mathbb{F}$ is a rectangular array of elements of $\mathbb{F}$. A matrix with m rows and n columns is *a* $m \times n$ *matrix*, and we notate the set of all such matrices by $M_{m\times n}(\mathbb{F})$, or just $M_{m\times n}$ if the field is clear from context. .

A $m \times n$ matrix is *square* if $m = n$, that is, it has the same number of rows as columns. We will sometimes represent the set of $n \times n$ square matrices by M_n.

We will generally describe the elements of a matrix with the notation

$$(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

We can define operations on these matrices:

- If $A = (a_{ij})$ is an $m \times n$ matrix over a field $\mathbb{F}$, and $r \in \mathbb{F}$, then we can multiply each entry of the matrix A by the r. This is called *scalar multiplication* and we say that r is a *scalar*.

$$rA = (ra_{ij}) = \begin{bmatrix} ra_{11} & ra_{12} & \dots & ra_{1n} \\ ra_{21} & ra_{22} & \dots & ra_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ra_{m1} & ra_{m2} & \dots & ra_{mn} \end{bmatrix}.$$

- If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices over a field $\mathbb{F}$, we can add the two matrices by adding each individual pair of coordinates together.

$$A + B = (a_{ij} + b_{ij}) = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}.$$

Example 2.13. The set $M_{m \times n}(\mathbb{F})$ of $m \times n$ matrices is a vector space under the addition and scalar multiplication defined above, with zero vector given by

$$\mathbf{0} = (0) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

I'm not going to prove this, but you can see that it should be true for the same reason FR^{mn} is a vector space: they're both just lists of numbers, but one is arranged in a column and the other in a rectangle. The operations are the same.

Example 2.14. Pick a field $\mathbb{F}$, and let $\mathcal{P}_{\mathbb{F}}(x) = \{a_0 + a_1 x + \dots + a_n x^n : n \in \mathbb{N}, a_i \in \mathbb{F}\}$ be the set of polynomials with coefficients in $\mathbb{F}$. Define addition by

$$(a_0 + a_1 x + \dots + a_n x^n) + (b_0 + b_1 x + \dots + b_n x^n) = (a_0 + b_0) + (a_1 + b_1)x + \dots + (a_n + b_n)x^n$$

and define scalar multiplication by

$$r(a_0 + a_1 x + \dots + a_n x^n) = ra_0 + ra_1 x + \dots + ra_n x^n.$$

Then $\mathcal{P}_{\mathbb{F}}(x)$ is a vector space.

Example 2.15. Fix a field $\mathbb{F}$, and let S be the space of all doubly infinite sequences $\{y_k\} = \{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots : y_i \in \mathbb{F}\}$. We call this the space of (discrete) *signals*: it represents a sequence of measurements taken at regular time intervals. These sorts of regular measurements are common in engineering and digital information applications (such as digital music).

We define addition and scalar multiplication on the space of signals componentwise, so that

$$\{\dots, x_{-1}, x_0, x_1, \dots\} + \{\dots, y_{-1}, y_0, y_1, \dots\} = \{\dots x_{-1} + y_{-1}, x_0 + y_0, x_1 + y_1, \dots\}$$

and

$$r\{\ldots, y_{-1}, y_0, y_1, \ldots\} = \{\ldots, ry_{-1}, ry_0, ry_1, \ldots\}.$$

(In essence, S is composed of vectors that are infinitely long in both directions). Then S is a vector space.

Example 2.16. Let $\mathcal{F}(\mathbb{R}, \mathbb{R}) = \mathcal{F}$ be the set of functions from $\mathbb{R}$ to $\mathbb{R}$—that is, functions that take in a real number and return a real number, the vanilla functions of single-variable calculus. Define addition by $(f + g)(x) = f(x) + g(x)$ and define scalar multiplication by $(rf)(x) = r \cdot f(x)$. Then $\mathcal{F}$ is a vector space. You will prove this is a vector space in your homework.

Example 2.17. The integers $\mathbb{Z}$ are *not* a vector space (under the usual definitions of addition and multiplication). For instance, $1 \in \mathbb{Z}$ but $.5 \cdot 1 = .5 \notin \mathbb{Z}$.

(We only need to find one axiom that doesn't hold to show that a set is not a vector space, since a vector space must satisfy all the axioms).

Example 2.18. The closed interval $[0, 5]$ is not a vector space (under the usual operations) , since $3, 4 \in [0, 5]$ but $3 + 4 = 7 \notin [0, 5]$.

Example 2.19. Let $V = \mathbb{R}$ with scalar multiplication given by $r \cdot x = rx$ and addition given by $x \oplus y = 2x + y$. Then V is not a vector space, since $x \oplus y = 2x + y \neq 2y + x = y \oplus x$; in particular, we see that $3 \oplus 5 = 11$ but $5 \oplus 3 = 13$.

There are many more examples of vector spaces, but as you can see it's fairly tedious to prove that any particular thing is a vector space. In section 2.4 we'll develop a *much* easier way to establish that something is a vector space, so we won't develop any more examples now.

2.3.1 Properties of Vector Spaces

The great thing about the formal approach is that we can show that anything that satisfies the axioms of a vector space must aso follow some other rules. We'll establish a few of those rules here, though of course, there's a sense in which the entire rest of this course will be spent establishing those rules.

As before, you shouldn't think of these rules as new facts; all of them are "obvious". The point is that if we get the list of properties from definition 2.9, then all of these other things still have to occur. It's a guarantee that vector spaces behave how we expect—that they all do behave like $\mathbb{F}^n$, or indeed like $\mathbb{R}^3$, in all the ways we will expect.

Proposition 2.20 (Cancellation)**.** *Let V be a vector space and suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ are vectors. If $\mathbf{u} + \mathbf{w} = \mathbf{v} + \mathbf{w}$, then $\mathbf{u} = \mathbf{v}$.*

Remark 2.21. We stated this law for *fields* earlier; now we're also claiming it holds for *vector spaces.* But the proof is essentially the same in both cases. (This is the shadow of something called "universal algebra"; there are many other algebraic structures we could define, which will all have this same cancellation law for the same reason.)

Proof. By axiom we know that $\mathbf{w}$ has an additive inverse $-\mathbf{w}$. Then we have

$$\begin{aligned}
\mathbf{u} + \mathbf{w} &= \mathbf{v} + \mathbf{w} \\
(\mathbf{u} + \mathbf{w}) + (-\mathbf{w}) &= (\mathbf{v} + \mathbf{w}) + (-\mathbf{w}) \\
\mathbf{u} + (\mathbf{w} + (-\mathbf{w})) &= \mathbf{v} + (\mathbf{w} + (-\mathbf{w})) && \text{Additive associativity} \\
\mathbf{u} + \mathbf{0} &= \mathbf{v} + \mathbf{0} && \text{Additive inverses} \\
\mathbf{u} &= \mathbf{v} && \text{Additive identity.}
\end{aligned}$$

□

Proposition 2.22. *The additive inverse $-\mathbf{v}$ of a vector $\mathbf{v}$ is unique. That is, if $\mathbf{v} + \mathbf{u} = \mathbf{0}$, then $\mathbf{u} = -\mathbf{v}$.*

Proof. Suppose $\mathbf{v} + \mathbf{u} = \mathbf{0}$. By the additive inverses property we know that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$, and thus $\mathbf{v} + \mathbf{u} = \mathbf{v} + (-\mathbf{v})$. By cancellation we have $\mathbf{u} = -\mathbf{v}$. □

Remark 2.23. In our axioms we asserted that every vector *has* an inverse, but didn't require that there be only one.

Proposition 2.24. *Suppose V is a vector space with $\mathbf{u} \in V$ a vector and $r \in \mathbb{R}$ a scalar. Then:*

1. $0\mathbf{u} = \mathbf{0}$
2. $r\mathbf{0} = \mathbf{0}$
3. $(-1)\mathbf{u} = -\mathbf{u}$.

Remark 2.25. We would actually be pretty sad if any of those statements were false, since it would make our notation look very strange. (Especially the last statement). The fact that these statements *are* true justifies us using the notation we use.

Proof. 1.

$$\begin{aligned}
\mathbf{u} = 1 \cdot \mathbf{u} &= (0+1)\mathbf{u} && \text{Multiplicative identity} \\
&= 0\mathbf{u} + 1\mathbf{u} && \text{Distributivity} \\
&= 0\mathbf{u} + \mathbf{u} && \text{Multiplicative identity} \\
\mathbf{0} + \mathbf{u} &= 0\mathbf{u} + \mathbf{u} && \text{Additive identity} \\
\mathbf{0} &= 0\mathbf{u} && \text{Cancellation}
\end{aligned}$$

2. We know that $\mathbf{0} = \mathbf{0}+\mathbf{0}$ by additive identity, so $r\mathbf{0} = r(\mathbf{0}+\mathbf{0}) = r\mathbf{0}+r\mathbf{0}$ by distributivity. Then we have

$$\begin{aligned}
\mathbf{0} + r\mathbf{0} &= r\mathbf{0} + r\mathbf{0} && \text{additive identity} \\
\mathbf{0} &= r\mathbf{0} && \text{cancellation.}
\end{aligned}$$

3. We have

$$\begin{aligned}
\mathbf{v} + (-1)\mathbf{v} &= 1\mathbf{1} + (-1)\mathbf{v} && \text{multiplicative inverses} \\
&= (1 + (-1))\mathbf{v} && \text{distributivity} \\
&= 0\mathbf{v} = \mathbf{0}.
\end{aligned}$$

Then by uniqueness of additive inverses, we have $(-1)\mathbf{v} = -\mathbf{v}$.

□

Example 2.26. We'll give one last example of a vector space, which is both important and silly.

We define the *zero vector space* to be the set $\{\mathbf{0}\}$ with addition given by $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and scalar multiplication given by $r \cdot \mathbf{0} = \mathbf{0}$. It's easy to check that this is in fact a vector space.

Notice that we didn't ask what "kind" of object this is; we just said it has the zero vector and nothing else. As such, this could be the zero vector of any vector space at all. In section 2.4 we will talk about vector spaces that fit inside other vector spaces, like this one.

2.4 Vector Space Subspaces

Our very first two examples of a vector space were the Cartesian plane and Euclidean three-space. But we see that while we can think of them as totally distinct vector spaces, the plane sits *inside* threespace, as a subset. In fact it sits inside it in a number of different ways; we can start by taking the xy plane, the xz plane, or the yz plane.

Definition 2.27. Let V be a vector space. A subset $W \subseteq V$ is a *subspace* of V if W is also a vector space with the same operations as V.

Example 2.28. The Cartesian plane $\mathbb{R}^2$ is a subset of threespace $\mathbb{R}^3$. Similarly the line $\mathbb{R}^1$ is a subset of the plane $\mathbb{R}^2$. (And we can stack this up as high as we want; $\mathbb{R}^7 \subseteq \mathbb{R}^8$.

In general, if $n < m$ then $\mathbb{F}^n$ is a subspace of $\mathbb{F}^m$.

Example 2.29.

Example 2.30. Let $V = \mathbb{R}^3$ and let $W = \{(x, y, x + y) \in \mathbb{R}^3\}$. Geometrically, this is a plane (given by $z = x + y$). We could in fact write $W = \{(x, y, z) : z = x + y\}$; this is a more useful way to write it for multivariable calculus, but less useful for lienar algebra. W is certainly a subset of V, so we just need to figure out if W is a subspace.

We could do this by checking all ten axioms, but that would take a very long time; we want a better tool. And it seems like we should be able to avoid a lot of that work since we *already* know many of the axioms hold in $\mathbb{R}^3$.

In fact, one major reason to care about subspaces is that it allows us to avoid a lot of work. If $W \subseteq V$, it seems like most of the vector space axioms should hold *automatically*. After all, if elements of V add commutatively, and elements of W are in V, then the elements of W must add commutatively. And in fact there's very little we have to check.

Proposition 2.31. *Let V be a vector space over a field $\mathbb{F}$ and $W \subseteq V$. Then W is a subspace of V if and only if the following three "subspace" conditions hold:*

1. $\mathbf{0} \in W$ *(zero vector);*
2. *Whenever $\mathbf{u}, \mathbf{v} \in W$ then $\mathbf{u} + \mathbf{v} \in W$ (Closed under addition); and*
3. *Whenever $r \in \mathbb{F}$ and $\mathbf{u} \in W$ then $r\mathbf{u} \in W$ (Closed under scalar multiplication).*

Proof. Suppose W is a subspace of V. Then W is a vector space, so it contains a zero vector and is closed under addition and multiplication by the definition of vector spaces.

Conversely, suppose $W \subseteq V$ and the three subspace conditions hold. We need to check the ten axioms of a vector space. But most of these properties are inherited from the fact that any element of W is also an element of V, and W has the same operations as V. The only really non-trivial one is that the additive inverse exists.

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in W$ (and thus $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$), and $r, s \in \mathbb{F}$.

1. W is closed under addition by hypothesis.

2. W is closed under scalar multiplication by hypothesis.
3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ since V is a vector space.
4. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ since V is a vector space.
5. $\mathbf{0} \in W$ by hypothesis, and $\mathbf{u} + \mathbf{0} = \mathbf{u}$ since V is a vector space.
6. $-\mathbf{u} = (-1)\mathbf{u} \in W$ by closure under scalar multiplication.
7. $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$ since V is a vector space.
8. $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$ since V is a vector space.
9. $(rs)\mathbf{u} = r(s\mathbf{u})$ since V is a vector space.
10. $1\mathbf{u} = \mathbf{u}$ since V is a vector space.

Thus W satisfies the axioms of a vector space, and is itself a vector space. □

Example 2.32 (Continued)**.** Let's continute to take $V = \mathbb{R}^3$ and $W = \{(x, y, x + y) \in \mathbb{R}^3\}$. To show that W is a subspace of V we only need to check three things.

If $(x_1, y_1, x_1 + y_1), (x_2, y_2, x_2 + y_2) \in W$ then

$$\begin{bmatrix} x_1 \\ y_1 \\ x_1 + y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ x_2 + y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ (x_1 + x_2) + (y_1 + y_2) \end{bmatrix} \in W.$$

If $r \in \mathbb{R}$, then

$$r \begin{bmatrix} x \\ y \\ x + y \end{bmatrix} = \begin{bmatrix} rx \\ ry \\ (rx) + (ry) \end{bmatrix} \in W.$$

And the zero vector is

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 + 0 \end{bmatrix} \in W.$$

Thus W is a subspace of V.

Example 2.33. If V is a vector space, then 0 and V are both subspaces of V. We don't actually need to check anything here, since both are clearly subsets of V, and both are already known to be vector spaces.

(When we want to ignore this possibility we will refer to "proper" or "nontrivial" subspaces, which are neither the trivial space nor the entire space).

Example 2.34. Let $V = \mathbb{R}^2$ and let $W = \{(x, x^2)\} = \{(x, y) : y = x^2\} \subseteq V$. Then W is *not* a subspace.

W does in fact contain the zero vector $(0, 0) = (0, 0^2)$. But we see that $(1, 1) \in W$, and $(1, 1) + (1, 1) = (2, 2) \notin W$. Thus W is not a subspace.

Example 2.35. Let $V = \mathbb{F}_3^2$ and let $W = \{(0, 0), (1, 2), (2, 1)\} \subset V$. Is W a subspace?

It's easy to see that $\mathbf{0} = (0, 0) \in W$. We just need to check it's closed under addition and scalar multiplication.

It's a little hard to check this without just testing elements. But we compute:

$$\begin{aligned} &(0,0) + (0,0) = (0,0) \in W && (0,0) + (1,2) = (1,2) \in W \\ &(0,0) + (2,1) = (2,1) \in W && (1,2) + (1,2) = (2,1) \in W \\ &(1,2) + (2,1) = (0,0) \in W && (2,1) + (2,1) = (1,2) \in W. \end{aligned}$$

And similarly

$$\begin{aligned} &0 \cdot (0,0) = (0,0) \in W && 0 \cdot (1,2) = (0,0) \in W && 0 \cdot (2,1) = (0,0) \in W \\ &1 \cdot (0,0) = (0,0) \in W && 1 \cdot (1,2) = (1,2) \in W && 1 \cdot (2,1) = (2,1) \in W \\ &2 \cdot (0,0) = (0,0) \in W && 2 \cdot (1,2) = (2,1) \in W && 2 \cdot (2,1) = (1,2) \in W \end{aligned}$$

So W is closed under addition and scalar multiplication, so it's a subspace.

Example 2.36. Let $V = \mathcal{P}(x)$ and let $W = \{a_1 x + \cdots + a_n x^n\} = x\mathcal{P}(x)$ be the set of polynomials with zero constant term. Is W a subspace of V?

1. The zero polynomial $0 + 0x + \cdots + 0x^n = 0$ certainly has zero constant term, so is in W.

2. If $a_1 x + \cdots + a_n x^n$ and $b_1 x + \cdots + b_n x^n \in W$, then

$$(a_1 x + \cdots + a_n x^n) + (b_1 x + \cdots + b_n x^n) = (a_1 + b_1)x + \cdots + (a_n + b_n)x^n \in W.$$

   Alternatively, we can say that if we add two polynomials with zero constant term, their sum will have zero constant term.

3. If $r \in \mathbb{R}$ and $a_1 x + \cdots + a_n x^n \in W$, then

$$r\left(a_1 x + \cdots + a_n x^n\right) = (ra_1)x + \cdots + (ra_n)x^n$$

   has zero constant term and is in W.

Thus W is a subspace of V.

Example 2.37. Let $V = \mathcal{P}(x)$ and let $W = \{a_0 + a_1 x\}$ be the space of linear polynomials. Then W is a subspace of V.

1. The zero polynomial $0 + 0x \in W$.
2. If $a_0 + a_1 x, b_0 + b_1 x \in W$, then $(a_0 + a_1 x) + (b_0 + b_1 x) = (a_0 + b_0) + (a_1 + b_1)x \in W$.
3. If $r \in \mathbb{R}$ and $a_0 + a_1 x \in W$, then $r(a_0 + a_1 x) = ra_0 + (ra_1)x \in W$.

Example 2.38. Let $V = \mathcal{P}(x)$ and let $W = \{1 + ax\}$ be the space of linear polynomials with constant term 1. Is W a subspace of V?

No, because $0 = 0 + 0x \notin W$.

Exercise 2.39. *Fix a natural number $n \geq 0$. Let $V = \mathcal{P}(x)$ and let $W = \mathcal{P}_n(x) = \{a_0 + a_1 x + \cdots + a_n x^n\}$ be the set of polynomials with degree at most n. Then $\mathcal{P}_n(x)$ is a subspace of $\mathcal{P}(x)$.*

Example 2.40. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ be the space of functions of one real variable, and let $W = \mathcal{D}(\mathbb{R}, \mathbb{R})$ be the space of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. Is W a subspace of V?

1. The zero function is differentiable, so the zero vector is in W.
2. From calculus we know that the derivative of the sums is the sum of the derivatives; thus the sum of differentiable functions is differentiable. That is, $(f+g)'(x) = f'(x) + g'(x)$.

   So if $f, g \in W$, then f and g are differentiable, and thus $f + g$ is differentiable and thus in W.
3. Again we know that $(rf)'(x) = rf'(x)$. If f is in W, then f is differentiable. Thus rf is differentiable and therefore in W.

Example 2.41. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and let $W = \mathcal{F}([a, b], \mathbb{R})$ be the space of functions from the closed interval $[a, b]$ to $\mathbb{R}$. We can view W as a subset of V by, say, looking at all the functions that are zero outside of $[a, b]$. Is W a subspace of V?

1. The zero function is in W.
2. If f and g are functions from $[a, b] \to \mathbb{R}$, then $(f + g)$ is as well.
3. If f is a function from $[a, b] \to \mathbb{R}$, then rf is as well.

Example 2.42. Let $V = \mathcal{F}(\mathbb{R}, R)$. Then $\mathcal{C}(\mathbb{R}, \mathbb{R})$ the space of *continuous* real-valued functions is a subspace of V. So iare $\mathcal{D}(\mathbb{R}, \mathbb{R})$ the space of differentiable functions and $\mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ the space of infinitely differentiable functions.

Example 2.43. Let $V = \mathcal{F}(\mathbb{R}, R)$ and let $W = \{f : f(x) = f(-x) \forall x \in \mathbb{R}\}$ be the set of *even* real-valued functions, the functions that are symmetric around 0. Then W is a subspace of V.

Example 2.44. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and let $W = \mathcal{F}(\mathbb{R}, [a, b])$ be the space of functions from $\mathbb{R}$ to the closed interval $[a, b]$. Is W a subspace of V?

No! The simplest condition to check is scalar multiplication. Let $f(x) = b$ be a function in V. Let $r = (b + 1)/b$. Then $(rf)(x) = fb = b + 1$ and thus $rf \notin W$.

Example 2.45. Let $V = S$ be the space of signals, and let W be the space of signals that are eventually zero. That is, $W = \{\{y_k\} : \exists n \text{ such that } y_m = 0 \,\forall\, m > n\}$. Then W is a subspace of V.

The space $\{\{y_k\} : y_0 = 0\}$ is a subspace of V. But the space $\{\{y_k\} : y_0 = 1\}$ is not.

Theorem 2.46. *Any intersection of subspaces of a vector space V is a subspace of V.*

Proof. Let $\mathcal{C}$ be any collection of subspaces of V (there might be two, or three, or infinitely many subspaces in $\mathcal{C}$). Let W be the intersection of all subspaces in $\mathcal{C}$.

Since every subspace contains $\mathbf{0}$, therefore, $\mathbf{z}ero \in W$. Now let $a \in \mathbb{F}$ and $x, y \in W$. Since x and y are in the intersection of every subspace of $\mathcal{C}$, they are contained in each subspace in $\mathcal{C}$. Because each subspace is closed under addition, therefore $x + y$ is contained in each subspace in $\mathcal{C}$ and so $x + y \in W$.

Similarly, each subspace in $\mathcal{C}$ is closed under scalar multiplication, so each subspace contains ax. Hence $ax \in W$. Since W contains $\mathbf{z}ero$ and is closed under addition and scalar multiplication, by our subspace theorem, W is a subspace of V. □

2.5 Linear Combinations and Linear Equations

We have defined many vector spaces, but we started by looking at $\mathbb{R}^n$, which is much easier to think about. One of the nicest and most helpful things about $\mathbb{R}^n$ is the existence of *coordinates*. Rather than, say, just drawing a point on a graph, or perhaps giving an angle and a distance, we can specify a point in $\mathbb{R}^3$ by giving its x-coordinate, its y-coordinate, and its z-coordinate. And similarly, we can specify a point in $\mathbb{R}^7$ by specifying seven real-number coordinates.

In contrast, it's not really clear what it means to talk about coordinates for $\mathcal{F}(\mathbb{R},\mathbb{R})$. But if we had coordinates there, it would make our life much easier. (In particular, physicists often want to talk about subspaces of $\mathcal{F}(\mathbb{R},\mathbb{R})$ and then put coordinates on them and treat them like $\mathbb{R}^n$). So we would like to find a way to put coordinates on any vector space V.

There are a few ideas that will mix in here, but the first one is that coordinate let us express a vector as a sum of simple vectors. If I have a vector $(1,3,2)$, one way I can think of this is

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

Definition 2.47. If V is a vector space $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a list of vectors in V, then a *linear combination* of of the vectors in S is a vector of the form

$$\sum_{i=1}^{n} a_i \mathbf{v}_i = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

where $a_i \in \mathbb{R}$ are (real number) scalars.

A linear combination of vectors in V will always itself be an element of V, since V is closed under scalar multiplication and under vector addition.

Geometrically, a linear combination of vectors represents some destination you can reach only going in the directions of your chosen vectors (for any distance. So if I can go north or west, any distance"northwest" will be a linear combination of those vectors. And "southeast" will as well, since we can always go in the "opposite" direction. But "up" will not be.

Remark 2.48. This is a "linear" combination because it combines the vectors in the same way a line or plane does—adding all the vectors together, but with some coefficient. We will revisit this terminology in the next section when we discuss linear functions.

It's totally possible to have a linear combination of infinitely many vectors. But studying these requires some sense of convergence, and thus calculus/analysis. So we won't talk about it in *this* class, except for the occasional aside.

Example 2.49. Here is a table of the number of grams of protein, fats, and carbohydrates in 10g portions of certain foods (rounded to give us easier numbers):[1]

Food (10g)	Protein (g)	Fats (g)	Carbs (g)
ground beef	4	4	0
lentils	2	1	3
rice (brown)	1	0	5
cauliflower	1	0	1

We could record each different food as a vector in $\mathbb{R}^3$. So we have

$$g = \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}, \quad \ell = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad r = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Then if we prepare a meal consisting of 100g of ground beef, 150g of rice, and 200g of cauliflower, the macronutrient content is the linear combination

$$10g + 15r + 20c = 10 \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix} + 15 \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} + 20 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 75 \\ 40 \\ 95 \end{bmatrix}.$$

A very reasonable question to ask here is: if we have a fixed vector $\mathbf{b}$, and a set of vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, can we express $\mathbf{b}$ as a linear combination of the other vectors?

Example 2.50. Can we write $(1, 3, 2)$ as a linear combination of $(1, 0, 0)$ and $(1, 1, 1)$?

In this case it's pretty easy to see that we can't, because any linear combination of these two vectors would have the same second and third coordinate. In other words, if we had

$$\begin{aligned} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} &= a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} a + b \\ b \\ b \end{bmatrix} \end{aligned}$$

which implies $3 = b = 2$.

Example 2.51. Is it possible to prepare a meal using the four ingredients if we want to get exactly 70g of protein, 30g of fat, and 40g of carbs? This is asking if the vector $(70, 30, 40)$ is a linear combination of the vectors g, ℓ, r, c.

In other words, we must determine if there are scalars a_1, a_2, a_3, a_4 such that

$$\begin{aligned} \begin{bmatrix} 70 \\ 30 \\ 40 \end{bmatrix} &= a_1 \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + a_3 \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} + a_4 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 4a_1 + 2a_2 + a_3 + a_4 \\ 4a_1 + a_2 \\ 3a_2 + 5a_3 + a_4 \end{bmatrix} \end{aligned}$$

And that means we need to solve the following *system of linear equations*:

$$\begin{aligned} 4a_1 + 2a_2 + a_3 + a_4 &= 70 \\ 4a_1 + a_2 &= 30 \\ 3a_2 + 5a_3 + a_4 &= 40. \end{aligned}$$

Definition 2.52. A *system of linear equations* is a system of the form

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

with the a_{ij} and b_is all real numbers. We say this is a system of m equations in n unknowns.

Importantly, these equations are restricted to be relatively simple. In each equation we multiply each variable by some constant real number, add them together, and set that equal to some constant real number. We aren't allowed to multiply variables together, or do anything else fancy with them. This means the equations can't get too complicated, and are relatively easy to work with.

Thus our question about vectors became a question about linear equations. (Or maybe originally our question about linear equations became a question about vectors; they're two ways of seeing the same thing. As the course develops we'll see a few other ways we can think of the same questions.)

There are a few ways to approach solving systems of equations like this. One is by substitution: solve for one variable in terms of the other variables, and substitute into another equation. But this gets quite cumbersome. A better way is to add copies of one equation to another.

Example 2.53 (Continued)**.** We have the system of equations

$$\begin{aligned} 4a_1 + 2a_2 + a_3 + a_4 &= 70 \\ 4a_1 + a_2 &= 30 \\ 3a_2 + 5a_3 + a_4 &= 40. \end{aligned}$$

We can eliminate the a_1 terms from all but the first equation by subtracting the first equation from the second, giving:

$$\begin{aligned} 4a_1 + 2a_2 + a_3 + a_4 &= 70 \\ - a_2 - a_3 - a_4 &= -40 \\ 3a_2 + 5a_3 + a_4 &= 40. \end{aligned}$$

We might flip the second equation to make it easier to look at:

$$\begin{aligned} 4a_1 + 2a_2 + a_3 + a_4 &= 70 \\ a_2 + a_3 + a_4 &= 40 \\ 3a_2 + 5a_3 + a_4 &= 40. \end{aligned}$$

Now we can get rid of most of the a_2 terms. We subtract 2 times the second equation from the first and 3 times the second equation from the third to obtain

$$\begin{aligned} 4a_1 \qquad - a_3 - a_4 &= -10 \\ a_2 + a_3 + a_4 &= 40 \\ 2a_3 - 2a_4 &= -80. \end{aligned}$$

Divide the third equation by 2 (or multiply by 1/2) to get

$$\begin{aligned} 3a_1 \qquad - a_3 - a_4 &= -10 \\ a_2 + a_3 + a_4 &= 40 \\ a_3 - a_4 &= -40. \end{aligned}$$

Then add it to the first equation and subtract it from the second equation to yield

$$\begin{aligned} 4a_1 \qquad\qquad - 2a_4 &= -50 \\ a_2 \qquad + 2a_4 &= 80 \\ a_3 - a_4 &= -40. \end{aligned}$$

And now we still have a system of three equations in four unknowns. But it should be clear now that if we pick *any* real number for a_4, that will give us exactly one solution to the whole system:

$$\left(\frac{-50 + 2a_4}{4}, 80 - 2a_4, -40 + a_4, a_4 \right) = \left(\frac{-25 + a_4}{2}, 80 - 2a_4, -40 + a_4, a_4 \right).$$

We want a general approach to solving these equations. We say that two systems of equations are *equivalent* if they have the same set of solutions. Thus the process of solving a system of equations is mostly the process of converting a system into an equivalent system that is simpler.

There are three basic operations we can perform on a system of equations to get an equivalent system:

1. We can write the equations in a different order.
2. We can multiply any equation by a nonzero scalar.
3. We can add a multiple of one equation to another.

All three of these operations are guaranteed not to change the solution set; proving this is a reasonable exercise. Our goal now is to find an efficient way to use these rules to get a useful solution to our system.

But, it's possible for us to be lazy about this by encoding our system in a matrix.

Right now, we will just use this as a convenient notational shortcut; we will see later on in the course that this has a number of theoretical and practical advantages.

Definition 2.54. The *coefficient matrix* of a system of linear equations given by

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

and the *augmented coefficient matrix* is

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right].$$

Example 2.55. Suppose we have a system

$$\begin{aligned} 4x + 2y + 2z &= 8 \\ 3x + 2y + z &= 6. \end{aligned}$$

Then the coefficient matrix is

$$\begin{bmatrix} 4 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

and the augmented coefficient matrix is

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 8 \\ 3 & 2 & 1 & 6. \end{array}\right]$$

Earlier we listed three operations we can perform on a system of equations without changing the solution set: we can reorder the equations, multiply an equation by a nonzero scalar, or add a multiple of one equation to another. We can do analogous things to the coefficient matrix.

Definition 2.56. The three *elementary row operations* on a matrix are

I Interchange two rows.

II Multiply a row by a nonzero real number.

III Replace a row by its sum with a multiple of another row.

Example 2.57. What can we do with our previous matrix? We can

$$\begin{bmatrix} 4 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} \overset{I}{\to} \begin{bmatrix} 3 & 2 & 1 \\ 4 & 2 & 2 \end{bmatrix} \overset{II}{\to} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \overset{III}{\to} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}.$$

So how do we use this to solve a system of equations? The basic idea is to remove variables from successive equations until we get one equation that contains only one variable—at which point we can substitute for that variable, and then the others. To do that with this matrix, we have

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 8 \\ 3 & 2 & 1 & 6 \end{array}\right] \overset{III}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 6 \end{array}\right] \overset{III}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 2 & -2 & 0 \end{array}\right] \overset{II}{\to} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 0 \end{array}\right].$$

What does this tell us? That our system of equations is equivalent to the system

$$x + z = 2$$
$$y - z = 0.$$

This gives us the answer: $z = 2 - x$ and $y = z = 2 - x$. So the set of solutions is the set of triples $\{(x, 2 - x, 2 - x)\}$.

Example 2.58. In $\mathcal{P}_3(\mathcal{R})$, we claim that the polynomial

$$f = 2x^3 - 2x^2 + 12x - 6$$

is a linear combination of the polynomials

$$g_1 = x^3 - 2x^2 - 5x - 3 \quad \text{and} \quad g_2 = 3x^3 - 5x^2 - 4x - 9$$

but that the polynomial

$$h = 3x^3 - 2x^2 + 7x + 8$$

is not.

To show that f is a linear combination of g_1 and g_2, we need to find scalars $a_1, a_2 \in \mathbb{R}$ such that $f = a_1 g_1 + a_2 g_2$, that is

$$\begin{aligned} 2x^3 - 2x^2 + 12x - 6 &= a_1(x^3 - 2x^2 - 5x - 3) + a_2(3x^3 - 5x^2 - 4x - 9) \\ &= (a_1 + 3a_2)x^3 + (-2a_1 - 5a_2)x^2 + (-5a_1 - 4a_2)x + (-3a_1 - 9a_2). \end{aligned}$$

Therefore, we want to solve the following system of linear equations for a_1 and a_2:

$$a_1 + 3a_2 = 2$$
$$-2a_1 - 5a_2 = -2$$
$$-5a_1 - 4a_2 = 12$$
$$-3a_1 - 9a_2 = -6.$$

We write this as a matrix:

$$\left[\begin{array}{cc|c} 1 & 3 & 2 \\ -2 & -5 & -2 \\ -5 & -4 & 12 \\ -3 & -9 & -6 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & 2 \\ 0 & 1 & 2 \\ 0 & 11 & 22 \\ 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

which corresponds to the system

$$\begin{aligned} a_1 \qquad\quad &= -4 \\ a_2 &= 2 \\ 0 &= 0 \\ 0 &= 0 \end{aligned}$$

and thus we have a single solution. And indeed we can verify that $f = -4g_1 + 2g_2$.

Now let's show that h is not a linear combination of g_1 and g_2? If it were, then there would be scalars $a_1, a_2 \in \mathbb{R}$ such that

$$\begin{aligned} 3x^3 - 2x^2 + 7x + 8 &= a_1(x^3 - 2x^2 - 5x - 3) + a_2(3x^3 - 5x^2 - 4x - 9) \\ &= (a_1 + 3a_2)x^3 + (-2a_1 - 5a_2)x^2 + (-5a_1 - 4a_2)x + (-3a_1 - 9a_2). \end{aligned}$$

In other words, there would be a solution to the following system:

$$\begin{aligned} a_1 + 3a_2 &= 3 \\ -2a_1 - 5a_2 &= -2 \\ -5a_1 - 4a_2 &= 7 \\ -3a_1 - 9a_2 &= 8. \end{aligned}$$

This becomes the matrix

$$\left[\begin{array}{cc|c} 1 & 3 & 3 \\ -2 & -5 & -2 \\ -5 & -4 & 7 \\ -3 & -9 & 8 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 3 & 3 \\ 0 & 1 & 4 \\ 0 & 11 & 22 \\ 0 & 0 & 17 \end{array}\right]$$

and we can already see this system will have no solutions, because the fourth line gives us $0 = 17$, which is false.

Definition 2.59. A matrix is in *row echelon form* if

- Every row containing nonzero elements is above every row containing only zeroes; and
- The first (leftmost) nonzero entry of each row is to the right of the first nonzero entry of the above row.

Remark 2.60. Some people require the first nonzero entry in each nonzero row to be 1. This is really a matter of taste and doesn't matter much, but you should do it to be safe; it's an easy extra step to take by simply dividing each row by its leading coefficient.

Example 2.61. The following matrices are all in Row Echelon Form:

$$\begin{bmatrix} 1 & 3 & 2 & 5 \\ 0 & 3 & -1 & 4 \\ 0 & 0 & -2 & 3 \end{bmatrix} \quad \begin{bmatrix} 5 & 1 & 3 & 2 & 8 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -7 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 5 \\ 0 & -2 & 3 \\ 0 & 0 & 7 \end{bmatrix}.$$

The following matrices are not in Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 & 5 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 5 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

A system of equations sometimes has a solution, but does not always. We say a system is *inconsistent* if there is no solution; we say a system is *consistent* if there is at least one solution.

Definition 2.62. A matrix is in *reduced row echelon form* if it is in row echelon form, and the first nonzero entry in each row is the only entry in its column.

This means that we will have some number of columns that each have a bunch of zeroes and one 1. Other than that we may or may not have more columns, which can contain basically anything; we've used up all our degrees of freedom to fix those columns that contain the leading term of some row.

Note that the columns we have fixed are not necessarily the first columns, as the next example shows.

Example 2.63. The following matrices are all in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 17 & 0 & 2 & 8 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The following matrices are not in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 2 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 15 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

2.6 Spanning and Linear Independence

Recall we want to put a set of "coordinates" on our vector spaces. Any "coordinate system" will need to have two basic properties: first, we want it to represent any vector in our vector space; second, we want it to represent each vector only once. So we first want to talk about the vectors that can be represented by a given collection of vectors.

Definition 2.64. Let V be a vector space $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a set of vectors in V. We say the *span* of S is the set of all linear combinations of vectors in S, and write it $\operatorname{span}(S)$ or $\operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$.

For notational consistency, we define the span of the empty set $\operatorname{span}(\{\})$ to be the trivial vector space $\mathbf{0} = \{\mathbf{0}\}$.

Example 2.65. As before, take $V = \mathbb{R}^3$ and $S = \{(1, 0, 0), (0, 1, 0)\}$. Then

$$\operatorname{span}(S) = \{a(1, 0, 0) + b(0, 1, 0)\} = \{(a, b, 0)\}.$$

Now let $T = \{(3, 2, 0), (13, 7, 0)\}$. Then

$$\operatorname{span}(T) = \{a(3, 2, 0) + b(13, 7, 0)\} = \{(3a + 13b, 2a + 7b, 0)\}.$$

Spans are really convenient to work with because the span of any set will always be a subspace.

Proposition 2.66. *If V is a vector space over a field $\mathbb{F}$ and $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\} \subset V$, then $\operatorname{span}(S)$ is a subspace of V.*

Proof. If $S = \varnothing$ then $\operatorname{span}(S) = \{\mathbf{0}\}$ by definition, so it is the trivial subspace of V.

So now suppose S is non-empty. We know that $S \subset V$, and since any linear combination of vectors in V is itself a vector in V, we know that $\operatorname{span}(S) \subset V$. So we just need to check the three subspace conditions.

1. Because $S \neq \varnothing$, there is some vector $\mathbf{v} \in S$, and then $0 \cdot \mathbf{v} = \mathbf{0}$. This is a linear combination of vectors in S, so it is in $\operatorname{span}(S)$.

2. Suppose $\mathbf{v}_1, \mathbf{v}_2 \in \operatorname{span}(S)$. This implies that we can write

$$\mathbf{v}_1 = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n \qquad \mathbf{v}_2 = b_1\mathbf{w}_1 + \cdots + b_m\mathbf{w}_m$$

for some $a_i, b_j \in \mathbb{F}$, and some $\mathbf{v}_i, \mathbf{w}_j \in S$. Thus

$$\begin{aligned} \mathbf{v} + \mathbf{v}_2 &= \big(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n\big) + \big(b_1\mathbf{w}_1 + \cdots + b_m\mathbf{w}_m\big) \\ &= a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n + b_1\mathbf{w}_1 + \cdots + b_m\mathbf{w}_m \end{aligned}$$

is a linear combination of vectors in S, and thus an element of $\operatorname{span}(S)$.

3. Suppose $r \in \mathbb{F}$ and $\mathbf{v} \in \operatorname{span}(S)$. Then we can write

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n$$

for some $a_i \in \mathbb{F}$. Then

$$r\mathbf{v} = r\left(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{v}_n\right) = (ra_1)\mathbf{u}_1 + \cdots + (ra_n)\mathbf{u}_n \in \operatorname{span}(S).$$

Thus we see that $\operatorname{span}(S)$ is a subspace of V. □

Corollary 2.67. *Let V be a vector space over $\mathbb{F}$, let W be a subspace of V, and let $S \subset V$. If W contains V then W contains* $\operatorname{span}(S)$.

Proof. We know that W is a vector space containing S, so $\operatorname{span}(S)$ must be a subspace of W. □

Corollary 2.68. *If V is a vector space and $S \subseteq V$, then* $\operatorname{span}(S)$ *is the smallest subspace of V containing S.*

Proof. We just showed in proposition 2.66 that $\operatorname{span}(S)$ is a subspace of V, and of course it contains S. So we just need to show that there's no smaller subspace. In particular, I'll prove that if W is a subspace of V, and $S \subseteq W$, then $\operatorname{span}(S) \subseteq W$.

So suppose W is a subspace of V and $S \subseteq W$. Let $\mathbf{v} \in \operatorname{span}(S)$. The $\mathbf{v}$ is a linear combination of vectors in S. But $S \subseteq W$, so $\mathbf{v}$ is a linear combination of vectors in W, and thus an element of W since W is a vector space. Thus any element of $\operatorname{span}(S)$ is an element of W, so $\operatorname{span}(S) \subseteq W$. □

Definition 2.69. Let V be a vector space and $S \subset V$. If $\operatorname{span}(S) = V$ then we say S *spans* V, or *generates* V, or is a *spanning set* for V.

If S spans V, then we can express any element of V purely in terms of elements of S. But this expression might not be unique! Thus we need to introduce a second concept.

Definition 2.70. Let V be a vector space over F, and $S \subset V$. We say S is *linearly independent* if, for any finite collection of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in S$, the only scalars solving the equation

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

are $a_1 = \cdots = a_n = 0$.

If a set of vectors is not linearly independent, we call it *linearly dependent* and there is a *linear dependence* relationship among the vectors.

Remark 2.71. This is one of the more subtle definitions in this course, and often gives people a lot of trouble when they first start working with it. In particular, it features a problem with *nested conditionals*: a set of vectors is linearly independent if, *if* there is a linear combination equal to zero, then all of the coefficients must be zero. I didn't use that phrasing in the formal definition because it's incredibly awkward to have to instances of the word "if" in a row, but that does highlight the problem.

In particular, to prove a set is linearly independent, you shouldn't try to prove that any linear combination is equal to zero. And you shouldn't try to prove that a particular set of coefficients is zero. Instead you should start out with the *hypothesis* that a finite linear combination of vectors produces zero, and then prove that all of the coefficients must have been zero.

(In practice this will almost always involve solving a system of linear equations, and thus row reducing a matrix.)

Example 2.72.

1. The set $S = \{(1,0,0),(0,1,0),(0,0,1)\}$ is linearly independent: suppose

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

Then we have the system of equations $a = 0, b = 0, c = 0$ and thus all the scalars are zero.

2. The set $S = \{(1,0,0),(0,1,0)\}$ is linearly independent. Suppose

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

Then we have the system of equations $a = 0, b = 0$ and thus all the scalars are zero.

3. The set S $= \{(1,0,0),(0,1,0),(1,1,0)\}$ is not linearly independent, since

$$1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

4. Any set containing the zero vector is linearly dependent, since $1 \cdot \mathbf{0} = \mathbf{0}$ but $1 \neq 0$.

Example 2.73. The set $S = \{1, x, x^2, x^3\}$ is linearly independent in $\mathcal{P}_3(x)$. So is the set $T = \{1 + x + x^2 + x^3, 1 + x + x^2, 1 + x, 1\}$.

Theorem 2.74. *Let V be a vector space and let $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then S_2 is linearly dependent.*

Proof. If S_1 is linearly dependent, then there are vectors $u_1, \dots, u_n \in S_1$ and scalars not all zero $a_1, \dots, a_n \in \mathbb{F}$ such that

$$a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n = \mathbf{0}.$$

But since $S_1 \subseteq S_2$, therefore each $\mathbf{u}_i \in S_2$. So the previous equation shows that S_2 is linearly dependent by definition. □

Corollary 2.75. *Let V be a vector space and let $S_1 \subseteq S_2 \subseteq V$. If S_2 is linearly independent, then S_1 is linearly independent.*

Proof. This is just the contrapositive of the previous theorem. □

From an intuitive standpoint, these two results make sense. If S_1 is linearly dependent, then has some sort of redundancy. But since $S_1 \subseteq S_2$, therefore S_2 also contains redundant vectors. Adding more vectors to a redundant set cannot make the set less redundant. So S_2 must be linearly dependent.

Similarly, if S_2 is linearly independent, then the vectors in S_2 point in “genuinely different directions”. Taking a subset, those vectors will still point in “genuinely different directions”.

Recall that earlier we saw that the set $\{(1, 0, 0), (0, 1, 0), (1, 1, 0)\}$ was linearly dependent. But we might notice that we can remove a vector and get a linearly independent set with the same span—we can just get rid of the redundancy. Conversely, we can start with the linearly independent set $\{(1, 0, 0), (0, 1, 0)\}$ and try to add a vector. If that vector is in the span, then it will be redundant, and we get a linearly dependent set. But if it’s *not* in the span, it’s not redundant, and we get an independent set.

Theorem 2.76. *Let S be a linearly independent subset of a vector space V, and let $\mathbf{v} \in V$. Then $S \cup \{\mathbf{v}\}$ is linearly dependent if and only if $\mathbf{v} \in \operatorname{span}(S)$.*

Proof. Suppose S is a linearly independent subset of a vector space V and let $v \in V$.

[⇒]. Suppose that $S \cup \{\mathbf{v}\}$ is linearly dependent. Then there are distinct vectors $\mathbf{u}_1, \dots, \mathbf{u}_n \in S$ and scalars $a_1, \dots, a_n, b \in \mathbb{F}$ not all zero such that

$$a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n + bv = \mathbf{0}.$$

Now we claim that $b \neq 0$, since if it were 0, then we could delete it from the above equation to get a nontrivial linear combination of the $\mathbf{u}_i$'s to equal $\mathbf{0}$, which is not possible since S itself is linearly independent.

Since $b \neq 0$, it has a multiplicative inverse, so we can write

$$\mathbf{v} = b^{-1}(-a_1\mathbf{u}_1 - a_2\mathbf{u}_2 - \cdots - a_n\mathbf{u}_n).$$

This shows that $\mathbf{v} \in \operatorname{span}(S)$.

[$\Leftarrow$]. Conversely, suppose $\mathbf{v} \in \operatorname{span}(S)$. Then there exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in S$ and scalars $b_1, \ldots, b_m \in \mathbb{F}$ such that

$$\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m.$$

Hence,

$$b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m - v = \mathbf{0}.$$

Note that $\mathbf{v} \neq \mathbf{v}_i$ for any i, since we are assuming $\mathbf{v} \notin S$. Hence, this is a nontrivial linear combination of the vectors in $S \cup \{\mathbf{v}\}$ which equals $\mathbf{0}$ (it is nontrivial since the coefficient on $\mathbf{v}$ is -1). Thus, $S \cup \{\mathbf{v}\}$ is linearly dependent. $\square$

2.7 Bases and Dimension

Now we're ready to introduce our idea of coordinates. Recall we wanted a set S such that we could write any vector in V as a sum of vectors in S, but only one way. With our new notation, we can define:

Definition 2.77. If V is a vector space and S is a spanning set for V that is also linearly independent, we say that S is a *basis* for V.

Example 2.78. The set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is a basis for $\mathbb{R}^3$, as we have seen before. We call this set the *standard basis* for $\mathbb{R}^3$, and we write the three elements $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$.

We can generalize this to $\mathbb{R}^n$. We define the *standard basis vectors* for $\mathbb{R}^n$ by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \ldots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

and the set of standard basis vectors is the *standard basis*. You can check that the standard basis is in fact a basis.

Example 2.79. Every (non-trivial) vector space has more than one basis. The set $S = \{(1,0,0),(1,1,0),(1,1,1)\}$ is a basis for $\mathbb{R}^3$:

First we show that it is a spanning set. Let $(a,b,c) \in \mathbb{R}^3$. Then we want to solve

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

which gives the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & a \\ 0 & 1 & 1 & b \\ 0 & 0 & 1 & c \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & a-b \\ 0 & 1 & 0 & b-c \\ 0 & 0 & 1 & c \end{array}\right]$$

which tells us that $\alpha_3 = c, \alpha_2 = b-c, \alpha_1 = a-b$. Thus there is a solution for any $(a,b,c) \in \mathbb{R}^3$, and the set spans.

We also need to prove linear indepencence. So suppose

$$\mathbf{0} = \alpha_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

This gives us a system of linear equations corresponding to the homogeneous system

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

so the only solution here is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Thus S is linear independent, and since it also spans, it is a basis.

Example 2.80. The set $S = \{(1,0,0),(0,1,0)\}$ is not a basis for $\mathbb{R}^3$. It is linearly independent (since it is a subset of the standard basis, which is linear independent), but it is not a spanning set, since $(0,0,1)$ is not in the span of S.

Example 2.81. The set $S = \{(2,3),(3,4),(4,4)\}$ is a spanning set for $\mathbb{R}^2$ but not a basis. To see that it's a spanning set we solve

$$\begin{bmatrix} a \\ b \end{bmatrix} = \alpha_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \alpha_3 \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2\alpha_1 + 3\alpha_2 + 4\alpha_3 \\ 3\alpha_1 + 4\alpha_2 + 4\alpha_3 \end{bmatrix}$$

giving the system of equations

$$a = 2\alpha_1 + 3\alpha_2 + 4\alpha_3 \qquad\qquad b = 3\alpha_1 + 4\alpha_2 + 4\alpha_3$$

and the augmented matrix

$$\left[\begin{array}{ccc|c} 2 & 3 & 4 & a \\ 3 & 4 & 4 & b \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & b-a \\ 2 & 3 & 4 & a \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & b-a \\ 0 & 1 & 4 & 3a-2b \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -4 & 3b-4a \\ 0 & 1 & 4 & 3a-2b \end{array}\right].$$

Thus for any $(a, b) \in \mathbb{R}^2$, at least one solution exists; in fact we can pick α_3 to be any real number and we get a corresponding solution $(3b - 4a + 4\alpha_3, 3a - 2b - 4\alpha_3, \alpha_3)$. Thus the set spans.

But S is not linearly independent. We can see this in a few ways. Most easily we can observe that $(2, 3) + (1/4)(4, 4) = (3, 4)$. If we can't see that on our own, we can do a couple things. We can find the nullspace:

$$\begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 4 \end{bmatrix}$$

and we see the nullspace $\{(4\alpha, -4\alpha, \alpha)\}$ is non-trivial, so the set is not linearly independent.

But if these row operations seem familiar, that's because we did exactly the same thing to check spanning! So we can look at our spanning equations and try to find all the solutions when we take $a = b = 0$. We see that there's more than one solution there, so the vectors aren't linearly independent.

Determining whether a set is a basis is sometimes annoying, but doesn't involve anything we haven't already done: a basis is just a set that both spans and is linearly independent, and we can check both properties individually. But we'd like to make things a little simpler.

Further, we want to talk about how "big" a space is, and this should plausibly be determined by how many elements there are in the basis. But since every space has more than one basis, talking about the size of "the" basis is potentially problematic. Fortunately, this is not an actual problem, as we shall see.

Lemma 2.82. *If $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ spans a vector space V, and $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is a collection of vectors in V with $m > n$, then T is linearly dependent.*

Proof. There are two possible ways to prove this. One involves simply writing out a bunch of linear equations and solving them; this works, but is more tedious than informative. We'll use a more formal and abstract approach to proving this instead, which, hopefully, will actually explain some of *why* this is true.

We will start with the set S, and one by one we will trade out vectors in S for vectors in T, and show that we always still have a spanning set. We will suppose T is linearly independent, and show that $m \leq n$.

Since S is a spanning set, we know that $\mathbf{u}_1 \in \operatorname{span}(S)$, and thus $\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}$ is linearly dependent.. Then we can rewrite our linear dependence equation to express $\mathbf{v}_1$ (without loss of generality) as a linear combination of $\{\mathbf{u}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = S_1$, and thus

$$\operatorname{span}(S) = \operatorname{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}) = \operatorname{span}(S_1).$$

We can repeat this process: at every step we add the next vector from T to get the set $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_k, \dots, \mathbf{v}_n\}$. Since S_{k-1} is a spanning set, this set is linearly dependent; since the $\mathbf{u}_i$ are linearly independent by hypothesis, we can remove one of the $\mathbf{v}_i$, and without loss of generality we can remove $\mathbf{v}_k$, to obtain the set $S_k = \{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$.

If $m > n$, we can continue until we have replaced every $\mathbf{v}_i$. Then we have $S_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a spanning set, and thus $\mathbf{u}_{n+1} \in \operatorname{span}(S_n)$ and so T is linearly dependent, which contradicts our assumption.

Thus if T is linearly independent, we must have $m \leq n$. Conversely, if $m > n$ then T is linearly dependent, as we stated. □

Corollary 2.83. *$S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ are two bases for a space V, then they are the same size, i.e. $m = n$.*

Proof. S is a spanning set and T is linearly independent, so we can't have $m > n$ by lemma 2.82. But T is a spanning set and S is linearly independent, so we can't have $n > m$ by lemma 2.82. Thus $n = m$. □

Definition 2.84. Let V be a vector space. If V has a basis consisting of n vectors, we say that V has *dimension* n and write $\dim V = n$. The trivial vector space $\{\mathbf{0}\}$ has dimension 0.

We say that V is *finite–dimensional* if there is a finite set of vectors that spans V. (Thus if V is n-dimensional it is finite-dimensional). Otherwise, we say that V is *infinite–dimensional*.

In this course we will primarily discuss finite dimensional vector spaces; but there are many important infinite-dimensional examples.

Example 2.85. The set of standard basis vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$, so $\mathbb{R}^n$ is n-dimensional.

The set $\{1, x, \dots, x^n\}$ is a basis for $\mathcal{P}_n(x)$. This set has $n+1$ vectors, so $\dim \mathcal{P}_n(x) = n+1$.

$\mathcal{P}(x)$ does not have a finite basis. We can see this since the set $S = \{1, x, \dots, x^n\}$ is linearly independent for any n; but every spanning set is at least as big as any linearly independent set, so we can never have a finite spanning set. However, if we allow infinite bases, then $\{1, x, \ldots, x^n, \ldots\}$ is a basis for $\mathcal{P}(x)$.

Remark 2.86. $\mathcal{C}([a,b], \mathbb{R})$ is infinite–dimensional, but if we allow infinite sums and make convergence arguments it is possible to think of the set $\{1, x, \ldots, x^n, \ldots\}$ as a sort of ("separable") basis. But this requires analysis and is outside the scope of this course. We can also build a (separable) basis out of the functions $\sin(nx)$ and $\cos(nx)$ for $n \in \mathbb{N}$; this is the foundation of Fourier analysis and Fourier series.

The set $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is absurdly huge, and does not have a countable basis. If you believe the axiom of choice it has a basis, as all sets do, but you can't possibly write it down. You can think of it has having "coordinates" given by functions like

$$f_r(x) = \begin{cases} 1 & x = r \\ 0 & x \neq r \end{cases}$$

but this isn't a basis because it would require uncountable sums, which you can't really define.

How do we find bases? There are two basic ways we can build them.

Lemma 2.87 (Basis Reduction)**.** *Suppose $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a spanning set for V. Then S can be reduced to a basis for V. That is, there is a subset $B \subseteq S$ that is a basis for V.*

Proof. If S is linearly independent, then it is a basis and we're done.

So suppose S is linearly dependent. Then we know at least one vector is redundant, so without loss of generality we can reorder the set so that we can write $\mathbf{v}_n$ as a linear combination of the other vectors in S.

But then $\operatorname{span}(S) = \operatorname{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\})$, and $S_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$ is a spanning set for V and a proper subset of S. If S_1 is linearly independent, then it is a basis; if not, we can repeat this process until we reach a linearly independent set, which is our basis B. □

Remark 2.88. This proof assumes that S is finite. The result is still (mostly) true if S is infinite, but if the space is finite-dimensional this isn't interesting, and if the space is infinite-dimensional things get very complicated and we don't want to worry about them here.

Example 2.89. Let $S = \{(1,1,0), (1,1,1), (0,0,1), (2,7,0)\}$ be a spanning set for $\mathbb{R}^3$. Find a basis $B \subseteq S$ for $\mathbb{R}^3$.

We'll take as given that this is a spanning set, which is not difficult to check. We see that we can write $(1,1,1) = (1,1,0) + (0,0,1)$, so we can remove $(1,1,1)$ without changing the span, and we have $B = \{(1,1,0), (0,0,1), (2,7,0)\} \subseteq S$ is a basis for $\mathbb{R}^3$.

Lemma 2.90 (Basis Padding). *Suppose $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent in V. Then if V has any finite spanning set $T = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$, we can obtain a basis by padding S. That is, there is a basis B for V with $S \subseteq B$.*

Proof. If $T \subset \operatorname{span}(S)$, then $\operatorname{span}(T) \subset \operatorname{span}(S)$, so S is a spanning set for V and thus a basis, so we're done.

So suppose without loss of generality that $\mathbf{u}_1 \notin \operatorname{span}(S)$. Then $S_1 = \{\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{u}_1\}$ is linearly independent since we can't write any element as a linear combination of the others.

If S_1 spans V, then it is a basis and we're done. If not, there is some other $\mathbf{u}_i \notin \operatorname{span}(S_1)$, so we can repeat the process, and after at most m steps this process will terminate (since we run out of elements in T). When we reach a spanning set, this is our basis.

□

Example 2.91. Let $S = \{1 + x, x^2 - 3\} \subset \mathcal{P}_2(x)$. Can we find a basis B for $\mathcal{P}_2(x)$ that contains T?

We need to find a vector (or quadratic polynomial) that isn't in S. There are lots of choices here, but it looks to me like 1 is not in the span of S. Then we check: suppose $a(1 + x) + b(x^2 - 3) = 1$. Then we have

$$(a - 3b) + ax + bx^2 = 1$$

which gives the system

$$\left[\begin{array}{cc|c} 1 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

which has no solution. Thus indeed $1 \notin \operatorname{span}(S)$, so $\{1, 1 + x, x^2 - 3\}$ is a basis for $\mathcal{P}_3(x)$.