Cambridge University Press & Assessment
978-1-108-83710-1 — Matrix Mathematics
2nd Edition
Excerpt
More Information

1 Vector Spaces

Many types of mathematical objects can be added and scaled: vectors in the plane, real-valued functions on a given real interval, polynomials, and real or complex matrices. Through long experience with these and other examples, mathematicians have identified a short list of essential features (axioms) that define a consistent and inclusive mathematical framework known as a vector space.

The theory of vector spaces and linear transformations provides a conceptual framework and vocabulary for linear mathematical models of diverse phenomena. Even inherently nonlinear physical theories may be well approximated for a broad range of applications by linear theories, whose natural setting is in real or complex vector spaces.

Examples of vector spaces include the two-dimensional real plane (the setting for plane analytic geometry and two-dimensional Newtonian mechanics) and three-dimensional real Euclidean space (the setting for solid analytic geometry, classical electromagnetism, and analytical dynamics). Other kinds of vector spaces abound in science and engineering. For example, standard mathematical models in quantum mechanics, electrical circuits, and signal processing use complex vector spaces. Many scientific theories exploit the formalism of vector spaces, which supplies powerful mathematical tools that are based only on the axioms for a vector space and their logical consequences, not on the details of a particular application.

In this chapter, we provide formal definitions of real and complex vector spaces, and many examples. Among the important concepts introduced are linear combinations, span, linear independence, and linear dependence.

1.1 What Is a Vector Space?

A vector space comprises four things that work together in harmony:

(a) A field $\mathbb{F}$ of *scalars*, which in this book is either $\mathbb{C}$ (complex numbers) or $\mathbb{R}$ (real numbers).

(b) A set $\mathcal{V}$ of objects called *vectors*.

(c) An operation of *vector addition* that takes any pair of vectors $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ and assigns to them a vector in $\mathcal{V}$ denoted by $\mathbf{u} + \mathbf{v}$ (their *sum*).

(d) An operation of *scalar multiplication* that takes any scalar $c \in \mathbb{F}$ and any vector $\mathbf{u} \in \mathcal{V}$ and assigns to them a vector in $\mathcal{V}$ denoted by $c\mathbf{u}$.

Definition 1.1.1 Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $\mathcal{V}$ is a *vector space over the field* $\mathbb{F}$ (alternatively, $\mathcal{V}$ is an $\mathbb{F}$-*vector space*) if the scalars $\mathbb{F}$, the vectors $\mathcal{V}$, and the operations of vector addition and scalar multiplication satisfy the following axioms:

(1) There is a unique element $\mathbf{0} \in \mathcal{V}$ that is the additive identity element for vector addition, that is, $\mathbf{0} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathcal{V}$. The vector $\mathbf{0}$ is the *zero vector*.

(2) Vector addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

(3) Vector addition is associative: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$.

(4) Additive inverses exist and are unique: for each $\mathbf{u} \in \mathcal{V}$, there is a unique $-\mathbf{u} \in \mathcal{V}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

(5) The number 1 is the identity element for scalar multiplication: $1\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathcal{V}$.

(6) Multiplication in $\mathbb{F}$ and scalar multiplication are compatible: $a(b\mathbf{u}) = (ab)\mathbf{u}$ for all $a, b \in \mathbb{F}$ and all $\mathbf{u} \in \mathcal{V}$.

(7) Scalar multiplication distributes over vector addition: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for all $c \in \mathbb{F}$ and all $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

(8) Addition in $\mathbb{F}$ distributes over scalar multiplication: $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$ for all $a, b \in \mathbb{F}$ and all $\mathbf{u} \in \mathcal{V}$.

A vector space over $\mathbb{R}$ is a *real vector space*; a vector space over $\mathbb{C}$ is a *complex vector space*. To help distinguish vectors from scalars, we often denote vectors (elements of the set $\mathcal{V}$) by boldface lowercase letters such as $\mathbf{a}$, $\mathbf{b}$, $\mathbf{u}$, and $\mathbf{v}$. In particular, this distinguishes the scalar 0 from the vector $\mathbf{0}$.

We often need to derive a conclusion from the fact that a vector $c\mathbf{u}$ is the zero vector, so we should look carefully at how that can happen.

Theorem 1.1.2 *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space, let $c \in \mathbb{F}$, and let $\mathbf{u} \in \mathcal{V}$. The following statements are equivalent:*

(a) $c = 0$ *or* $\mathbf{u} = \mathbf{0}$.

(b) $c\mathbf{u} = \mathbf{0}$.

Proof (a) ⇒ (b) First suppose that $\mathbf{u} = \mathbf{0}$. Then

$$\begin{aligned}
c\mathbf{0} &= c\mathbf{0} + \mathbf{0} && \text{Axiom (1)}\\
&= c\mathbf{0} + \big(c\mathbf{0} + (-(c\mathbf{0}))\big) && \text{Axiom (4)}\\
&= (c\mathbf{0} + c\mathbf{0}) + (-(c\mathbf{0})) && \text{Axiom (3)}\\
&= c(\mathbf{0} + \mathbf{0}) + (-(c\mathbf{0})) && \text{Axiom (7)}\\
&= c\mathbf{0} + (-(c\mathbf{0})) && \text{Axiom (1)}\\
&= \mathbf{0} && \text{Axiom (4).}
\end{aligned}$$

In particular, observe that

$$c\mathbf{0} = \mathbf{0} \text{ for any } c \in \mathbb{F}. \tag{1.1.3}$$

Now let $c = 0$ and compute

$$\begin{aligned}
0\mathbf{u} &= 0\mathbf{u} + \mathbf{0} && \text{Axiom (1)}\\
&= 0\mathbf{u} + \big(0\mathbf{u} + (-(0\mathbf{u}))\big) && \text{Axiom (4)}\\
&= (0\mathbf{u} + 0\mathbf{u}) + (-(0\mathbf{u})) && \text{Axiom (3)}\\
&= (0 + 0)\mathbf{u} + (-(0\mathbf{u})) && \text{Axiom (8)}\\
&= 0\mathbf{u} + (-(0\mathbf{u})) && \\
&= \mathbf{0} && \text{Axiom (4).}
\end{aligned}$$

(b) ⇒ (a) Suppose that $c\mathbf{u} = \mathbf{0}$. If $c = 0$, we are done. If $c \neq 0$, then

$$\begin{aligned} \mathbf{u} &= 1\mathbf{u} && \text{Axiom (5)} \\ &= (c^{-1}c)\mathbf{u} \\ &= c^{-1}(c\mathbf{u}) && \text{Axiom (6)} \\ &= c^{-1}\mathbf{0} \\ &= \mathbf{0} && \text{by (1.1.3).} \end{aligned}$$ □

Corollary 1.1.4 *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space. Then $(-1)\mathbf{u} = -\mathbf{u}$ for every $\mathbf{u} \in \mathcal{V}$.*

Proof Let $\mathbf{u} \in \mathcal{V}$. We must show that $(-1)\mathbf{u} + \mathbf{u} = \mathbf{0}$. Use the vector-space Axioms (5) and (8), together with the preceding theorem, and compute

$$\begin{aligned} (-1)\mathbf{u} + \mathbf{u} &= (-1)\mathbf{u} + 1\mathbf{u} && \text{Axiom (5)} \\ &= (-1 + 1)\mathbf{u} && \text{Axiom (8)} \\ &= 0\mathbf{u} \\ &= \mathbf{0} && \text{Theorem 1.1.2.} \end{aligned}$$ □

Addition in a vector space is an operation on only two vectors. We define addition of three or more vectors via a sequence of two-vector additions. We can define $\mathbf{u} + \mathbf{v} + \mathbf{w}$ to be

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} \quad \text{or} \quad \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

because Axiom (4) (associativity) says that these expressions are equal. We can define $\mathbf{u}+\mathbf{v}+\mathbf{w} + \mathbf{x}$ in a similar fashion (insert suitable parentheses) to be

$$(\mathbf{u} + \mathbf{v}) + (\mathbf{w} + \mathbf{x}), \quad \mathbf{u} + (\mathbf{v} + (\mathbf{w} + \mathbf{x})), \quad \text{or} \quad (\mathbf{u} + (\mathbf{v} + \mathbf{w})) + \mathbf{x}$$

because we can prove that these expressions are equal. For example, two applications of Axiom (4) show that

$$(\mathbf{u} + (\mathbf{v} + \mathbf{w})) + \mathbf{x} = ((\mathbf{u} + \mathbf{v}) + \mathbf{w}) + \mathbf{x} = (\mathbf{u} + \mathbf{v}) + (\mathbf{w} + \mathbf{x}).$$

If $n \geq 3$, we define $\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n$ (a finite sum) via any sequence of two-vector additions obtained by insertion of suitable parentheses. It follows from Axiom (4) that the sum obtained does not depend on how the parentheses are inserted.

1.2 Examples of Vector Spaces

Axiom (1) ensures that every vector space contains a zero vector, so a vector space cannot be empty. However, the axioms for a vector space permit $\mathcal{V}$ to contain only the zero vector. Such a vector space is not interesting, and we often need to exclude it when formulating theorems.

Definition 1.2.1 Let $\mathcal{V}$ be an $\mathbb{F}$-vector space. If $\mathcal{V} = \{\mathbf{0}\}$, then $\mathcal{V}$ is a *zero vector space*; if $\mathcal{V} \neq \{\mathbf{0}\}$, then $\mathcal{V}$ is a *nonzero vector space*.

In each of the following examples, we describe the elements of the set $\mathcal{V}$ (the *vectors*), the zero vector, and the operations of scalar multiplication and vector addition. In this book, the field $\mathbb{F}$ is always either $\mathbb{C}$ or $\mathbb{R}$.

Example 1.2.2 Let $\mathcal{V} = \mathbb{F}^n$, the set of $n \times 1$ matrices (column vectors) with entries from $\mathbb{F}$. For typographical convenience, we often write $\mathbf{u} = [u_i]$ or $\mathbf{u} = [u_1\ u_2\ \ldots\ u_n]^\mathsf{T}$ instead of[1]

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{F}^n \quad \text{for } u_1, u_2, \ldots, u_n \in \mathbb{F}.$$

Vector addition of $\mathbf{u} = [u_i]$ and $\mathbf{v} = [v_i]$ is defined by $\mathbf{u}+\mathbf{v} = [u_i+v_i]$, and scalar multiplication by elements of $\mathbb{F}$ is defined by $c\mathbf{u} = [cu_i]$; we refer to these as *entrywise operations*. The zero vector in $\mathbb{F}^n$ is $\mathbf{0}_n = [0\ 0\ \ldots\ 0]^\mathsf{T}$. We often omit the subscript from a zero vector when its size can be inferred from context.

Example 1.2.3 Let $\mathcal{V} \in \mathsf{M}_{1\times n}(\mathbb{F})$, the set of $1 \times n$ matrices (row vectors) with entries from $\mathbb{F}$. Vector addition and scalar multiplication are defined entrywise, as in the preceding example. The zero vector is the row vector $\mathbf{0}^\mathsf{T}$.

Example 1.2.4 Let $\mathcal{V} = \mathsf{M}_{m\times n}(\mathbb{F})$, the set of $m \times n$ matrices with entries from $\mathbb{F}$. Vector addition and scalar multiplication are defined entrywise, as in the preceding two examples. The zero vector in $\mathsf{M}_{m\times n}(\mathbb{F})$ is the matrix $0_{m\times n} \in \mathsf{M}_{m\times n}(\mathbb{F})$, all entries of which are zero. We often omit the subscripts from a zero matrix if its size can be inferred from context.

Example 1.2.5 Let $\mathcal{V} = \mathcal{P}_n$, the set of polynomials of degree at most n with complex coefficients. The set of polynomials of degree at most n with real coefficients is denoted by $\mathcal{P}_n(\mathbb{R})$. Addition of polynomials is defined by adding the coefficients of corresponding monomials. For example, if $p(z) = iz^2 - 5$ and $q(z) = -7z^2 + 3z + 2$ in $\mathcal{P}_2$, then $(p+q)(z) = (i-7)z^2 + 3z - 3$. Scalar multiplication of a polynomial by a scalar c is defined by multiplying each coefficient by c. For example, $(4p)(z) = 4iz^2 - 20$. The zero vector in $\mathcal{P}_n$ is the zero polynomial; see Appendix B.1.

Example 1.2.6 Let $\mathcal{V} = \mathcal{P}$, the set of all polynomials with complex coefficients. The operations of vector addition and scalar multiplication are the same as in the preceding example, and the zero vector in $\mathcal{P}$ is again the zero polynomial.

Example 1.2.7 Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$. The *Cartesian product* $\mathcal{V} \times \mathcal{W}$ is the set of all ordered pairs $(\mathbf{v}, \mathbf{w})$ in which $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$. Vector addition of $(\mathbf{v}_1, \mathbf{w}_1)$ and $(\mathbf{v}_2, \mathbf{w}_2)$ is defined by $(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2)$, in which $\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{w}_1 + \mathbf{w}_2$ denote the results of vector addition operations in $\mathcal{V}$ and $\mathcal{W}$, respectively. Scalar multiplication by elements of $\mathbb{F}$ is defined by $c(\mathbf{v}, \mathbf{w}) = (c\mathbf{v}, c\mathbf{w})$, in which $c\mathbf{v}$ and $c\mathbf{w}$ denote the results of scalar multiplication operations in $\mathcal{V}$ and $\mathcal{W}$, respectively. The zero vector in $\mathcal{V} \times \mathcal{W}$ is $(\mathbf{0}, \mathbf{0})$, which employs the respective zero vectors in $\mathcal{V}$ and $\mathcal{W}$. The elements of the vector spaces $\mathcal{V}$ and $\mathcal{W}$ can come from different sets, but it is essential that both vector spaces be over the same field.

Example 1.2.8 Let $\mathcal{V} = C_\mathbb{F}[a, b]$, the set of continuous $\mathbb{F}$-valued functions on an interval $[a, b] \subset \mathbb{R}$ with $a < b$. If the field designator is absent, it is understood that $\mathbb{F} = \mathbb{C}$, that is, $C[0, 1]$ is $C_\mathbb{C}[0, 1]$. The operations of vector addition and scalar multiplication are defined

[1] If you encounter an unfamiliar symbol in this book, consult the Notation section for a cross-reference to a definition. The Notation section entry for the symbol A^T identifies it as the transpose of a matrix A and points to Appendix C.2.

pointwise. If $f, g \in C_{\mathbb{F}}[a,b]$, then $f+g$ is the $\mathbb{F}$-valued function on $[a,b]$ defined by $(f+g)(t) = f(t) + g(t)$ for each $t \in [a,b]$. If $c \in \mathbb{F}$, the $\mathbb{F}$-valued function cf is defined by $(cf)(t) = cf(t)$ for each $t \in [a,b]$. A theorem from calculus ensures that $f+g$ and cf are continuous if f and g are continuous, so sums and scalar multiples of elements of $C_{\mathbb{F}}[a,b]$ are in $C_{\mathbb{F}}[a,b]$. The zero vector in $C_{\mathbb{F}}[a,b]$ is the *zero function*, which takes the value zero at every point in $[a,b]$.

Example 1.2.9 Let $\mathcal{V}$ be the set of all infinite sequences $\mathbf{u} = (u_1, u_2, \ldots)$, in which each $u_i \in \mathbb{F}$ and $u_i \neq 0$ for only finitely many values of the index i. The operations of vector addition and scalar multiplication are defined entrywise. The zero vector in $\mathcal{V}$ is the *zero infinite sequence* $\mathbf{0} = (0, 0, \ldots)$. We say that $\mathcal{V}$ is the $\mathbb{F}$-vector space of *finitely nonzero sequences*.

1.3 Subspaces

Definition 1.3.1 A *subspace* of an $\mathbb{F}$-vector space $\mathcal{V}$ is a subset $\mathcal{U} \subseteq \mathcal{V}$ that is an $\mathbb{F}$-vector space with the same vector addition and scalar multiplication operations as in $\mathcal{V}$.

A subspace is nonempty; it is a vector space, so it contains a zero vector.

Example 1.3.2 If $\mathcal{V}$ is an $\mathbb{F}$-vector space, then $\{\mathbf{0}\}$ and $\mathcal{V}$ itself are subspaces of $\mathcal{V}$.

To show that a subset $\mathcal{U}$ of an $\mathbb{F}$-vector space $\mathcal{V}$ is a subspace, we do not need to verify the vector-space Axioms (2)–(3) and (5)–(8) because they are automatically satisfied; we say that $\mathcal{U}$ *inherits* these properties from $\mathcal{V}$. However, we must show the following:

(a) Sums and scalar multiples of elements of $\mathcal{U}$ are in $\mathcal{U}$ (that is, $\mathcal{U}$ is *closed under vector addition and scalar multiplication*).

(b) $\mathcal{U}$ contains the zero vector of $\mathcal{V}$.

(c) $\mathcal{U}$ contains an additive inverse for each of its elements.

The following theorem describes a streamlined way to verify these conditions.

Theorem 1.3.3 *Let $\mathcal{V}$ be an $\mathbb{F}$-vector space and let $\mathcal{U}$ be a nonempty subset of $\mathcal{V}$. Then $\mathcal{U}$ is a subspace of $\mathcal{V}$ if and only if $c\mathbf{u} + \mathbf{v} \in \mathcal{U}$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ and all $c \in \mathbb{F}$.*

Proof If $\mathcal{U}$ is a subspace, $\mathbf{u}, \mathbf{v} \in \mathcal{U}$, and $c \in \mathbb{F}$, then $c\mathbf{u} + \mathbf{v} \in \mathcal{U}$ because a subspace is closed under scalar multiplication and vector addition.

Conversely, suppose that $c\mathbf{u} + \mathbf{v} \in \mathcal{U}$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ and every $c \in \mathbb{F}$. We must verify the properties (a), (b), and (c) in the preceding list.

(a) We have $c\mathbf{u} = c\mathbf{u} + \mathbf{0} \in \mathcal{U}$ and $\mathbf{u} + \mathbf{v} = 1\mathbf{u} + \mathbf{v} \in \mathcal{U}$ for all $c \in \mathbb{F}$ and all $\mathbf{u}, \mathbf{v} \in \mathcal{U}$.

(b) Let $\mathbf{u} \in \mathcal{U}$. Corollary 1.1.4 ensures that $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$, so $\mathbf{0} = (-1)\mathbf{u} + \mathbf{u} \in \mathcal{U}$.

(c) Since $(-1)\mathbf{u} = (-1)\mathbf{u} + \mathbf{0}$, it follows that the additive inverse of $\mathbf{u}$ is in $\mathcal{U}$. ☐

The following examples use the criterion in the preceding theorem to verify that a certain subset of a vector space is a subspace.

Example 1.3.4 Let $A \in \mathsf{M}_{m\times n}(\mathbb{F})$. The *null space* of A is

$$\operatorname{null} A = \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{F}^n. \tag{1.3.5}$$

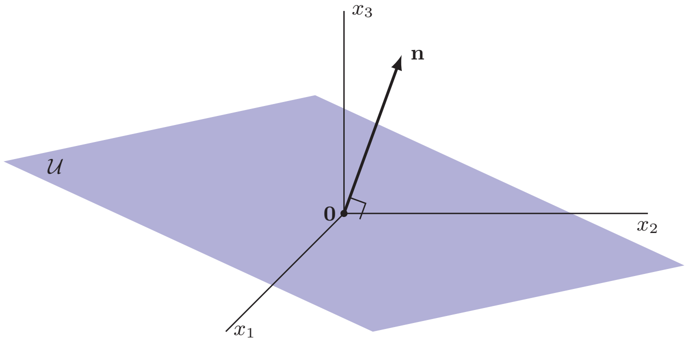

Figure 1.1 A plane $\mathcal{U}$ through the origin is a subspace of $\mathbb{R}^3$.

Since $A\mathbf{0}_n = \mathbf{0}_m$, the zero vector of $\mathbb{F}^n$ is in null A, which is therefore not empty. If $\mathbf{x}, \mathbf{y} \in$ null A, then $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. For any $c \in \mathbb{F}$, we have $A(c\mathbf{x} + \mathbf{y}) = cA\mathbf{x} + A\mathbf{y} = c\mathbf{0} + \mathbf{0} = \mathbf{0}$, so $c\mathbf{x} + \mathbf{y} \in$ null A. The preceding theorem ensures that null A is a subspace of $\mathbb{F}^n$.

Example 1.3.6 Let $\mathbf{n} = [a\ b\ c]^\mathsf{T} \in \mathbb{R}^3$ be nonzero and refer to Figure 1.1. The set $\mathcal{U} = \{\mathbf{x} = [x_1\ x_2\ x_3]^\mathsf{T} \in \mathbb{R}^3 : ax_1 + bx_2 + cx_3 = 0\}$ is the plane in $\mathbb{R}^3$ that contains the zero vector and has normal vector $\mathbf{n}$. Since $\mathbf{n}^\mathsf{T} \in \mathsf{M}_{1\times 3}(\mathbb{R})$ and $\mathbf{n}^\mathsf{T}\mathbf{x} = ax_1 + bx_2 + cx_3$, it follows that $\mathcal{U} = \operatorname{null} \mathbf{n}^\mathsf{T}$. The preceding example ensures that $\mathcal{U}$ is a subspace of $\mathbb{R}^3$.

Example 1.3.7 Let $A \in \mathsf{M}_{m\times n}(\mathbb{F})$. The *column space* of A is

$$\operatorname{col} A = \{A\mathbf{x} : \mathbf{x} \in \mathbb{F}^n\} \subseteq \mathbb{F}^m. \tag{1.3.8}$$

Since $A\mathbf{0}_n = \mathbf{0}_m$, the zero vector of $\mathbb{F}^m$ is in col A, which is therefore not empty. If $\mathbf{u}, \mathbf{v} \in$ col A, then there are $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ such that $\mathbf{u} = A\mathbf{x}$ and $\mathbf{v} = A\mathbf{y}$. For any $c \in \mathbb{F}$, we have $c\mathbf{u} + \mathbf{v} = cA\mathbf{x} + A\mathbf{y} = A(c\mathbf{x} + \mathbf{y})$, so $c\mathbf{u} + \mathbf{v} \in$ col A. Theorem 1.3.3 ensures that col A is a subspace of $\mathbb{F}^m$.

Example 1.3.9 Let $A \in \mathsf{M}_{m\times n}(\mathbb{F})$. The *row space* of A is

$$\operatorname{row} A = \{\mathbf{x}^\mathsf{T} A : \mathbf{x} \in \mathbb{F}^m\} \subseteq \mathsf{M}_{1\times n}(\mathbb{F}). \tag{1.3.10}$$

Arguments similar to those in the preceding example show that row A is a subspace of $\mathsf{M}_{1\times n}(\mathbb{F})$. The row vector $\mathbf{x}^\mathsf{T} A$ and the column vector $A^\mathsf{T}\mathbf{x}$ are transposes of each other. This one-to-one correspondence between the elements of row A and col A^T permits us to deduce properties of one of these subspaces from properties of the other.

Example 1.3.11 Let $\mathcal{V} = \mathsf{M}_{m\times n}(\mathbb{F})$ and let $\mathcal{U} \subseteq \mathcal{V}$ be the subset of matrices whose last row has only zero entries. The zero matrix is in $\mathcal{U}$. Sums and scalar multiples of elements of $\mathcal{U}$ have zero last row. It follows that $\mathcal{U}$ is a subspace of $\mathcal{V}$.

Example 1.3.12 Let $A \in \mathsf{M}_m(\mathbb{F})$ and let $\mathcal{U}$ be a subspace of $\mathsf{M}_{m\times n}(\mathbb{F})$. We claim that

$$A\mathcal{U} = \{AX : X \in \mathcal{U}\}$$

is a subspace of $\mathsf{M}_{m\times n}(\mathbb{F})$. Since $0 \in \mathcal{U}$, we have $0 = A0 \in A\mathcal{U}$, which is therefore not empty. Moreover, $cAX + AY = A(cX + Y) \in A\mathcal{U}$ for any scalar c and any $X, Y \in \mathcal{U}$. Theorem 1.3.3 ensures that $A\mathcal{U}$ is a subspace of $\mathsf{M}_{m\times n}(\mathbb{F})$. For example, $A\mathsf{M}_{m\times 1}(\mathbb{F}) = \operatorname{col} A$.

The next four examples involve subspaces whose elements are polynomials.

Example 1.3.13 $\mathcal{P}_5$ is a subspace of $\mathcal{P}$; see Examples 1.2.5 and 1.2.6. Sums and scalar multiples of polynomials of degree 5 or less are in $\mathcal{P}_5$.

Example 1.3.14 $\mathcal{P}_5(\mathbb{R})$ is a subset of $\mathcal{P}_5$, but it is not a subspace. For example, the scalar 1 is in $\mathcal{P}_5(\mathbb{R})$, but $i1 = i \notin \mathcal{P}_5(\mathbb{R})$. The issue here is that the scalars for the vector space $\mathcal{P}_5(\mathbb{R})$ are real numbers and the scalars for the vector space $\mathcal{P}_5$ are complex numbers. A subspace and the vector space that contains it must use the same field of scalars.

Example 1.3.15 A polynomial $p \in \mathcal{P}$ is *even* if $p(-z) = p(z)$ for all z. We denote the set of even polynomials by $\mathcal{P}_{\text{even}}$. A polynomial p is *odd* if $p(-z) = -p(z)$ for all z. We denote the set of odd polynomials by $\mathcal{P}_{\text{odd}}$. For example, $p(z) = 2 + 3z^2$ is even and $p(z) = 5z + 4z^3$ is odd. Constant polynomials are even; the zero polynomial is both even and odd. Each of $\mathcal{P}_{\text{even}}$ and $\mathcal{P}_{\text{odd}}$ is a subspace of $\mathcal{P}$.

Example 1.3.16 The complex vector space $\mathcal{P}$ is a subspace of the complex vector space $C[a,b]$. Every polynomial is a continuous function, and $cp + q \in \mathcal{P}$ whenever $p, q \in \mathcal{P}$ and $c \in \mathbb{C}$. Theorem 1.3.3 ensures that $\mathcal{P}$ is a subspace of $C[a,b]$.

1.4 Linear Combinations, Lists, and Span

The basic operations in an $\mathbb{F}$-vector space permit us to multiply vectors by scalars and then add them. For example, in the real vector space $\mathbb{R}^2$, consider the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$ illustrated in Figure 1.2. A computation reveals that

$$7\mathbf{u} - 5\mathbf{v} + \mathbf{w} = 7\begin{bmatrix}1\\1\end{bmatrix} - 5\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix} = \mathbf{z}, \tag{1.4.1}$$

so $\mathbf{z}$ is a sum of scalar multiples of the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. We also have

$$-\mathbf{u} + \mathbf{v} - \mathbf{w} = -\begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix} = \mathbf{z}, \tag{1.4.2}$$

which expresses $\mathbf{z}$ in two different ways as a sum of scalar multiples of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, respectively. The following definition provides vocabulary to describe computations like these.

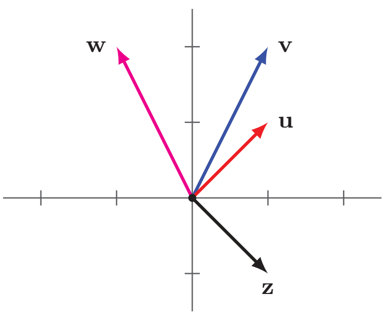

Figure 1.2 Vectors in $\mathbb{R}^2$: $\mathbf{u} = \begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v} = \begin{bmatrix}1\\2\end{bmatrix}$, $\mathbf{w} = \begin{bmatrix}-1\\2\end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix}1\\-1\end{bmatrix}$.

Definition 1.4.3 Let $\mathcal{U}$ be a nonempty subset of an $\mathbb{F}$-vector space $\mathcal{V}$. A *linear combination* of elements of $\mathcal{U}$ is an expression of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r, \tag{1.4.4}$$

in which r is a positive integer, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in \mathcal{U}$, and $c_1, c_2, \ldots, c_r \in \mathbb{F}$. A linear combination (1.4.4) is *trivial* if $c_1 = c_2 = \cdots = c_r = 0$; otherwise, it is *nontrivial*.

A linear combination is, by definition, a sum of finitely many scalar multiples of vectors. For example, (1.4.1) shows that the vector $\mathbf{z}$ in Figure 1.2 is a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$; (1.4.2) expresses $\mathbf{z}$ as a different linear combination of these vectors.

Example 1.4.5 Every element of $\mathcal{P}$ is a linear combination of $1, z, z^2, \ldots$.

Definition 1.4.6 A *list* of vectors in an $\mathbb{F}$-vector space $\mathcal{V}$ is a nonempty finite sequence $\beta = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ of vectors in $\mathcal{V}$. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are the *elements* of the list β. A *nonzero list* has at least one nonzero element. We often denote a list by a lowercase Greek letter such as β or γ.

A subtle, but important, point is that a vector can appear more than once in a list. For example, $\beta = z, z^2, z^2, z^2, z^3$ is a list of five vectors in $\mathcal{P}_3$. However, the set of vectors in the list β is $\{z, z^2, z^3\}$. Sets do not have multiplicities; see Appendix C.1. Accounting for the multiplicities of scalars and vectors is often important in linear algebra; lists help us do this.

A second important point is that order matters in a list. For example, $\beta = z, z^2, z^3$ and $\gamma = z^3, z^2, z$ are different lists of vectors in $\mathcal{P}_3$.

Definition 1.4.7 Let $\mathcal{U}$ be a subset of an $\mathbb{F}$-vector space $\mathcal{V}$. If $\mathcal{U} \neq \varnothing$, then $\operatorname{span}\mathcal{U}$ is the set all of linear combinations of elements of $\mathcal{U}$; we define $\operatorname{span}\varnothing = \{\mathbf{0}\}$. If $\beta = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ is a list of vectors in $\mathcal{V}$, we define $\operatorname{span}\beta = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, that is, the span of a list is the span of the set of vectors in the list.

Suppose that a list of vectors β is obtained from a list γ by reordering its elements. The commutativity of vector addition ensures that $\operatorname{span}\beta = \operatorname{span}\gamma$.

Example 1.4.8 If $\mathbf{u} \in \mathcal{V}$, then Theorem 1.3.3 ensures that $\operatorname{span}\{\mathbf{u}\} = \{c\mathbf{u} : c \in \mathbb{F}\}$ is a subspace of $\mathcal{V}$. In particular, $\operatorname{span}\{\mathbf{0}\} = \{\mathbf{0}\}$.

Example 1.4.9 Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n] \in \mathsf{M}_{m\times n}(\mathbb{F})$ (see (C.2.2) or (3.1.1) for this presentation of a matrix, partitioned according to its columns) and consider the list $\beta = \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ of vectors in the $\mathbb{F}$-vector space $\mathbb{F}^m$. Then

$$\operatorname{span}\beta = \{x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n : x_1, x_2, \ldots, x_n \in \mathbb{F}\} = \{A\mathbf{x} : \mathbf{x} \in \mathbb{F}^n\} = \operatorname{col} A,$$

that is, the span of the columns of a matrix is its column space. A vector $\mathbf{y} \in \mathbb{F}^m$ is in the span of the columns of A if and only if $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$.

The preceding example suggests a powerful result: an inclusion of column spaces is equivalent to the existence of a certain matrix factorization.

Theorem 1.4.10 *Let $Y = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_p] \in \mathsf{M}_{m\times p}(\mathbb{F})$ and let $A \in \mathsf{M}_{m\times n}(\mathbb{F})$. Then $\operatorname{col} Y \subseteq \operatorname{col} A$ if and only if $Y = AX$ for some $X \in \mathsf{M}_{n\times p}(\mathbb{F})$.*

Proof If $\operatorname{col} Y \subseteq \operatorname{col} A$, then each column of Y is in the column space of A. Example 1.4.9 ensures that each $\mathbf{y}_j = A\mathbf{x}_j$ for some $\mathbf{x}_j \in \mathbb{F}^n$. If we let $X = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_p]$, then

$$Y = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_p] = [A\mathbf{x}_1\ A\mathbf{x}_2\ \ldots\ A\mathbf{x}_p] = A[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_p] = AX.$$

Cambridge University Press & Assessment
978-1-108-83710-1 — Matrix Mathematics
2nd Edition
Excerpt
More Information

Conversely, if $Y = AX$, $X \in \mathsf{M}_{n\times p}(\mathbb{F})$, and $\mathbf{u} \in \operatorname{col} Y$, then Example 1.4.9 tells us that $\mathbf{u} = Y\mathbf{v}$ for some $\mathbf{v} \in \mathbb{F}^p$. Consequently, $\mathbf{u} = Y\mathbf{v} = AX\mathbf{v} = A(X\mathbf{v}) \in \operatorname{col} A$, so $\operatorname{col} Y \subseteq \operatorname{col} A$. ☐

Example 1.4.11 Let

$$A = \begin{bmatrix} \mathbf{a}_1^\mathsf{T} \\ \vdots \\ \mathbf{a}_m^\mathsf{T} \end{bmatrix} \in \mathsf{M}_{m\times n}(\mathbb{F})$$

(see (C.2.3) or (3.1.11) for this presentation of a matrix, partitioned according to its rows). If $\beta = \mathbf{a}_1^\mathsf{T}, \mathbf{a}_2^\mathsf{T}, \ldots, \mathbf{a}_m^\mathsf{T}$, then

$$\operatorname{span} \beta = \{x_1\mathbf{a}_1^\mathsf{T} + x_2\mathbf{a}_2^\mathsf{T} + \cdots + x_m\mathbf{a}_m^\mathsf{T} : x_1, x_2, \ldots, x_m \in \mathbb{F}\} = \{\mathbf{x}^\mathsf{T} A : \mathbf{x} \in \mathbb{F}^m\} = \operatorname{row} A,$$

that is, the span of the rows of A is its row space. It can also be thought of as (the transpose of) the span of the columns of A^T.

Example 1.4.12 Consider the list $\beta = z, z^3$ of elements of $\mathcal{P}_3$. Then $\operatorname{span} \beta = \{c_1 z + c_2 z^3 : c_1, c_2 \in \mathbb{C}\}$ is a subspace of $\mathcal{P}_3$ because it is nonempty and

$$c(a_1 z + a_2 z^3) + (b_1 z + b_2 z^3) = (ca_1 + b_1)z + (ca_2 + b_2)z^3$$

is a linear combination of vectors in the list β for all $c, a_1, a_2, b_1, b_2 \in \mathbb{C}$.

The span of a subset of a vector space is always a subspace.

Theorem 1.4.13 *Let $\mathscr{U}$ be a subset of an $\mathbb{F}$-vector space $\mathcal{V}$.*

(a) $\operatorname{span} \mathscr{U}$ *is a subspace of* $\mathcal{V}$.

(b) $\mathscr{U} \subseteq \operatorname{span} \mathscr{U}$.

(c) $\mathscr{U} = \operatorname{span} \mathscr{U}$ *if and only if* $\mathscr{U}$ *is a subspace of* $\mathcal{V}$.

(d) $\operatorname{span}(\operatorname{span} \mathscr{U}) = \operatorname{span} \mathscr{U}$.

Proof First suppose that $\mathscr{U} = \varnothing$. Then Definition 1.4.7 says that $\operatorname{span} \varnothing = \{\mathbf{0}\}$, which is a subspace of $\mathcal{V}$. The empty set is a subset of every set, so $\varnothing \subset \{\mathbf{0}\} = \operatorname{span} \varnothing$. Both implications in (c) are vacuous. For (d), we have $\operatorname{span}(\operatorname{span} \varnothing) = \operatorname{span}\{\mathbf{0}\} = \{\mathbf{0}\} = \operatorname{span} \varnothing$; see Example 1.4.8.

Now suppose that $\mathscr{U} \neq \varnothing$. If $\mathbf{u}, \mathbf{v} \in \operatorname{span} \mathscr{U}$ and $c \in \mathbb{F}$, then each of $\mathbf{u}$, $\mathbf{v}$, $c\mathbf{u}$, and $c\mathbf{u} + \mathbf{v}$ is a linear combination of elements of $\mathscr{U}$, so each is in $\operatorname{span} \mathscr{U}$. Theorem 1.3.3 ensures that $\operatorname{span} \mathscr{U}$ is a subspace of $\mathcal{V}$. The assertion in (b) follows from the fact that $1\mathbf{u} = \mathbf{u}$ is an element of $\operatorname{span} \mathscr{U}$ for each $\mathbf{u} \in \mathscr{U}$. To prove the two implications in (c), first suppose that $\mathscr{U} = \operatorname{span} \mathscr{U}$. Then (a) ensures that $\mathscr{U}$ is a subspace of $\mathcal{V}$. Conversely, if $\mathscr{U}$ is a subspace of $\mathcal{V}$, then it is closed under vector addition and scalar multiplication, so $\operatorname{span} \mathscr{U} \subseteq \mathscr{U}$. The containment $\mathscr{U} \subseteq \operatorname{span} \mathscr{U}$ in (b) ensures that $\mathscr{U} = \operatorname{span} \mathscr{U}$. The assertion in (d) follows from (a) and (c). ☐

Theorem 1.4.14 *Let $\mathscr{U}$ and $\mathscr{W}$ be subsets of an $\mathbb{F}$-vector space $\mathcal{V}$. If $\mathscr{U} \subseteq \mathscr{W}$, then* $\operatorname{span} \mathscr{U} \subseteq \operatorname{span} \mathscr{W}$.

Proof If $\mathscr{U} = \varnothing$, then $\operatorname{span} \mathscr{U} = \{\mathbf{0}\} \subseteq \operatorname{span} \mathscr{W}$. If $\mathscr{U} \neq \varnothing$, then every linear combination of elements of $\mathscr{U}$ is a linear combination of elements of $\mathscr{W}$. ☐

Example 1.4.15 Let $\mathscr{U} = \{1,\, z - 2z^2,\, z^2 + 5z^3,\, z^3,\, 1 + 4z^2\}$. We claim that span $\mathscr{U} = \mathcal{P}_3$. To verify this, observe that

$$\begin{aligned} 1 &= 1, \\ z &= (z - 2z^2) + 2(z^2 + 5z^3) - 10z^3, \\ z^2 &= (z^2 + 5z^3) - 5z^3, \text{ and} \\ z^3 &= z^3. \end{aligned}$$

Thus, $\{1, z, z^2, z^3\} \subseteq \operatorname{span} \mathscr{U} \subseteq \mathcal{P}_3$. Now invoke the two preceding theorems, compute

$$\mathcal{P}_3 = \operatorname{span}\{1, z, z^2, z^3\} \subseteq \operatorname{span}(\operatorname{span} \mathscr{U}) = \operatorname{span} \mathscr{U} \subseteq \mathcal{P}_3,$$

and conclude that span $\mathscr{U} = \mathcal{P}_3$.

Definition 1.4.16 Let $\mathcal{V}$ be an $\mathbb{F}$-vector space. Let $\mathscr{U}$ be a subset of $\mathcal{V}$ and let β be a list of vectors in $\mathcal{V}$. Then $\mathscr{U}$ *spans* $\mathcal{V}$ ($\mathscr{U}$ is a *spanning set* for $\mathcal{V}$) if span $\mathscr{U} = \mathcal{V}$. The list β *spans* $\mathcal{V}$ (β is a *spanning list*) if span $\beta = \mathcal{V}$.

It is convenient to say that "$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ span $\mathcal{V}$" rather than "the list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ spans $\mathcal{V}$." If a list of vectors in $\mathbb{F}^m$ is the list of columns of a matrix $A \in \mathsf{M}_{m\times n}(\mathbb{F})$, it is also convenient to say "the columns of A span $\mathcal{V}$" rather than "the list of columns of A span $\mathcal{V}$."

Example 1.4.17 Let $B \in \mathsf{M}_{n\times p}(\mathbb{F})$. The columns of B span $\mathbb{F}^n$ if and only if $\operatorname{col} B = \mathbb{F}^n$.

Example 1.4.18 Each of the sets $\{1, z, z^2, z^3\}$ and $\{1, z - 2z^2, z^2 + 5z^3, z^3, 1 + 4z^2\}$ spans $\mathcal{P}_3$; see Example 1.4.15.

Example 1.4.19 Let $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n] \in \mathsf{M}_n(\mathbb{F})$ be invertible, let $\mathbf{y} \in \mathbb{F}^n$, and let $A^{-1}\mathbf{y} = [x_1\ x_2\ \ldots\ x_n]^\mathsf{T}$. Then $\mathbf{y} = A(A^{-1}\mathbf{y}) = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ is a linear combination of the columns of A. We conclude that if $A \in \mathsf{M}_n(\mathbb{F})$ is invertible, then its columns span $\mathbb{F}^n$.

Example 1.4.20 The identity matrix I_n is invertible and its columns are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \qquad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{1.4.21}$$

Consequently, $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\} = \mathbb{F}^n$. Any $\mathbf{u} = [u_i] \in \mathbb{F}^n$ can be expressed as $\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \cdots + u_n\mathbf{e}_n$. For example, the *all-ones vector* $\mathbf{e} \in \mathbb{F}^n$ can be expressed as $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n = [1\ 1\ \ldots\ 1]^\mathsf{T}$.

Example 1.4.22 Consider the vectors $\mathbf{u}$ and $\mathbf{v}$ in Figure 1.2, and let $A = [\mathbf{u}\ \mathbf{v}] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. A computation using (C.2.7) reveals that $A^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$, so A is invertible. The preceding example ensures that $\operatorname{span}\{\mathbf{u}, \mathbf{v}\} = \mathbb{R}^2$, so each vector in $\mathbb{R}^2$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Equivalently, the system of linear equations $A\mathbf{x} = \mathbf{y}$ is consistent for each $\mathbf{y} \in \mathbb{R}^2$.

Example 1.4.23 Let $A \in \mathsf{M}_{m\times n}(\mathbb{F})$ and $B \in \mathsf{M}_{n\times p}(\mathbb{F})$. If the columns of B span $\mathbb{F}^n$, then Example 1.4.19 ensures that

$$\operatorname{col} AB = \{AB\mathbf{x} : \mathbf{x} \in \mathbb{F}^p\} = \{A\mathbf{y} : \mathbf{y} \in \mathbb{F}^n\} = \operatorname{col} A.$$