Vector spaces and bases

COMS 3251 Fall 2022 (Daniel Hsu)

1 Vector spaces

The n-dimensional Cartesian space $\mathbb{R}^n$ is an example of a vector space. Informally, vector spaces are collections of objects that can be "added" together and that can be "scaled", and these notions of adding and scaling behave in the way we would expect from standard algebra. Besides n-tuples of numbers, many other collections of objects share these properties, including polynomials, matrices, functions, and sequences; we just need to suitably define what it means to add and scale these objects.

Formally, a (real) *vector space* is a mathematical structure consisting of a set $\mathbb{V}$ of objects, which we call *vectors*, bundled together with two operations, $\mathsf{add}\colon \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ and $\mathsf{scale}\colon \mathbb{R} \times \mathbb{V} \to \mathbb{V}$, described as follows.

- *Vector addition*: Given any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{V}$, $\mathsf{add}(\mathbf{u}, \mathbf{v})$ returns another vector in $\mathbb{V}$. We usually write this as $\mathbf{u} + \mathbf{v}$.
- *Scalar multiplication*: Given any scalar $c \in \mathbb{R}$ and vector $\mathbf{v} \in \mathbb{V}$, $\mathsf{scale}(c, \mathbf{v})$ returns another vector in $\mathbb{V}$. We usually write this as $c\mathbf{v}$.

These operations must satisfy the following so-called vector space properties:

VS1 (Vector addition is commutative.) For all $\mathbf{u}, \mathbf{v} \in \mathbb{V}$,

$$\mathsf{add}(\mathbf{u}, \mathbf{v}) \;=\; \mathsf{add}(\mathbf{v}, \mathbf{u}).$$

VS2 (Vector addition is associative.) For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{V}$,

$$\mathsf{add}(\mathbf{u}, \mathsf{add}(\mathbf{v}, \mathbf{w})) \;=\; \mathsf{add}(\mathsf{add}(\mathbf{u}, \mathbf{v}), \mathbf{w}).$$

VS3 (Vector addition has an identity.) There exists a vector $\mathbf{0}_{\mathbb{V}} \in \mathbb{V}$ (called the *zero vector*) such that for all $\mathbf{v} \in \mathbb{V}$,

$$\mathsf{add}(\mathbf{0}_{\mathbb{V}}, \mathbf{v}) \;=\; \mathbf{v}.$$

(We'll often omit the subscript $\mathbb{V}$ in $\mathbf{0}_{\mathbb{V}}$.)

VS4 (Vector addition has inverses.) For every $\mathbf{v} \in \mathbb{V}$, there exists a vector $\mathbf{u} \in \mathbb{V}$, called an *additive inverse* of $\mathbf{v}$, such that

$$\mathsf{add}(\mathbf{u}, \mathbf{v}) \;=\; \mathbf{0}_{\mathbb{V}}.$$

An additive inverse of $\mathbf{v}$ is usually written as $-\mathbf{v}$.[1]

VS5 (Scalar multiplication has an identity.) For all $\mathbf{v} \in \mathbb{V}$,

$$\mathsf{scale}(1, \mathbf{v}) \;=\; \mathbf{v}$$

(where $1 \in \mathbb{R}$ is the multiplicative identity in $\mathbb{R}$).

VS6 (Scalar multiplication is associative.) For all $\mathbf{v} \in \mathbb{V}$ and $c, d \in \mathbb{R}$,

$$\mathsf{scale}(c, \mathsf{scale}(d, \mathbf{v})) \;=\; \mathsf{scale}(c\,d, \mathbf{v}).$$

VS7 (Scalar multiplication distributes over vector addition.) For all $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ and $c \in \mathbb{R}$,

$$\mathsf{scale}(c, \mathsf{add}(\mathbf{u}, \mathbf{v})) \;=\; \mathsf{add}(\mathsf{scale}(c, \mathbf{u}), \mathsf{scale}(c, \mathbf{v})).$$

VS8 (Scalar multiplication distributes over scalar addition.) For all $\mathbf{v} \in \mathbb{V}$ and $c, d \in \mathbb{R}$,

$$\mathsf{scale}(c + d, \mathbf{v}) \;=\; \mathsf{add}(\mathsf{scale}(c, \mathbf{v}), \mathsf{scale}(d, \mathbf{v})).$$

What we have defined above are vector spaces over the real numbers $\mathbb{R}$. One can generalize even further (though we will not do so) to vector spaces over certain other "number systems" called *fields*, such as the rational numbers, the complex numbers, and the integers modulo a prime power q.

The vector space properties may seem rather "obvious" for $\mathbb{R}^n$. But their purpose is to spell-out the required properties for (other) vector spaces that will make them behave like $\mathbb{R}^n$ in the sense that we have studied it so far. Thus, we can develop techniques for working with all of these vector spaces in a unified way. After all, we don't have different arithmetic rules for kilograms and meters, even though mass and distance are different types of scalars.

The vector space properties also help us understand which (possibly intuitive) properties of $\mathbb{R}^n$ are not necessarily shared by other vector spaces. For instance, the notion of "distance" between vectors is something that goes beyond vector spaces, but it is one we may be accustomed to in $\mathbb{R}^n$.

[1] We don't need to assume that additive inverses are unique, as it will be implied by other properties.

Note. Some of the notation we have been using are not technically operations explicitly specified by a vector space, such as $\mathbf{u}+\mathbf{v}+\mathbf{w}$ and $\mathbf{u}-\mathbf{v}$. We are really using shorthand and appealing to some vector space properties to ensure that these notations are unambiguous:

$$\mathbf{u}+\mathbf{v}+\mathbf{w} \;=\; \mathbf{u}+(\mathbf{v}+\mathbf{w}) \;=\; (\mathbf{u}+\mathbf{v})+\mathbf{w},$$

where the latter equality is by VS2; and

$$\mathbf{u}-\mathbf{v} \;=\; \mathbf{u}+(-\mathbf{v}),$$

where the meaning of $-\mathbf{v}$ comes from VS4. ∎

In this course, our primary focus will remain on $\mathbb{R}^n$ (and subspaces of $\mathbb{R}^n$, defined next), although it is useful to keep in mind other common vector spaces to appreciate the wide applicability of linear algebra. We will take examples from some of these other vector spaces.

Examples.

1. *The polynomials (with real coefficients)*, denoted by $\mathsf{P}(\mathbb{R})$. A *polynomial* in $\mathsf{P}(\mathbb{R})$ is an expression of the form

$$a_0 t^0 + a_1 t + a_2 t^2 + \cdots + a_d t^d.$$

In this expression,

- t is a *variable* (a.k.a. *indeterminant*, *formal variable*); and t^k, for any non-negative integer k, is the kth power of t;
- d can be any non-negative integer; and
- the a_k's can be any real numbers, and a_k is called the *coefficient* of t^k in the polynomial f.
- The coefficient for t^k is 0 if t^k does not appear in the expression. We also usually drop t^0 from the expression and just write $a_0 + a_1 t + a_2 t^2 + \cdots + a_d t^d$.

The *degree* of a polynomial $f(t) \in \mathsf{P}(\mathbb{R})$ is the largest k such that f has a non-zero coefficient for t^k. Two polynomials $f(t)$ and $g(t)$ are equal

if they have the same degree, and if, for each k, their coefficients for t^k are the same. The zero polynomial $0(t)$ has coefficient 0 for all t^k (and, by convention, we say its degree is -1).

To add polynomials $f(t)$ and $g(t)$, we create a new polynomial for which the coefficient of each t^k is the sum of the corresponding coefficients of $f(t)$ and $g(t)$. And to scale a polynomial $f(t)$ by the scalar c, we create a new polynomial for which the coefficient of each t^k is obtained by multiplying the corresponding coefficient of $f(t)$ by c.

2. *The polynomials (with real coefficients) of degree at most d*, denoted by $\mathsf{P}_d(\mathbb{R})$.

   The rules for adding and scaling are just as they are for $\mathsf{P}(\mathbb{R})$. We just need to ensure that these operations always leave us with a polynomial of degree at most d, and indeed they do.

   The vector space $\mathsf{P}_d(\mathbb{R})$ resembles $\mathbb{R}^{d+1}$.[2] On the other hand, $\mathsf{P}(\mathbb{R})$ is rather different.

3. *The $m \times n$ matrices (with real entries)*, denoted by $\mathsf{M}_{m\times n}(\mathbb{R})$.

   We have already defined how to add and scale such matrices.

   The vector space $\mathsf{M}_{m\times n}(\mathbb{R})$ resembles $\mathbb{R}^{mn}$.

4. *The (real-valued) continuous functions on $\mathbb{R}$*, denoted by $\mathsf{C}(\mathbb{R}, \mathbb{R})$.

   Recall, from calculus, that functions $f \in \mathsf{C}(\mathbb{R}, \mathbb{R})$, such as such as $f(t) = \sin(t)$ and $f(t) = e^t$, are those that satisfy some tedious property involving epsilons and deltas.[3]

   To add two functions $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$, we create a new function $h\colon \mathbb{R} \to \mathbb{R}$ such that $h(t) = f(t) + g(t)$ for all $t \in \mathbb{R}$. And to scale a function $f\colon \mathbb{R} \to \mathbb{R}$ by $c \in \mathbb{R}$, we create a new function $h\colon \mathbb{R} \to \mathbb{R}$ such that $h(t) = cf(t)$ for all $t \in \mathbb{R}$. Checking that $f + g \in \mathsf{C}(\mathbb{R}, \mathbb{R})$ and $cf \in \mathsf{C}(\mathbb{R}, \mathbb{R})$ can be done using an argument involving epsilons and deltas.

[2] In particular, $\mathsf{P}_d(\mathbb{R})$ and $\mathbb{R}^{d+1}$ are *isomorphic*, which means that there is a bijective linear map between the two spaces.

[3] A function $f\colon \mathbb{R} \to \mathbb{R}$ is *continuous* if, for every $t_0 \in \mathbb{R}$ and every $\epsilon > 0$, there exists $\delta > 0$ such that $f(t) \in (f(t_0) - \epsilon, f(t_0) + \epsilon)$ for every $t \in (t_0 - \delta, t_0 + \delta)$.

The vector space $\mathsf{C}(\mathbb{R}, \mathbb{R})$ is very different from $\mathbb{R}^n$, and also very different from $\mathsf{P}(\mathbb{R})$.

5. *The (real-valued) continuous functions on* $[0, 1]$, denoted by $\mathsf{C}([0, 1], \mathbb{R})$.

   This similar to $\mathsf{C}(\mathbb{R}, \mathbb{R})$, except that we only consider the behavior of functions on the interval $[0, 1]$. For example, $f(t) = |t|$ and $g(t) = t$ are different functions when considered in $\mathsf{C}(\mathbb{R}, \mathbb{R})$, but they are the same function in $\mathsf{C}([0, 1], \mathbb{R})$.

6. *The (real-valued) functions on* $\mathcal{S}$, denoted by $\mathbb{R}^{\mathcal{S}}$. Here, $\mathcal{S}$ may be any set (e.g., $\mathcal{S} = \mathbb{R}$, $\mathcal{S} = \mathbb{N}$, $\mathcal{S} = \{1, \ldots, n\}$, $\mathcal{S} = \{1, \ldots, m\} \times \{1, \ldots, n\}$).

   The rules for adding and scaling functions in $\mathbb{R}^{\mathcal{S}}$ are the same as that for $\mathsf{C}(\mathbb{R}, \mathbb{R})$, except we consider $t \in \mathcal{S}$.

   Notice that $\mathbb{R}^{\mathcal{S}}$ for $\mathcal{S} = \{1, \ldots, n\}$ resembles $\mathbb{R}^n$, and $\mathbb{R}^{\mathcal{S}}$ for $\mathcal{S} = \{1, \ldots, m\} \times \{1, \ldots, n\}$ resembles $\mathsf{M}_{m \times n}(\mathbb{R})$. The set of all real-valued functions on $[0, 1]$ is $\mathbb{R}^{[0,1]}$, which contains $\mathsf{C}([0, 1], \mathbb{R})$.

7. *Vector space containing only the zero vector* $\{\mathbf{0}\}$. A minimal vector space is one that only contains the zero vector (mandated by VS3). Note that the zero vector can be interpreted however you like (e.g., a zero polynomial, a matrix with all entries equal to 0). There's not any point in defining how to add or scale vectors in this space, since the output must always be $\mathbf{0}$. ∎

Because general vector spaces permit linear combinations, the concept of linear independence—which we had defined for $\mathbb{R}^n$—can be directly "ported" over to general vector spaces. Also, many of the theorems we have seen for $\mathbb{R}^n$ also hold for general vector spaces (with cosmetic changes). Here are some of them:

Theorem 1 (Unique Representations Theorem)**.** *Suppose* $\mathbb{V}$ *is a vector space,* $\{\mathbf{b}_1, \ldots \mathbf{b}_n\}$ *is a set of* n *linearly independent vectors from* $\mathbb{V}$*, and*

$$x_1 \mathbf{b}_1 + \cdots + x_n \mathbf{b}_n \;=\; y_1 \mathbf{b}_1 + \cdots + y_n \mathbf{b}_n.$$

Then $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$*.*

Theorem 2 (Growth Theorem)**.** *Let* $\mathcal{S}$ *be a set of vectors from a vector space* $\mathbb{V}$*, and let* $\mathbf{v}$ *be a vector not in* $\mathcal{S}$*.*

- *If* $\mathbf{v} \in \mathrm{span}(\mathcal{S})$*, then* $\mathcal{S} \cup \{\mathbf{v}\}$ *is linearly dependent and*

$$\mathrm{span}(\mathcal{S}) \;=\; \mathrm{span}(\mathcal{S} \cup \{\mathbf{v}\}).$$

- *If* $\mathbf{v} \notin \mathrm{span}(\mathcal{S})$*, then*

$$\mathrm{span}(\mathcal{S}) \;\subsetneq\; \mathrm{span}(\mathcal{S} \cup \{\mathbf{v}\});$$

  and if, additionally, $\mathcal{S}$ *is linearly independent, then so is* $\mathcal{S} \cup \{\mathbf{v}\}$*.*

Theorem 3 (Removal Theorem)**.** *Let* $\mathcal{S}$ *be a set of vectors from a vector space* $\mathbb{V}$*.*

- *If* $\mathcal{S}$ *is linearly dependent, then there is a vector* $\mathbf{v} \in \mathcal{S}$ *such that*

$$\mathrm{span}(\mathcal{S} \setminus \{\mathbf{v}\}) \;=\; \mathrm{span}(\mathcal{S}).$$

- *If* $\mathcal{S}$ *is linearly independent, then every proper subset* $\mathcal{S}' \subsetneq \mathcal{S}$ *is linearly independent and*

$$\mathrm{span}(\mathcal{S}') \;\subsetneq\; \mathrm{span}(\mathcal{S}).$$

Theorem 4 (Exchange Theorem)**.** *Let* $\mathcal{E}$ *and* $\mathcal{W}$ *be finite sets of vectors from a vector space* $\mathbb{V}$ *such that* $\mathrm{span}(\mathcal{E}) = \mathbb{V}$*. If* $\mathcal{W}$ *is a linearly independent subset of* $\mathbb{V}$*, then*

- $|\mathcal{W}| \leq |\mathcal{E}|$*, and*
- *there is a subset* $\mathcal{F} \subseteq \mathcal{E}$ *with* $|\mathcal{F}| = |\mathcal{E}| - |\mathcal{W}|$ *such that* $\mathbb{V} = \mathrm{span}(\mathcal{W} \cup \mathcal{F})$*.*

2 Subspaces

Suppose $\mathbb{V}$ is a vector space (over $\mathbb{R}$). We say a subset $\mathbb{W} \subseteq \mathbb{V}$ is a *subspace* of $\mathbb{V}$ if it is a vector space with the same vector addition and scalar multiplication operations as $\mathbb{V}$.

To check that $\mathbb{W} \subseteq \mathbb{V}$ is a vector space, we need to verify the following:

- The vector addition and scalar multiplication operations, which we defined for $\mathbb{V}$, only return vectors in $\mathbb{W}$ when used on vectors from $\mathbb{W}$.

- The vector space properties VS1–VS8 hold for $\mathbb{W}$.

Fortunately, many of the vector space properties hold for $\mathbb{W}$ because they hold for vectors in $\mathbb{V}$. So, it turns out to be necessary and sufficient for $\mathbb{W} \subseteq \mathbb{V}$ to satisfy the following:

SS1 ($\mathbb{W}$ is closed under vector addition.) $\mathsf{add}(\mathbf{u}, \mathbf{v}) \in \mathbb{W}$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{W}$.

SS2 ($\mathbb{W}$ is closed under scalar multiplication.) $\mathsf{scale}(c, \mathbf{v}) \in \mathbb{W}$ for all $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{W}$.

SS3 (VS3 holds for $\mathbb{W}$.) $\mathbf{0} \in \mathbb{W}$.

SS1 and SS2 make sure that the vector addition and scalar multiplication only return vectors in $\mathbb{W}$ when used on vectors in $\mathbb{W}$. We already know that VS1, VS2, and VS5–VS8 hold for vectors in $\mathbb{V}$, so they also hold for vectors in $\mathbb{W}$; SS3 ensures that VS3 holds for $\mathbb{W}$.

It may seem like we forgot to ensure that VS4 holds for $\mathbb{W}$. But it turns out that we get it for free (after ensuring SS1–SS3). Indeed, for any $\mathbf{v} \in \mathbb{V}$, the unique additive inverse of $\mathbf{v}$ is $\mathsf{scale}(-1, \mathbf{v})$ (as proved in Theorem 11 from Appendix B). For $\mathbf{v} \in \mathbb{W}$, SS2 ensures that $\mathsf{scale}(-1, \mathbf{v}) \in \mathbb{W}$ as well, so the unique additive inverse of any vector in $\mathbb{W}$ is also contained in $\mathbb{W}$.

Examples.

1. Both $\mathbb{V}$ and $\{\mathbf{0}\}$ are subspaces of $\mathbb{V}$ for any vector space $\mathbb{V}$.
2. $\mathsf{P}_d(\mathbb{R})$ is a subspace of $\mathsf{P}(\mathbb{R})$ for any non-negative integer d.
3. $\mathsf{P}_d(\mathbb{R})$ is a subspace of $\mathsf{P}_{d'}(\mathbb{R})$ whenever $d \leq d'$.
4. We say a polynomial $f(t) \in \mathsf{P}(\mathbb{R})$ is *even* if $f(t) = f(-t)$. An example of an even polynomial is $f(t) = 2t^4 + t^2 + 3$, and an example of a non-even polynomial is $f(t) = t^5$. The even polynomials $\mathsf{P}^{\mathsf{even}}(\mathbb{R})$ is a subspace of $\mathsf{P}(\mathbb{R})$.
5. $\mathsf{C}([0, 1], \mathbb{R})$ is a subspace of $\mathbb{R}^{[0,1]}$.

6. We say a function $f : \mathbb{R} \to \mathbb{R}$ is *periodic* with *period* $P > 0$ if $f(t+P) = f(t)$ for all $t \in \mathbb{R}$. The subset $\mathsf{C}_{\mathsf{periodic}}([0,1], \mathbb{R})$ of $\mathsf{C}([0,1], \mathbb{R})$ that are periodic with period 1 is a subspace of $\mathsf{C}([0,1], \mathbb{R})$. In the context of functions defined on $[0,1]$, periodicity simply requires $f(0) = f(1)$.

   (What about additional constraints, such as $f(0) = f(1/3) = f(1)$?)

7. $\{(x, y, 0) : (x, y) \in \mathbb{R}^2\}$ is a subspace of $\mathbb{R}^3$.

   (What about $\{(x, y, 1) : (x, y) \in \mathbb{R}^2\}$?)

8. If $\mathbb{V}$ is a vector space, and if $\mathbb{W}_1$ and $\mathbb{W}_2$ are both subspaces of $\mathbb{V}$, then $\mathbb{W}_1 \cap \mathbb{W}_2$ is also a subspace of $\mathbb{V}$.

   (What about $\mathbb{W}_1 \cup \mathbb{W}_2$?)

9. If $\mathbb{V}$ is a vector space and $\mathcal{S} \subseteq \mathbb{V}$, then $\mathrm{span}(\mathcal{S})$ is a subspace of $\mathbb{V}$.

   Let us verify this example, since it is particularly important in this course. First we establish SS1 and SS2. Suppose $\mathbf{u}$ and $\mathbf{v}$ are in $\mathrm{span}(\mathcal{S})$. This means we can write each of $\mathbf{u}$ and $\mathbf{v}$ as a linear combination of vectors from $\mathcal{S}$. Let $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subseteq \mathcal{S}$ be the union of the vectors involved in these linear combinations, and write $\mathbf{u} = x_1\mathbf{s}_1 + \cdots + x_n\mathbf{s}_n$ and $\mathbf{v} = y_1\mathbf{s}_1 + \cdots + y_n\mathbf{s}_n$. Then for any scalar $c \in \mathbb{R}$, we can write $c\mathbf{u} + \mathbf{v} = (cx_1 + y_1)\mathbf{s}_1 + \cdots + (cx_n + y_n)\mathbf{s}_n$. (Here, we have used several vector space properties enjoyed by $\mathbb{V}$.) Therefore $c\mathbf{u} + \mathbf{v} \in \mathrm{span}(\mathcal{S})$ as well; this establishes SS1 and SS2. To establish SS3, observe that the "empty" linear combination of vectors from $\mathcal{S}$ results in $\mathbf{0}$, and hence $\mathbf{0} \in \mathrm{span}(\mathcal{S})$. ∎

   (Can a subspace of $\mathbb{V}$ contain $\mathcal{S}$ but not $\mathrm{span}(\mathcal{S})$?)

10. If A is an $m \times n$ matrix, then $\mathsf{CS}(A)$ is a subspace of $\mathbb{R}^m$.

    This is really just a special case of the previous example, since $\mathsf{CS}(A)$ is the span of the columns of A, and each column of A is an m-vector.

Because a subspace is, in its own right, a vector space, the theorems we have enumerated above for general vector spaces are also applicable to subspaces of general vector spaces.

3 Bases

We say a set of vectors $\mathcal{B}$ from a vector space $\mathbb{V}$ is a *basis* for $\mathbb{V}$ if:

1. $\mathcal{B}$ is linearly independent, and
2. $\text{span}(\mathcal{B}) = \mathbb{V}$.

(The plural form of "basis" is "bases".) Thus, a basis for a vector space $\mathbb{V}$ is a minimal collection of vectors by which we can construct all of $\mathbb{V}$ simply via linear combination. A basis is *finite* if its size (a.k.a. cardinality) is finite.

Example. A basis for $\mathbb{R}^4$ is $\mathcal{E}_4 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$, where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Observe that $\mathbf{e}_i$ cannot be obtained as a linear combination of the other $\mathbf{e}_j$'s, since the other $\mathbf{e}_j$'s (for $j \neq i$) have 0's in the ith component. So $\mathcal{E}_4$ is linearly independent. Any 4-vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ can be written as $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + x_4\mathbf{e}_4$. So every 4-vector is in $\text{span}(\mathcal{E}_4)$. (This naturally generalizes to $\mathcal{E}_n = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, for any n.) ∎

The basis from the previous example is called the *standard basis* (a.k.a. *coordinate basis*). And it is easy to generalize the example to $\mathbb{R}^n$ for any n (and with it, the definition of the standard basis for $\mathbb{R}^n$; the symbol "$\mathbf{e}_i$" is commonly used to denote the ith standard basis vector, regardless of n). But the standard basis is not the only basis for $\mathbb{R}^n$, as the next example shows.

Example. Another basis for $\mathbb{R}^4$ is $\mathcal{F}_4 = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$, where

$$\mathbf{f}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{f}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{f}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{f}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

For instance, the vector $\mathbf{x} = (5, 4, 8, -1)$ can be expressed as a linear combination of $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ in the following way:

$$\mathbf{x} \;=\; \mathbf{f}_1 - 4\,\mathbf{f}_2 + 9\,\mathbf{f}_3 - \mathbf{f}_4.$$

In general, we define $\mathcal{F}_n = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ for $\mathbb{R}^n$ for any n. We verify that $\mathcal{F}_n$ is a basis for $\mathbb{R}^n$. (Caution: Unlike the standard basis, $\mathcal{F}_n$ is not a commonly-used basis; we only define it here for this example.)

Note that $\mathbf{e}_1 = \mathbf{f}_1$, and $\mathbf{e}_i = \mathbf{f}_i - \mathbf{f}_{i-1}$ for every $i > 1$. So every standard basis vector can be obtained as a linear combinations of vectors in $\mathcal{F}_n$. This implies that $\mathrm{span}(\mathcal{F}_n) = \mathbb{R}^n$.

Using mathematical induction, we prove that $\{\mathbf{f}_1, \ldots, \mathbf{f}_i\}$ is linearly independent for every i. The singleton set $\{\mathbf{f}_1\}$ is linearly independent since $\mathbf{f}_1$ is not the zero vector. This covers the base case. And for any $i > 1$, the vector $\mathbf{f}_i$ is not in $\mathrm{span}(\{\mathbf{f}_1, \ldots, \mathbf{f}_{i-1}\})$ on account of the 1 in the ith component in $\mathbf{f}_i$. So by the inductive hypothesis that $\{\mathbf{f}_1, \ldots, \mathbf{f}_{i-1}\}$ is linearly independent, and the Growth Theorem, the set $\{\mathbf{f}_1, \ldots, \mathbf{f}_i\}$ is linearly independent. ∎

Even though there may be many different bases for a vector space $\mathbb{V}$, they all have exactly the same number of vectors.

Theorem 5 (Size-of-Basis Theorem)**.** *Suppose a vector space $\mathbb{V}$ has a finite basis. Then all bases for $\mathbb{V}$ have exactly the same number of vectors.*

Proof. Let $\mathcal{E}$ denote a finite basis for $\mathbb{V}$, and consider any other basis $\mathcal{B}$ for $\mathbb{V}$. (Here, we denote cardinality of a set S by $|S|$.)

Suppose for sake of contradiction that $|\mathcal{B}| > |\mathcal{E}|$. Let $\mathcal{B}'$ be a subset of $\mathcal{B}$ containing exactly $|\mathcal{E}| + 1$ vectors. Since $\mathcal{B}$ is linearly independent, so is $\mathcal{B}'$ by the Removal Theorem (Theorem 3). And since $\mathcal{B}'$ is a set of vectors from $\mathbb{V} = \mathrm{span}(\mathcal{E})$, the Exchange Theorem implies that $|\mathcal{B}'| \leq |\mathcal{E}|$. But this is impossible because $|\mathcal{B}'| = |\mathcal{E}| + 1$: So it must be that $\mathcal{B}$ is finite and $|\mathcal{B}| \leq |\mathcal{E}|$.

Now we reverse the roles of $\mathcal{E}$ and $\mathcal{B}$. Since $\mathcal{B}$ is a basis for $\mathbb{V}$, we have $\mathrm{span}(\mathcal{B}) = \mathbb{V}$. And since $\mathcal{E}$ is a linearly independent set of vectors from $\mathbb{V} = \mathrm{span}(\mathcal{B})$, the Exchange Theorem (Theorem 4) implies that $|\mathcal{E}| \leq |\mathcal{B}|$.

Hence, we conclude that $|\mathcal{B}| = |\mathcal{E}|$. □

The size of a finite basis for $\mathbb{V}$, which is the same for every such basis, is called the *dimension* of $\mathbb{V}$, written $\dim(\mathbb{V})$.[4] For example, $\dim(\mathbb{R}^n) = n$

[4] In this class, infinite dimensional vector spaces will only come up occasionally in examples. The main focus of the class will be on finite dimensional vector spaces, and $\mathbb{R}^n$ in particular.

because the standard basis has n vectors, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$.

Theorem 6 (Basis Sufficiency Theorem)**.** *Let $\mathbb{V}$ be an n-dimensional vector space. Then the following statements are equivalent:*

1. *$\mathcal{B}$ is a basis for $\mathbb{V}$.*
2. *$\mathcal{B}$ is a set of n linearly independent vectors from $\mathbb{V}$.*
3. *$\mathcal{B}$ is a set of n vectors from $\mathbb{V}$ with $\mathrm{span}(\mathcal{B}) = \mathbb{V}$.*

Proof. The first statement implies the other two by the definition of basis and the Size-of-Basis Theorem (Theorem 5).

We now show that the second statement implies the first statement. Let $\mathcal{B}$ be a set of n linearly independent vectors from $\mathbb{V}$, and let $\mathcal{E}$ denote a basis for $\mathbb{V}$ of size n. By the Exchange Theorem (Theorem 4), there is a subset $\mathcal{F}$ of $|\mathcal{E}| - |\mathcal{B}|$ vectors from $\mathcal{E}$ such that $\mathbb{V} = \mathrm{span}(\mathcal{B} \cup \mathcal{F})$. Since $|\mathcal{E}| = |\mathcal{B}|$, we can take $\mathcal{F} = \emptyset$, so $\mathbb{V} = \mathrm{span}(\mathcal{B})$. So $\mathcal{B}$ is a basis for $\mathbb{V}$.

Finally, we show that the third statement implies the first statement. Let $\mathcal{B}$ be a set of n vectors from $\mathbb{V}$ with $\mathrm{span}(\mathcal{B}) = \mathbb{V}$, and suppose for sake of contradiction that $\mathcal{B}$ is not linearly independent. Then by (repeated application of) the Removal Theorem (Theorem 3), there is a strict subset $\mathcal{B}' \subsetneq \mathcal{B}$ of $\mathcal{B}$ that is linearly independent and $\mathrm{span}(\mathcal{B}') = \mathbb{V}$. So $\mathcal{B}'$ is a basis for $\mathbb{V}$ with fewer than n vectors. By the Size-of-Basis Theorem (Theorem 5), this is impossible. So we conclude that $\mathcal{B}$ is linearly independent, and since $\mathrm{span}(\mathcal{B}) = \mathbb{V}$, $\mathcal{B}$ is a basis for $\mathbb{V}$. □

If a linearly independent set of vectors from a finite-dimensional vector space is not a basis, then it can be "completed" to become a basis—i.e., augmented with additional vectors so that the resulting set is a basis. This is the a direct consequence of the Exchange Theorem (Theorem 4) and the definition of basis.

Theorem 7 (Basis Completion Theorem)**.** *Let $\mathcal{W}$ be a linearly independent set of k vectors from an n-dimensional vector space $\mathbb{V}$. There exists a subset $\mathcal{F}$ of $n - k$ vectors such that $\mathcal{W} \cup \mathcal{F}$ is a basis for $\mathbb{V}$.*

More examples.

1. A basis for $\mathsf{P}_2(\mathbb{R})$ is
$$\{1, t, t^2\},$$
so the dimension of $\mathsf{P}_2(\mathbb{R})$ is 3.

2. A basis for $\mathsf{M}_{3\times 2}(\mathbb{R})$ is
$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$
so the dimension of $\mathsf{M}_{3\times 2}(\mathbb{R})$ is 6.

3. A basis for $\{\mathbf{0}\}$ is the empty set, so the dimension of $\{\mathbf{0}\}$ is 0.

4. A basis for $\mathsf{P}(\mathbb{R})$ is
$$\{1, t, t^2, t^3, \ldots\},$$
an infinite collection of vectors. The dimension of $\mathsf{P}(\mathbb{R})$ is infinite.

5. It turns out $\mathsf{C}(\mathbb{R}, \mathbb{R})$ also has a basis, but not one we can explicitly describe.[5]

Finally, a important use of a basis is to (fully) characterize a linear transformation between vector spaces by its action on a basis for the input space.

Theorem 8 (Unique Linear Transformation Theorem)**.** *Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces, and suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{V}$. For any $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbb{W}$, there is exactly one linear transformation $T \colon \mathbb{V} \to \mathbb{W}$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for all $i \in \{1, \ldots, n\}$.*

Thus, if linear transformations $T \colon \mathbb{V} \to \mathbb{W}$ and $U \colon \mathbb{V} \to \mathbb{W}$ agree on their behavior on a basis for $\mathbb{V}$, then they must agree everywhere—i.e., they must be the same linear transformation.

If $\mathbb{W}$ is a subspace of the vector space $\mathbb{V}$, and if $\mathcal{B}$ is a finite basis for $\mathbb{W}$, then by the Size-of-Basis Theorem (Theorem 5), every basis for $\mathbb{W}$ has $|\mathcal{B}|$ vectors—i.e., the dimension of $\mathbb{W}$ is $|\mathcal{B}|$.

[5]The existence of a basis for an arbitrary vector space is guaranteed by the Axiom of Choice.

Examples.

1. Suppose $\mathbb{V} = \mathbb{R}^3$ and $\mathbb{W} = \{(x, y, 0) : (x, y) \in \mathbb{R}^2\}$. Then, the set
$$\{(1, 0, 0), (0, 1, 0)\}$$
is a basis for $\mathbb{W}$, and $\dim(\mathbb{W}) = 2$.

2. Suppose $\mathbb{V} = \mathbb{R}^3$ and $\mathbb{W} = \{(x, y, x) : (x, y) \in \mathbb{R}^2\}$. Then, the set
$$\{(1, 0, 1), (0, 1, 0)\}$$
is a basis for $\mathbb{W}$, and $\dim(\mathbb{W}) = 2$.

3. Suppose $\mathbb{V} = \mathbb{R}^3$ and $\mathbb{W} = \{(x, 2x, 3x) : x \in \mathbb{R}\}$. Then, the set
$$\{(1, 2, 3)\}$$
is a basis for $\mathbb{W}$, and $\dim(\mathbb{W}) = 1$.

4. Recall that the even polynomials $\mathsf{P}^{\mathsf{even}}(\mathbb{R})$ is a subspace of $\mathsf{P}(\mathbb{R})$. A basis for $\mathsf{P}^{\mathsf{even}}(\mathbb{R})$ is
$$\{1, t^2, t^4, t^6, \ldots\}.$$
Like $\mathsf{P}(\mathbb{R})$, the even polynomials also has infinite dimension.

5. For any matrix A, a basis for $\mathsf{CS}(A)$ is given as the columns of the matrix C from the CR factorization of A. (Recall that these columns of C are also columns of A.) ∎

Theorem 9 (Subspace Dimension Theorem)**.** *If $\mathbb{W}$ is a subspace of a vector space $\mathbb{V}$ with $\dim(\mathbb{V}) < \infty$, then $\dim(\mathbb{W}) \leq \dim(\mathbb{V})$. Moreover, if $\dim(\mathbb{W}) = \dim(\mathbb{V})$, then $\mathbb{W} = \mathbb{V}$.*

Proof. Let $d = \dim(\mathbb{V})$. Starting with the empty set, greedily construct a maximal linearly independent set $\mathcal{B}$ of vectors from $\mathbb{W}$ (which are also from $\mathbb{V}$). By the Size-of-Basis Theorem (Theorem 5), no linearly independent set from $\mathbb{V}$ can contain more than d vectors, so $|\mathcal{B}| \leq d$. So this process terminates with a linearly independent set $\mathcal{B} \subset \mathbb{W}$, with $0 \leq |\mathcal{B}| \leq d$, such that for any other vector $\mathbf{w} \in \mathbb{W} \setminus \mathcal{B}$, the set $\mathcal{B} \cup \{\mathbf{w}\}$ is linearly dependent. By the Growth Theorem (Theorem 2), this means that $\mathbf{w} \in \mathrm{span}(\mathcal{B})$ for all

$\mathbf{w} \in \mathbb{W} \setminus \mathcal{B}$, which, in turn, implies $\mathrm{span}(\mathcal{B}) = \mathbb{W}$. So $\mathcal{B}$ is a basis for $\mathbb{W}$, and $|\mathcal{B}| = \dim(\mathbb{W}) \leq d$.

If $|\mathcal{B}| = \dim(\mathbb{W}) = d$, then the Basis Sufficiency Theorem (Theorem 6) implies that $\mathcal{B}$ is a basis for $\mathbb{V}$, and hence $\mathbb{W} = \mathrm{span}(\mathcal{B}) = \mathbb{V}$. □

4 Coordinate representations

As we mentioned, a basis for a vector space $\mathbb{V}$ provides essential building blocks for constructing all vectors in $\mathbb{V}$ by linear combinations. But because a vector space $\mathbb{V}$ can have many bases, a vector can be constructed in many different ways.

When we write an n-vector $\mathbf{v}$ as $\mathbf{v} = (v_1, \ldots, v_n)$, we implicitly interpret the coordinates $(v_1, \ldots, v_n)$ as specifying a linear combination of vectors from the *standard ordered basis* $\mathcal{E}_n = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$:

$$\mathbf{v} \;=\; v_1\,\mathbf{e}_1 + \cdots + v_n\,\mathbf{e}_n.$$

Here, the ordering of the basis vectors in $\mathcal{E}_n$ is important, since we really need $\mathbf{e}_i$ to be the ith basis vector to match up with the coordinate v_i. We use the round (non-curly) brackets to emphasize this.[6] An *ordered basis* for a vector space $\mathbb{V}$ is a basis for $\mathbb{V}$ in which we specify the ordering of the basis vectors. So $(v_1, \ldots, v_n)$ is the *coordinate representation* (a.k.a. *coordinates*) of $\mathbf{v}$ with respect to the ordered basis $\mathcal{E}_n$.

Suppose $\mathcal{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ is an ordered basis for $\mathbb{V} = \mathbb{R}^n$, where the coordinates for $\mathbf{f}_j$ with respect to $\mathcal{E}_n$ is $\mathbf{f}_j = (f_{1,j}, \ldots, f_{n,j})$:

$$\mathbf{f}_j \;=\; f_{1,j}\,\mathbf{e}_1 + \cdots + f_{n,j}\,\mathbf{e}_n.$$

How can we find the coordinates of a given vector $\mathbf{v} \in \mathbb{R}^n$ with respect to $\mathcal{F}$, where $\mathbf{v}$ is given as coorindates $\mathbf{v} = (v_1, \ldots, v_n)$ with respect to $\mathcal{E}_n$? That is, how do we find $(x_1, \ldots, x_n)$ such that $\mathbf{v} = x_1\mathbf{f}_1 + \cdots + x_n\mathbf{f}_n$?

Answer: Solve (e.g., using Elimination) the system of linear equations where the jth column of the coefficient matrix $F \in \mathsf{M}_{n\times n}(\mathbb{R})$ is $(f_{1,j}, \ldots, f_{n,j})$, and the right-hand side vector is $(v_1, \ldots, v_n)$. Since F has n linearly independent columns, it is invertible by the Invertibility Theorem. The solution can be expressed algebraically as $F^{-1}\mathbf{v}$.

[6] The use of curly braces in $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ usually implies that the set is unordered. Round brackets (a.k.a. parentheses) are usually used to specify tuples or lists, which are always ordered.

Example. Recall that

$$\mathcal{F}_4 = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

is an ordered basis for the vector space $\mathbb{R}^4$. To obtain the coordinates of $\mathbf{v} = (5, 4, 8, -1)$ with respect to $\mathcal{F}_4$, we solve

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 8 \\ -1 \end{bmatrix}.$$

The unique solution is $(x_1, x_2, x_3, x_4) = (1, -4, 9, -1)$. ∎

We generalize the concept of coordinates in $\mathbb{R}^n$ to arbitrary n-dimensional vector spaces $\mathbb{V}$ with respect to an ordered basis $\mathcal{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ for $\mathbb{V}$ as follows. The *coordinate representation* (a.k.a. *coordinates*) of a vector $\mathbf{v} \in \mathbb{V}$ with respect to $\mathcal{F}$ is the (unique) vector of coefficients $(c_1, \ldots, c_n) \in \mathbb{R}^n$ such that $\mathbf{v} = c_1\mathbf{f}_1 + \cdots + c_n\mathbf{f}_n$. We use the notation $[\mathbf{v}]_{\mathcal{F}}$ to denote this coefficient vector. The transformation $[\,\cdot\,]_{\mathcal{F}} \colon \mathbb{V} \to \mathbb{R}^n$ is called the *standard representation map* for $\mathbb{V}$ with respect to $\mathcal{F}$.

An important property of the standard representation map is linearity.[7]

Theorem 10. *Let $\mathcal{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ be an ordered basis for a vector space $\mathbb{V}$. Then $[\,\cdot\,]_{\mathcal{F}} \colon \mathbb{V} \to \mathbb{R}^n$ is linear.*

Proof. Pick any $\mathbf{u}, \mathbf{v} \in \mathbb{V}$ and scalar c. We need to show that $[c\,\mathbf{u} + \mathbf{v}]_{\mathcal{F}} = c\,[\mathbf{u}]_{\mathcal{F}} + [\mathbf{v}]_{\mathcal{F}}$. Let $(x_1, \ldots, x_n) = [\mathbf{u}]_{\mathcal{F}}$, $(y_1, \ldots, y_n) = [\mathbf{v}]_{\mathcal{F}}$, and $(z_1, \ldots, z_n) = c\,[\mathbf{u}]_{\mathcal{F}} + [\mathbf{v}]_{\mathcal{F}}$. We claim that $(z_1, \ldots, z_n) = [c\,\mathbf{u} + \mathbf{v}]_{\mathcal{F}}$:

$$\begin{aligned} z_1\,\mathbf{f}_1 + \cdots + z_n\,\mathbf{f}_n &= (c\,x_1 + y_1)\,\mathbf{f}_1 + \cdots + (c\,x_n + y_n)\,\mathbf{f}_n \\ &= c\,(x_1\,\mathbf{f}_1 + \cdots + x_n\,\mathbf{f}_n) + (y_1\,\mathbf{f}_1 + \cdots + y_n\,\mathbf{f}_n) \\ &= c\,\mathbf{u} + \mathbf{v}. \end{aligned}$$

Hence $[c\,\mathbf{u} + \mathbf{v}]_{\mathcal{F}} = c\,[\mathbf{u}]_{\mathcal{F}} + [\mathbf{v}]_{\mathcal{F}}$, so $[\,\cdot\,]_{\mathcal{F}}$ is linear. □

[7] Going in the opposite direction (from coordinates to vectors in $\mathbb{V}$) is also linear. In fact, for any vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{V}$, the linear transformation $T \colon \mathbb{R}^n \to \mathbb{V}$ defined by $T(x_1, \ldots, x_n) = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ is linear. This is a generalization of matrix-vector multiplication: $T(\mathbf{x}) = [\mathbf{v}_1, \ldots, \mathbf{v}_n]\mathbf{x}$.

The standard representation map with respect to an ordered basis is useful for computation; it allows us to work with coordinate vectors, just as we have been doing with vectors in Cartesian spaces.

Example. Let $\mathbb{V} = \mathsf{P}_2(\mathbb{R})$, and consider the polynomial $f(t) = 3t^2+2t+1 \in \mathbb{V}$. One ordered basis for $\mathbb{V}$ is $\mathcal{F} = (1, t, t^2)$, and another is $\mathcal{H} = (1, t, t^2 - 1)$. We have $[f(t)]_{\mathcal{F}} = (1, 2, 3)$. To get $[f(t)]_{\mathcal{H}}$, we first determine

$$[1]_{\mathcal{H}} \;=\; (1, 0, 0), \qquad [t]_{\mathcal{H}} \;=\; (0, 1, 0), \qquad [t^2]_{\mathcal{H}} \;=\; (1, 0, 1),$$

and then we compute the matrix-vector product

$$[f(t)]_{\mathcal{H}} \;=\; \begin{bmatrix} [1]_{\mathcal{H}} & [t]_{\mathcal{H}} & [t^2]_{\mathcal{H}} \end{bmatrix} \begin{bmatrix} \\ [f(t)]_{\mathcal{F}} \\ \\ \end{bmatrix} \;=\; \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \;=\; \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}. \quad \blacksquare$$

The matrix in the previous example is an example of a *change-of-coordinates matrix*. It converts representations with respect to one basis $\mathcal{F}$ to representations with respect to another basis $\mathcal{H}$. In general, if $\mathcal{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_n)$ and $\mathcal{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_n)$ are ordered bases for a vector space $\mathbb{V}$, then the change-of-coordinates matrix from $\mathcal{F}$ to $\mathcal{H}$ is the $n \times n$ matrix

$$\begin{bmatrix} \\ [\mathbf{f}_1]_{\mathcal{H}} & \cdots & [\mathbf{f}_n]_{\mathcal{H}} \\ \\ \end{bmatrix}.$$

This matrix is invertible, and its inverse is the $n \times n$ matrix $[[\mathbf{h}_1]_{\mathcal{F}}, \ldots, [\mathbf{h}_n]_{\mathcal{F}}]$.

A Unique linear transformation theorem

Proof of Theorem 8. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, and define $T \colon \mathbb{V} \to \mathbb{W}$ as

$$T(\mathbf{x}) \;=\; [\mathbf{x}]_{\mathcal{B},1}\mathbf{w}_1 + \cdots + [\mathbf{x}]_{\mathcal{B},n}\mathbf{w}_n,$$

where $[\mathbf{x}]_{\mathcal{B},i}$ is the ith component of the n-vector $[\mathbf{x}]_{\mathcal{B}}$. This is a composition of linear transformations and hence is linear. It satisfies $T(\mathbf{v}_i) = \mathbf{w}_i$ since $[\mathbf{v}_i]_{\mathcal{B}} = \mathbf{e}_i$ for all $i \in \{1, \ldots, n\}$.

To show that T is the only linear transformation witih this property, we suppose there is another linear transformation $U \colon \mathbb{V} \to \mathbb{W}$ with $U(\mathbf{v}_i) = \mathbf{w}_i$ for all $i \in \{1, \ldots, n\}$. We show that $U(\mathbf{x}) = T(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{V}$:

$$\begin{aligned} U(\mathbf{x}) &= U([\mathbf{x}]_{\mathcal{B},1}\,\mathbf{v}_1 + \cdots + [\mathbf{x}]_{\mathcal{B},n}\,\mathbf{v}_n) \\ &= [\mathbf{x}]_{\mathcal{B},1}\,U(\mathbf{v}_1) + \cdots + [\mathbf{x}]_{\mathcal{B},n}\,U(\mathbf{v}_n) \quad \text{(since } U \text{ is linear)} \\ &= [\mathbf{x}]_{\mathcal{B},1}\,\mathbf{w}_1 + \cdots + [\mathbf{x}]_{\mathcal{B},n}\,\mathbf{w}_n \\ &= T(\mathbf{x}). \end{aligned}$$

□

B Extra vector space properties

Theorem 11. *Let $\mathbb{V}$ be a vector space. The following are true for all vectors $\mathbf{v} \in \mathbb{V}$ and all scalars $c \in \mathbb{R}$.*

(a) $\mathsf{scale}(0, \mathbf{v}) = \mathbf{0}_{\mathbb{V}}$.

(b) $\mathsf{scale}(c, \mathbf{0}_{\mathbb{V}}) = \mathbf{0}_{\mathbb{V}}$.

(c) $\mathsf{scale}(-1, \mathbf{v})$ *is the unique additive inverse of* $\mathbf{v}$.

(d) If $\mathsf{scale}(c, \mathbf{v}) = \mathbf{0}_{\mathbb{V}}$ *and* $c \neq 0$, *then* $\mathbf{v} = \mathbf{0}_{\mathbb{V}}$.

Proof.

(a) Let $-\mathbf{v}$ denote an additive inverse of $\mathbf{v}$ (as guaranteed to exist by VS4).

Then

$$\begin{aligned}
\mathsf{scale}(0, \mathbf{v}) &= \mathsf{add}(\mathbf{0}_{\mathbb{V}}, \mathsf{scale}(0, \mathbf{v})) && \text{(by VS3)} \\
&= \mathsf{add}(\mathsf{scale}(0, \mathbf{v}), \mathbf{0}_{\mathbb{V}}) && \text{(by VS1)} \\
&= \mathsf{add}(\mathsf{scale}(0, \mathbf{v}), \mathsf{add}(\mathbf{v}, -\mathbf{v})) && \text{(by VS4)} \\
&= \mathsf{add}(\mathsf{add}(\mathsf{scale}(0, \mathbf{v}), \mathbf{v}), -\mathbf{v}) && \text{(by VS2)} \\
&= \mathsf{add}(\mathsf{add}(\mathsf{scale}(0, \mathbf{v}), \mathsf{scale}(1, \mathbf{v})), -\mathbf{v}) && \text{(by VS5)} \\
&= \mathsf{add}(\mathsf{scale}(0 + 1, \mathbf{v}), -\mathbf{v}) && \text{(by VS8)} \\
&= \mathsf{add}(\mathsf{scale}(1, \mathbf{v}), -\mathbf{v}) && \text{(by scalar arithmetic)} \\
&= \mathsf{add}(\mathbf{v}, -\mathbf{v}) && \text{(by VS5)} \\
&= \mathbf{0}_{\mathbb{V}} && \text{(by VS4).}
\end{aligned}$$

(b)

$$\begin{aligned}
\mathsf{scale}(c, \mathbf{0}_{\mathbb{V}}) &= \mathsf{scale}(c, \mathsf{scale}(0, \mathbf{0}_{\mathbb{V}})) && \text{(by Part (a))} \\
&= \mathsf{scale}(c \cdot 0, \mathbf{0}_{\mathbb{V}}) && \text{(by VS6)} \\
&= \mathsf{scale}(0, \mathbf{0}_{\mathbb{V}}) && \text{(by scalar arithmetic)} \\
&= \mathbf{0}_{\mathbb{V}} && \text{(by Part (a)).}
\end{aligned}$$

(c) Our goal is to show that $\mathsf{scale}(-1, \mathbf{v})$ is the (unique) additive inverse of $\mathbf{v}$. Let $-\mathbf{v}$ denote an additive inverse of $\mathbf{v}$ (as guaranteed to exist by VS4). Then

$$\begin{aligned}
\mathsf{scale}(-1, \mathbf{v}) &= \mathsf{add}(\mathbf{0}_{\mathbb{V}}, \mathsf{scale}(-1, \mathbf{v})) && \text{(by VS3)} \\
&= \mathsf{add}(\mathsf{scale}(-1, \mathbf{v}), \mathbf{0}_{\mathbb{V}}) && \text{(by VS1)} \\
&= \mathsf{add}(\mathsf{scale}(-1, \mathbf{v}), \mathsf{add}(\mathbf{v}, -\mathbf{v})) && \text{(by VS4)} \\
&= \mathsf{add}(\mathsf{add}(\mathsf{scale}(-1, \mathbf{v}), \mathbf{v}), -\mathbf{v}) && \text{(by VS2)} \\
&= \mathsf{add}(\mathsf{add}(\mathsf{scale}(-1, \mathbf{v}), \mathsf{scale}(1, \mathbf{v})), -\mathbf{v}) && \text{(by VS5)} \\
&= \mathsf{add}(\mathsf{scale}(-1 + 1, \mathbf{v}), -\mathbf{v}) && \text{(by VS8)} \\
&= \mathsf{add}(\mathsf{scale}(0, \mathbf{v}), -\mathbf{v}) && \text{(by scalar arithmetic)} \\
&= \mathsf{add}(\mathbf{0}_{\mathbb{V}}, -\mathbf{v}) && \text{(by Part (a))} \\
&= -\mathbf{v} && \text{(by VS3).}
\end{aligned}$$

So every additive inverse of $\mathbf{v}$ must be $\mathsf{scale}(-1, \mathbf{v})$, and hence $\mathsf{scale}(-1, \mathbf{v})$ is the only additive inverse of $\mathbf{v}$.

(d) Assume $\mathsf{scale}(c, \mathbf{v}) = \mathbf{0}_{\mathbb{V}}$ and $c \neq 0$. The latter assumption implies that there is a scalar c^{-1} such that $c^{-1}c = 1$. Then

$$\begin{aligned}
\mathbf{v} &= \mathsf{scale}(1, \mathbf{v}) && \text{(by VS5)} \\
&= \mathsf{scale}(c^{-1}c, \mathbf{v}) && \text{(by choice of } c^{-1}) \\
&= \mathsf{scale}(c^{-1}, \mathsf{scale}(c, \mathbf{v})) && \text{(by VS6)} \\
&= \mathsf{scale}(c^{-1}, \mathbf{0}_{\mathbb{V}}) && \text{(by assumption)} \\
&= \mathbf{0}_{\mathbb{V}} && \text{(by Part (b)).}
\end{aligned}$$

□