

Chapter 2

Linear Algebra Essentials

When elementary school students first leave the solid ground of arithmetic for the more abstract world of algebra, the first objects they encounter are generally linear expressions. Algebraically, linear equations can be solved using elementary field properties, namely the existence of additive and multiplicative inverses. Geometrically, a nonvertical line in the plane through the origin can be described completely by one number—the slope. Linear functions $f : \mathbf{R} \rightarrow \mathbf{R}$ enjoy other nice properties: They are (in general) invertible, and the composition of linear functions is again linear.

Yet marching through the progression of more complicated functions and expressions—polynomial, algebraic, transcendental—many of these basic properties of linearity can become taken for granted. In the standard calculus sequence, sophisticated techniques are developed that seem to yield little new information about linear functions. Linear algebra is generally introduced after the basic calculus sequence has been nearly completed, and is presented in a self-contained manner, with little reference to what has been seen before. A fundamental insight is lost or obscured: that differential calculus is the study of nonlinear phenomena by “linearization.”

The main goal of this chapter is to present the basic elements of linear algebra needed to understand this insight of differential calculus. We also present some geometric applications of linear algebra with an eye toward later constructions in differential geometry.

While this chapter is written for readers who have already been exposed to a first course in linear algebra, it is self-contained enough that the only essential prerequisites will be a working knowledge of matrix algebra, Gaussian elimination, and determinants.

2.1 Vector Spaces

Modern mathematics can be described as the study of sets with some extra associated “structure.” In linear algebra, the sets under consideration have enough structure to allow elements to be added and multiplied by scalars. These two operations should behave and interact in familiar ways.

Definition 2.1.1. A (real) *vector space* consists of a set V together with two operations, addition and scalar multiplication.¹ Scalars are understood here as real numbers. Elements of V are called *vectors* and will often be written in bold type, as $\mathbf{v} \in V$. Addition is written using the conventional symbolism $\mathbf{v} + \mathbf{w}$. Scalar multiplication is denoted by $s\mathbf{v}$ or $s \cdot \mathbf{v}$.

The triple $(V, +, \cdot)$ must satisfy the following axioms:

- (V1) For all $\mathbf{v}, \mathbf{w} \in V$, $\mathbf{v} + \mathbf{w} \in V$.
- (V2) For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
- (V3) For all $\mathbf{v}, \mathbf{w} \in V$, $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
- (V4) There exists a distinguished element of V , called the *zero vector* and denoted by $\mathbf{0}$, with the property that for all $\mathbf{v} \in V$, $\mathbf{0} + \mathbf{v} = \mathbf{v}$.
- (V5) For all $\mathbf{v} \in V$, there exists an element called the *additive inverse of $\mathbf{v}$* and denoted $-\mathbf{v}$, with the property that $(-\mathbf{v}) + \mathbf{v} = \mathbf{0}$.
- (V6) For all $s \in \mathbf{R}$ and $\mathbf{v} \in V$, $s\mathbf{v} \in V$.
- (V7) For all $s, t \in \mathbf{R}$ and $\mathbf{v} \in V$, $s(t\mathbf{v}) = (st)\mathbf{v}$.
- (V8) For all $s, t \in \mathbf{R}$ and $\mathbf{v} \in V$, $(s + t)\mathbf{v} = s\mathbf{v} + t\mathbf{v}$.
- (V9) For all $s \in \mathbf{R}$ and $\mathbf{v}, \mathbf{w} \in V$, $s(\mathbf{v} + \mathbf{w}) = s\mathbf{v} + s\mathbf{w}$.
- (V10) For all $\mathbf{v} \in V$, $1\mathbf{v} = \mathbf{v}$.

We will often suppress the explicit ordered triple notation $(V, +, \cdot)$ and simply refer to “the vector space V .”

In an elementary linear algebra course, a number of familiar properties of vector spaces are derived as consequences of the 10 axioms. We list several of them here.

Theorem 2.1.2. *Let V be a vector space. Then:*

1. *The zero vector $\mathbf{0}$ is unique.*
2. *For all $\mathbf{v} \in V$, the additive inverse $-\mathbf{v}$ of $\mathbf{v}$ is unique.*
3. *For all $\mathbf{v} \in V$, $0 \cdot \mathbf{v} = \mathbf{0}$.*
4. *For all $\mathbf{v} \in V$, $(-1) \cdot \mathbf{v} = -\mathbf{v}$.*

Proof. Exercise. □

Physics texts often discuss vectors in terms of the two properties of magnitude and direction. These are not in any way related to the vector space axioms. Both of

¹More formally, addition can be described as a function $V \times V \rightarrow V$ and scalar multiplication as a function $\mathbf{R} \times V \rightarrow V$.

these concepts arise naturally in the context of *inner product spaces*, which we treat in Sect. 2.9.

In a first course in linear algebra, a student is exposed to a number of examples of vector spaces, familiar and not-so-familiar, in order to gain better acquaintance with the axioms. Here we introduce just two examples.

Example 2.1.3. For any positive integer n , define the set $\mathbf{R}^n$ to be the set of all n -tuples of real numbers:

$$\mathbf{R}^n = \{(a_1, \dots, a_n) \mid a_i \in \mathbf{R} \text{ for } i = 1, \dots, n\}$$

Define vector addition componentwise by

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n),$$

and likewise define scalar multiplication by

$$s(a_1, \dots, a_n) = (sa_1, \dots, sa_n).$$

It is a straightforward exercise to show that $\mathbf{R}^n$ with these operations satisfies the vector space axioms. These vector spaces (one for each natural number n) will be called *Euclidean spaces*.

The Euclidean spaces can be thought of as the “model” finite-dimensional vector spaces in at least two senses. First, they are the most familiar examples, generalizing the set $\mathbf{R}^2$ that is the setting for the most elementary analytic geometry that most students first encounter in high school. Second, we show later that every finite-dimensional vector space is “equivalent” (in a sense we will make precise) to $\mathbf{R}^n$ for some n .

Much of the work in later chapters will concern $\mathbf{R}^3$, $\mathbf{R}^4$, and other Euclidean spaces. We will be relying on additional structures of these sets that go beyond the bounds of linear algebra. Nevertheless, the vector space structure remains essential to the tools of calculus that we will employ later.

The following example gives a class of vector spaces that are in general not equivalent to Euclidean spaces.

Example 2.1.4 (Vector spaces of functions). For any set X , let $\mathcal{F}(X)$ be the set of all real-valued functions $f : X \rightarrow \mathbf{R}$. For every two such $f, g \in \mathcal{F}(X)$, define the sum $f + g$ pointwise as $(f + g)(x) = f(x) + g(x)$. Likewise, define scalar multiplication $(sf)(x) = s(f(x))$. The set $\mathcal{F}(X)$ equipped with these operations is a vector space. The zero vector is the function $O : X \rightarrow \mathbf{R}$ that is identically zero: $O(x) = 0$ for all $x \in X$. Confirmation of the axioms depends on the corresponding field properties in the codomain, the set of real numbers.

We will return to this class of vector spaces in the next section.

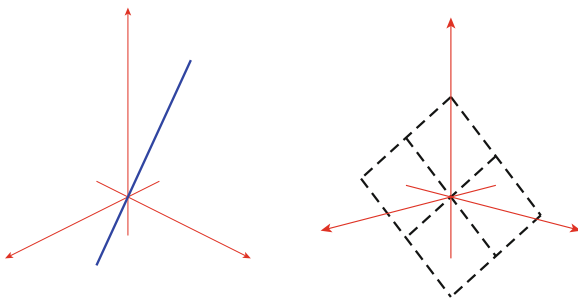


Fig. 2.1 Subspaces in $\mathbf{R}^3$.

2.2 Subspaces

A mathematical structure on a set distinguishes certain subsets of special significance. In the case of a set with the structural axioms of a vector space, the distinguished subsets are those that are themselves vector spaces under the same operations of vector addition and scalar multiplication as in the larger set.

Definition 2.2.1. Let W be a subset of a vector space $(V, +, \cdot)$. Then W is a *vector subspace* (or just *subspace*) of V if $(W, +, \cdot)$ satisfies the vector space axioms (V1)–(V10).

A subspace can be pictured as a vector space “within” a larger vector space. See Fig. 2.1.

Before illustrating examples of subspaces, we immediately state a theorem that ensures that most of the vector space axioms are in fact inherited from the larger ambient vector space.

Theorem 2.2.2. Suppose $W \subset V$ is a nonempty subset of a vector space V satisfying the following two properties:

- (W1) For all $\mathbf{v}, \mathbf{w} \in W$, $\mathbf{v} + \mathbf{w} \in W$.
- (W2) For all $\mathbf{w} \in W$ and $s \in \mathbf{R}$, $s\mathbf{w} \in W$.

Then W is a subspace of V .

Proof. Exercise. □

We note that for every vector space V , the set $\{\mathbf{0}\}$ is a subspace of V , known as the *trivial subspace*. Similarly, V is a subspace of itself, which is known as the *improper subspace*.

We now illustrate some nontrivial, proper subspaces of the vector space $\mathbf{R}^3$. We leave the verifications that they are in fact subspaces to the reader.

Example 2.2.3. Let $W_1 = \{(s, 0, 0) \mid s \in \mathbf{R}\}$. Then W_1 is a subspace of $\mathbf{R}^3$.

Example 2.2.4. Let $\mathbf{v} = (a, b, c) \neq \mathbf{0}$ and let $W_2 = \{s\mathbf{v} \mid s \in \mathbf{R}\}$. Then W_2 is a subspace of $\mathbf{R}^3$. Note that Example 2.2.3 is a special case of this example when $\mathbf{v} = (1, 0, 0)$.

Example 2.2.5. Let $W_3 = \{(s, t, 0) \mid s, t \in \mathbf{R}\}$. Then W_3 is a subspace of $\mathbf{R}^3$.

Example 2.2.6. As in Example 2.2.4, let $\mathbf{v} = (a, b, c) \neq \mathbf{0}$. Relying on the usual “dot product” in $\mathbf{R}^3$, define

$$\begin{aligned} W_4 &= \{\mathbf{x} \in \mathbf{R}^3 \mid \mathbf{v} \cdot \mathbf{x} = 0\} \\ &= \{(x_1, x_2, x_3) \mid ax_1 + bx_2 + cx_3 = 0\}. \end{aligned}$$

Then W_4 is a subspace of $\mathbf{R}^3$. Note that Example 2.2.5 is a special case of this example when $\mathbf{v} = (0, 0, 1)$.

We will show at the end of Sect. 2.4 that all proper, nontrivial subspaces of $\mathbf{R}^3$ can be realized either in the form of W_2 or W_4 .

Example 2.2.7 (Subspaces of $\mathcal{F}(\mathbf{R})$). We list here a number of vector subspaces of $\mathcal{F}(\mathbf{R})$, the space of real-valued functions $f : \mathbf{R} \rightarrow \mathbf{R}$. The verifications that they are in fact subspaces are straightforward exercises using the basic facts of algebra and calculus.

- $P_n(\mathbf{R})$, the subspace of polynomial functions of degree n or less;
- $P(\mathbf{R})$, the subspace of all polynomial functions (of any degree);
- $C(\mathbf{R})$, the subspace of functions that are continuous at each point in their domain;
- $C^r(\mathbf{R})$, the subspace of functions whose first r derivatives exist and are continuous at each point in their domain;
- $C^\infty(\mathbf{R})$, the subspace of functions all of whose derivatives exist and are continuous at each point in their domain.

Our goal in the next section will be to exhibit a method for constructing vector subspaces of any vector space V .

2.3 Constructing Subspaces I: Spanning Sets

The two vector space operations give a way to produce new vectors from a given set of vectors. This, in turn, gives a basic method for constructing subspaces. We mention here that for the remainder of the chapter, when we specify that a set is *finite* as an assumption, we will also assume that the set is nonempty.

Definition 2.3.1. Suppose $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a finite set of vectors in a vector space V . A vector $\mathbf{w}$ is a *linear combination* of S if there are scalars $c_1, \dots, c_n$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n.$$

A basic question in a first course in linear algebra is this: For a vector $\mathbf{w}$ and a set S as in Definition 2.3.1, decide whether $\mathbf{w}$ is a linear combination of S . In practice, this can be answered using the tools of matrix algebra.

Example 2.3.2. Let $S = \{\mathbf{v}_1, \mathbf{v}_2\} \subset \mathbf{R}^3$, where $\mathbf{v}_1 = (1, 2, 3)$ and $\mathbf{v}_2 = (-1, 4, 2)$. Let us decide whether $\mathbf{w} = (29, -14, 27)$ is a linear combination of S . To do this means solving the vector equation $\mathbf{w} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2$ for the two scalars s_1, s_2 , which in turn amounts to solving the system of linear equations

$$\begin{cases} s_1(1) + s_2(-1) = 29, \\ s_1(2) + s_2(4) = -14, \\ s_1(3) + s_2(2) = 27. \end{cases}$$

Gaussian elimination of the corresponding augmented matrix yields

$$\left[\begin{array}{cc|c} 1 & 0 & 17 \\ 0 & 1 & -12 \\ 0 & 0 & 0 \end{array} \right],$$

corresponding to the unique solution $s_1 = 17$, $s_2 = -12$. Hence, $\mathbf{w}$ is a linear combination of S .

The reader will notice from this example that deciding whether a vector is a linear combination of a given set ultimately amounts to deciding whether the corresponding system of linear equations is consistent.

We will now use Definition 2.3.1 to obtain a method for constructing subspaces. **Definition 2.3.3.** Let V be a vector space and let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset V$ be a finite set of vectors. The *span* of S , denoted by $\text{Span}(S)$, is defined to be the set of all linear combinations of S :

$$\text{Span}(S) = \{s_1\mathbf{v}_1 + \dots + s_n\mathbf{v}_n \mid s_1, \dots, s_n \in \mathbf{R}\}.$$

We note immediately the utility of this construction.

Theorem 2.3.4. Let $S \subset V$ be a finite set of vectors. Then $W = \text{Span}(S)$ is a subspace of V .

Proof. The proof is an immediate application of Theorem 2.2.2. $\square$

We will say that S *spans* the subspace W , or that S is a *spanning set* for the subspace W .

Example 2.3.5. Let $S = \{\mathbf{v}_1\} \subset \mathbf{R}^3$, where $\mathbf{v}_1 = (1, 0, 0)$. Then $\text{Span}(S) = \{s(1, 0, 0) \mid s \in \mathbf{R}\} = \{(s, 0, 0) \mid s \in \mathbf{R}\}$. Compare to Example 2.2.3.

Example 2.3.6. Let $S = \{\mathbf{v}_1, \mathbf{v}_2\} \subset \mathbf{R}^4$, where $\mathbf{v}_1 = (1, 0, 0, 0)$ and $\mathbf{v}_2 = (0, 0, 1, 0)$. Then

$$\text{Span}(S) = \{s(1, 0, 0, 0) + t(0, 0, 1, 0) \mid s, t \in \mathbf{R}\} = \{(s, 0, t, 0) \mid s, t \in \mathbf{R}\}.$$

Example 2.3.7. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset \mathbf{R}^3$ where $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, and $\mathbf{v}_3 = (0, 0, 1)$. Then

$$\begin{aligned}\text{Span}(S) &= \{s_1(1, 0, 0) + s_2(0, 1, 0) + s_3(0, 0, 1) \mid s_1, s_2, s_3 \in \mathbf{R}\} \\ &= \{(s_1, s_2, s_3) \mid s_1, s_2, s_3 \in \mathbf{R}\} \\ &= \mathbf{R}^3.\end{aligned}$$

Example 2.3.8. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \subset \mathbf{R}^3$, where $\mathbf{v}_1 = (1, 1, 1)$, $\mathbf{v}_2 = (-1, 1, 0)$, $\mathbf{v}_3 = (1, 3, 2)$, and $\mathbf{v}_4 = (-3, 1, -1)$. Then

$$\begin{aligned}\text{Span}(S) &= \{s_1(1, 1, 1) + s_2(-1, 1, 0) \\ &\quad + s_3(1, 3, 2) + s_4(-3, 1, -1) \mid s_1, s_2, s_3, s_4 \in \mathbf{R}\} \\ &= \{(s_1 - s_2 + s_3 - 3s_4, s_1 + s_2 + 3s_3 + s_4, \\ &\quad s_1 + 2s_3 - s_4) \mid s_1, s_2, s_3, s_4 \in \mathbf{R}\}.\end{aligned}$$

For example, consider $\mathbf{w} = (13, 3, 8) \in \mathbf{R}^3$. Then $\mathbf{w} \in \text{Span}(S)$, since $\mathbf{w} = \mathbf{v}_1 - \mathbf{v}_2 + 2\mathbf{v}_3 - 3\mathbf{v}_4$.

Note that this set of four vectors S in $\mathbf{R}^3$ *does not* span $\mathbf{R}^3$. To see this, take an arbitrary $\mathbf{w} \in \mathbf{R}^3$, $\mathbf{w} = (w_1, w_2, w_3)$. If $\mathbf{w}$ is a linear combination of S , then there are scalars s_1, s_2, s_3, s_4 such that $\mathbf{w} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + s_3\mathbf{v}_3 + s_4\mathbf{v}_4$. In other words, if $\mathbf{w} \in \text{Span}(S)$, then the system

$$\begin{cases} s_1 - s_2 + s_3 - 3s_4 = w_1, \\ s_1 + s_2 + 3s_3 + s_4 = w_2, \\ s_1 + 2s_3 - s_4 = w_3, \end{cases}$$

is consistent: we can solve for s_1, s_2, s_3, s_4 in terms of w_1, w_2, w_3 . Gaussian elimination of the corresponding augmented matrix

$$\left[\begin{array}{cccc|c} 1 & -1 & 1 & -3 & w_1 \\ 1 & 1 & 3 & 1 & w_2 \\ 1 & 0 & 2 & -1 & w_3 \end{array} \right]$$

yields

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & -1 & w_3 \\ 0 & 1 & 1 & 2 & -w_1 + w_3 \\ 0 & 0 & 0 & 0 & w_1 + w_2 - 2w_3 \end{array} \right].$$

Hence for every vector $\mathbf{w}$ such that $w_1 + w_2 - 2w_3 \neq 0$, the system is not consistent and $\mathbf{w} \notin \text{Span}(S)$. For example, $(1, 1, 2) \notin \text{Span}(S)$.

We return to this example below.

Note that a given subspace may have many different spanning sets. For example, consider $S = \{(1, 0, 0), (1, 1, 0), (1, 1, 1)\} \subset \mathbf{R}^3$. The reader may verify that S is a spanning set for $\mathbf{R}^3$. But in Example 2.3.7, we exhibited a different spanning set for $\mathbf{R}^3$.

2.4 Linear Independence, Basis, and Dimension

In the preceding section, we started with a finite set $S \subset V$ in order to generate a subspace $W = \text{Span}(S)$ in V . This procedure prompts the following question: For a subspace W , can we find a spanning set for W ? If so, what is the “smallest” such set? These questions lead naturally to the notion of a basis. Before defining that notion, however, we introduce the concepts of linear dependence and independence.

For a vector space V , a finite set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, and a vector $\mathbf{w} \in V$, we have already considered the question whether $\mathbf{w} \in \text{Span}(S)$. Intuitively, we might say that $\mathbf{w}$ “depends linearly” on S if $\mathbf{w} \in \text{Span}(S)$, i.e., if $\mathbf{w}$ can be written as a linear combination of elements of S . In the simplest case, for example, that $S = \{\mathbf{v}\}$, then $\mathbf{w}$ “depends on” S if $\mathbf{w} = s\mathbf{v}$, or, what is the same, $\mathbf{w}$ is “independent” of S if $\mathbf{w}$ is not a scalar multiple of $\mathbf{v}$.

The following definition aims to make this sense of dependence precise.

Definition 2.4.1. A finite set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is *linearly dependent* if there are scalars $s_1, \dots, s_n$, *not all zero*, such that

$$s_1\mathbf{v}_1 + \dots + s_n\mathbf{v}_n = \mathbf{0}.$$

If S is not linearly dependent, then it is *linearly independent*.

The positive way of defining linear independence, then, is that a finite set of vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent if the condition that there are scalars $s_1, \dots, s_n$ satisfying $s_1\mathbf{v}_1 + \dots + s_n\mathbf{v}_n = \mathbf{0}$ implies that

$$s_1 = \dots = s_n = 0.$$

Example 2.4.2. We refer back to the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \subset \mathbf{R}^3$, where $\mathbf{v}_1 = (1, 1, 1)$, $\mathbf{v}_2 = (-1, 1, 0)$, $\mathbf{v}_3 = (1, 3, 2)$, and $\mathbf{v}_4 = (-3, 1, -1)$, in Example 2.3.8. We will show that the set S is linearly dependent. In other words, we will find scalars s_1, s_2, s_3, s_4 , not all zero, such that $s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + s_3\mathbf{v}_3 + s_4\mathbf{v}_4 = \mathbf{0}$.

This amounts to solving the homogeneous system

$$\begin{cases} s_1 - s_2 + s_3 - 3s_4 = 0, \\ s_1 + s_2 + 3s_3 + s_4 = 0, \\ s_1 + 2s_3 - s_4 = 0. \end{cases}$$

Gaussian elimination of the corresponding augmented matrix yields

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & -1 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

This system has nontrivial solutions of the form $s_1 = -2t + u$, $s_2 = -t - 2u$, $s_3 = t$, $s_4 = u$. The reader can verify, for example, that

$$(-1)\mathbf{v}_1 + (-3)\mathbf{v}_2 + (1)\mathbf{v}_3 + (1)\mathbf{v}_4 = \mathbf{0}.$$

Hence S is linearly dependent.

Example 2.4.2 illustrates the fact that deciding whether a set is linearly dependent amounts to deciding whether a corresponding *homogeneous* system of linear equations has nontrivial solutions.

The following facts are consequences of Definition 2.4.1. The reader is invited to supply proofs.

Theorem 2.4.3. *Let S be a finite set of vectors in a vector space V . Then:*

1. *If $\mathbf{0} \in S$, then S is linearly dependent.*
2. *If $S = \{\mathbf{v}\}$ and $\mathbf{v} \neq \mathbf{0}$, then S is linearly independent.*
3. *Suppose S has at least two vectors. Then S is a linearly dependent set of nonzero vectors if and only if there exists a vector in S that can be written as a linear combination of the others.*

Linear dependence or independence has important consequences related to the notion of spanning sets. For example, the following theorem asserts that enlarging a set by adding linearly dependent vectors does not change the spanning set.

Theorem 2.4.4. *Let S be a finite set of vectors in a vector space V . Let $\mathbf{w} \in \text{Span}(S)$, and let $S' = S \cup \{\mathbf{w}\}$. Then $\text{Span}(S') = \text{Span}(S)$.*

Proof. Exercise. □

Generating “larger” subspaces thus requires adding vectors that are *linearly independent* of the original spanning set.

We return to a version of the question at the outset of this section: If we are given a subspace, what is the “smallest” subset that can serve as a spanning set for this subspace? This motivates the definition of a basis.

Definition 2.4.5. Let V be a vector space. A *basis* for V is a set $B \subset V$ such that (1) $\text{Span}(B) = V$ and (2) B is a linearly independent set.

Example 2.4.6. For the vector space $V = \mathbf{R}^n$, the set $B_0 = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, where $\mathbf{e}_1 = (1, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, $\dots$, $\mathbf{e}_n = (0, \dots, 0, 1)$, is a basis for $\mathbf{R}^n$. The set B_0 is called the *standard basis* for $\mathbf{R}^n$.

Example 2.4.7. Let $V = \mathbf{R}^3$ and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = (1, 4, -1)$, $\mathbf{v}_2 = (1, 1, 1)$, and $\mathbf{v}_3 = (2, 0, -1)$. To show that S is a basis for $\mathbf{R}^3$, we need to show that S spans $\mathbf{R}^3$ and that S is linearly independent. To show that S spans $\mathbf{R}^3$ requires choosing an arbitrary vector $\mathbf{w} = (w_1, w_2, w_3) \in \mathbf{R}^3$ and finding scalars c_1, c_2, c_3 such that $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. To show that S is linearly independent requires showing that the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$ has only the trivial solution $c_1 = c_2 = c_3 = 0$.

Both requirements involve analyzing systems of linear equations with coefficient matrix

$$A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3] = \begin{bmatrix} 1 & 1 & 2 \\ 4 & 1 & 0 \\ -1 & 1 & -1 \end{bmatrix},$$

in the first case the equation $A\mathbf{c} = \mathbf{w}$ (to determine whether it is consistent for all $\mathbf{w}$) and in the second case $A\mathbf{c} = \mathbf{0}$ (to determine whether it has only the trivial solution). Here $\mathbf{c} = (c_1, c_2, c_3)$ is the vector of coefficients. Both conditions are established by noting that $\det(A) \neq 0$. Hence S spans $\mathbf{R}^3$ and S is linearly independent, so S is a basis for $\mathbf{R}^3$.

The computations in Example 2.4.7 in fact point to a proof of a powerful technique for determining whether a set of vectors in $\mathbf{R}^n$ forms a basis for $\mathbf{R}^n$.

Theorem 2.4.8. *A set of n vectors $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbf{R}^n$ forms a basis for $\mathbf{R}^n$ if and only if $\det(A) \neq 0$, where $A = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ is the matrix formed by the column vectors $\mathbf{v}_i$.*

Just as we noted earlier that a vector space may have many spanning sets, the previous two examples illustrate that a vector space does not have a *unique* basis.

By definition, a basis B for a vector space V spans V , and so every element of V can be written as a linear combination of elements of B . However, the requirement that B be a linearly independent set has an important consequence.

Theorem 2.4.9. *Let B be a finite basis for a vector space V . Then each vector $\mathbf{v} \in V$ can be written uniquely as a linear combination of elements of B .*

Proof. Suppose that there are two different ways of expressing a vector $\mathbf{v}$ as a linear combination of elements of $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, so that there are scalars $c_1, \dots, c_n$ and $d_1, \dots, d_n$ such that

$$\mathbf{v} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$$

$$\mathbf{v} = d_1\mathbf{b}_1 + \cdots + d_n\mathbf{b}_n.$$

Then

$$(c_1 - d_1)\mathbf{b}_1 + \cdots + (c_n - d_n)\mathbf{b}_n = \mathbf{0}.$$

By the linear independence of the set B , this implies that

$$c_1 = d_1, \dots, c_n = d_n;$$

in other words, the two representations of $\mathbf{v}$ were in fact the same. □

As a consequence of Theorem 2.4.9, we introduce the following notation. Let $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V . Then for every $\mathbf{v} \in V$, let $[\mathbf{v}]_B \in \mathbf{R}^n$ be defined to be

$$[\mathbf{v}]_B = (v_1, \dots, v_n),$$

where $\mathbf{v} = v_1\mathbf{b}_1 + \dots + v_n\mathbf{b}_n$.

The following theorem is fundamental.

Theorem 2.4.10. *Let V be a vector space and let B be a basis for V that contains n vectors. Then no set with fewer than n vectors spans V , and no set with more than n vectors is linearly independent.*

Proof. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a finite set of nonzero vectors in V . Since B is a basis, for each $i = 1, \dots, m$ there are unique scalars $a_{i1}, \dots, a_{in}$ such that

$$\mathbf{v}_i = a_{i1}\mathbf{b}_1 + \dots + a_{in}\mathbf{b}_n.$$

Let A be the $m \times n$ matrix of components $A = [a_{ij}]$.

Suppose first that $m < n$. For $\mathbf{w} \in V$, suppose that there are scalars $c_1, \dots, c_m$ such that

$$\begin{aligned} \mathbf{w} &= c_1\mathbf{v}_1 + \dots + c_m\mathbf{v}_m \\ &= c_1(a_{11}\mathbf{b}_1 + \dots + a_{1n}\mathbf{b}_n) + \dots + c_m(a_{m1}\mathbf{b}_1 + \dots + a_{mn}\mathbf{b}_n) \\ &= (c_1a_{11} + \dots + c_ma_{m1})\mathbf{b}_1 + \dots + (c_1a_{1n} + \dots + c_ma_{mn})\mathbf{b}_n. \end{aligned}$$

Writing $[\mathbf{w}]_B = (w_1, \dots, w_n)$ relative to the basis B , the above vector equation can be written in matrix form $A^T\mathbf{c} = [\mathbf{w}]_B$, where $\mathbf{c} = (c_1, \dots, c_m)$. But since $m < n$, the row echelon form of the $(n \times m)$ matrix A^T must have a row of zeros, and so there exists a vector $\mathbf{w}_0 \in V$ such that $A^T\mathbf{c} = [\mathbf{w}_0]_B$ is not consistent. But this means that $\mathbf{w}_0 \notin \text{Span}(S)$, and so S does not span V .

Likewise, if $m > n$, then the row echelon form of A^T has at most n leading ones. Then the vector equation $A^T\mathbf{c} = \mathbf{0}$ has nontrivial solutions, and S is not linearly independent. $\square$

Corollary 2.4.11. *Let V be a vector space and let B be a basis of n vectors for V . Then every other basis B' of V must also have n elements.*

The corollary prompts the following definition.

Definition 2.4.12. Let V be a vector space. If there is no finite subset of V that spans V , then V is said to be *infinite-dimensional*. On the other hand, if V has a basis of n vectors (and hence, by Corollary 2.4.11, every basis has n vectors), then V is *finite-dimensional*. We call n the *dimension* of V and we write $\dim(V) = n$. By definition, $\dim(\{\mathbf{0}\}) = 0$.

Most of the examples we consider here will be finite-dimensional. However, of the vector spaces listed in Example 2.2.7, only P_n is finite-dimensional.

We conclude this section by considering the dimension of a subspace. Since a subspace is itself a vector space, Definition 2.4.12 makes sense in this context.

Theorem 2.4.13. *Let V be a finite-dimensional vector space, and let W be a subspace of V . Then $\dim(W) \leq \dim(V)$, with $\dim(W) = \dim(V)$ if and only if $W = V$. In particular, W is finite-dimensional.*

Proof. Exercise. □

Example 2.4.14. Recall $W_2 \subset \mathbf{R}^3$ from Example 2.2.4:

$$W_2 = \{(sa, sb, sc) \mid s \in \mathbf{R}\},$$

where $(a, b, c) \neq \mathbf{0}$. We have $W_2 = \text{Span}(\{(a, b, c)\})$, and also the set $\{(a, b, c)\}$ is linearly independent by Theorem 2.4.3, so $\dim(W_2) = 1$.

Example 2.4.15. Recall $W_4 \subset \mathbf{R}^3$ from Example 2.2.6:

$$W_4 = \{(x, y, z) \mid ax + by + cz = 0\}$$

for some $(a, b, c) \neq \mathbf{0}$. Assume without loss of generality that $a \neq 0$. Then W_4 can be seen to be spanned by the set $S = \{(-b, a, 0), (-c, 0, a)\}$. Since S is a linearly independent set, $\dim(W_4) = 2$.

Example 2.4.16. We now justify the statement at the end of Sect. 2.3: Every proper, nontrivial subspace of $\mathbf{R}^3$ is of the form W_2 or W_4 above. Let W be a subspace of $\mathbf{R}^3$. If it is a proper subspace, then $\dim(W) = 1$ or $\dim(W) = 2$. If $\dim(W) = 1$, then W has a basis consisting of one element $\mathbf{a} = (a, b, c)$, and so W has the form of W_2 .

If $\dim(W) = 2$, then W has a basis of two linearly independent vectors $\{\mathbf{a}, \mathbf{b}\}$, where $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$. Let

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1),$$

obtained using the vector cross product in $\mathbf{R}^3$. Note that $\mathbf{c} \neq \mathbf{0}$ by virtue of the linear independence of $\mathbf{a}$ and $\mathbf{b}$. The reader may verify that $\mathbf{w} = (x, y, z) \in W$ exactly when

$$\mathbf{c} \cdot \mathbf{w} = 0,$$

and so W has the form W_4 above.

Example 2.4.17. Recall the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \subset \mathbf{R}^3$, where $\mathbf{v}_1 = (1, 1, 1)$, $\mathbf{v}_2 = (-1, 1, 0)$, $\mathbf{v}_3 = (1, 3, 2)$, and $\mathbf{v}_4 = (-3, 1, -1)$, from Example 2.3.8. In that example we showed that S did not span $\mathbf{R}^3$, and so S cannot be a basis for $\mathbf{R}^3$. In fact, in Example 2.4.2, we showed that S is linearly dependent. A closer look at that example shows that the rank of the matrix $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \mathbf{v}_4]$ is two. A basis

for $W = \text{Span}(S)$ can be obtained by choosing vectors in S whose corresponding column in the row-echelon form has a leading one. In this case, $S' = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W , and so $\dim(W) = 2$.

2.5 Linear Transformations

For a set along with some extra structure, the next notion to consider is a function between the sets that in some suitable sense “preserves the structure.” In the case of linear algebra, such functions are known as *linear transformations*. The structure they preserve should be the vector space operations of addition and scalar multiplication.

In what follows, we consider two vector spaces V and W . The reader might benefit at this point from reviewing Sect. 1.2 on functions in order to review the terminology and relevant definitions.

Definition 2.5.1. A function $T : V \rightarrow W$ is a *linear transformation* if (1) for all $\mathbf{u}, \mathbf{v} \in V$, $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$; and (2) for all $s \in \mathbf{R}$ and $\mathbf{v} \in V$, $T(s\mathbf{v}) = sT(\mathbf{v})$ (Fig. 2.2).

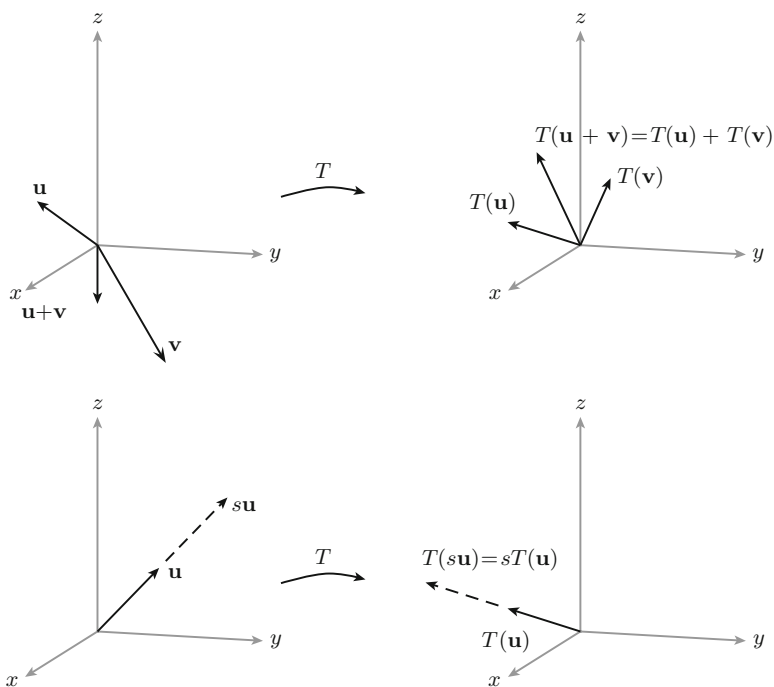


Fig. 2.2 The two conditions defining a linear transformation.

The two requirements for a function to be a linear transformation correspond exactly to the two vector space operations—the “structure”—on the sets V and W . The correct way of understanding these properties is to think of the function as “commuting” with the vector space operations: Performing the operation first (in V) and then applying the function yields the same result as applying the function first and then performing the operations (in W). It is in this sense that linear transformations “preserve the vector space structure.”

We recall some elementary properties of linear transformations that are consequences of Definition 2.5.1.

Theorem 2.5.2. *Let V and W be vector spaces with corresponding zero vectors $\mathbf{0}_V$ and $\mathbf{0}_W$. Let $T : V \rightarrow W$ be a linear transformation. Then*

1. $T(\mathbf{0}_V) = \mathbf{0}_W$.
2. For all $\mathbf{u} \in V$, $T(-\mathbf{u}) = -T(\mathbf{u})$.

Proof. Keeping in mind Theorem 2.1.2, both of these statements are consequences of the second condition in Definition 2.5.1, using $s = 0$ and $s = -1$ respectively. $\square$

The one-to-one, onto linear transformations play a special role in linear algebra. They allow one to say that two different vector spaces are “the same.”

Definition 2.5.3. Suppose V and W are vector spaces. A linear transformation $T : V \rightarrow W$ is a *linear isomorphism* if it is one-to-one and onto. Two vector spaces V and W are said to be *isomorphic* if there is a linear isomorphism $T : V \rightarrow W$.

The most basic example of a linear isomorphism is the identity transformation $\text{Id}_V : V \rightarrow V$ given by $\text{Id}_V(\mathbf{v}) = \mathbf{v}$. We shall see other examples shortly.

The concept of linear isomorphism is an example of a recurring notion in this text. The fact that an isomorphism between vector spaces V and W is one-to-one and onto says that V and W are the “same” *as sets*; there is a pairing between vectors in V and W . The fact that a linear isomorphism is in fact a linear transformation further says that V and W have the same *structure*. Hence when V and W are isomorphic as vector spaces, they have the “same” sets and the “same” structure, making them mathematically the same (different only possibly in the names or characterizations of the vectors). This notion of isomorphism as sameness pervades mathematics. We shall see it again later in the context of geometric structures.

One important feature of one-to-one functions is that they admit an inverse function from the range of the original function to the domain of the original function. In the case of a one-to-one, onto function $T : V \rightarrow W$, the inverse $T^{-1} : W \rightarrow V$ is defined on all of W , where $T \circ T^{-1} = \text{Id}_W$ and $T^{-1} \circ T = \text{Id}_V$. We summarize this in the following theorem.

Theorem 2.5.4. *Let $T : V \rightarrow W$ be a linear isomorphism. Then there is a unique linear isomorphism $T^{-1} : W \rightarrow V$ such that $T \circ T^{-1} = \text{Id}_W$ and $T^{-1} \circ T = \text{Id}_V$.*

Proof. Exercise. The most important fact to be proved is that the inverse of a linear transformation, which exists purely on set-theoretic grounds, is in fact a linear transformation. $\square$

We conclude with one sense in which isomorphic vector spaces have the same structure. We will see others throughout the chapter.

Theorem 2.5.5. *Suppose that V and W are finite-dimensional vector spaces, and suppose there is a linear isomorphism $T : V \rightarrow W$. Then $\dim V = \dim W$.*

Proof. If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V , the reader can show that

$$\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)\}$$

is a basis for W and that $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are distinct. $\square$

2.6 Constructing Linear Transformations

In this section we present two theorems that together generate a wealth of examples of linear transformations. In fact, for pairs of finite-dimensional vector spaces, these give a method that generates all possible linear transformations between them.

The first theorem should be familiar to readers who have been exposed to a first course in linear algebra. It establishes a basic correspondence between $m \times n$ matrices and linear transformations between Euclidean spaces.

Theorem 2.6.1. *Every linear transformation $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ can be expressed in terms of matrix multiplication in the following sense: There exists an $m \times n$ matrix $A_T = [T]$ such that $T(\mathbf{x}) = A_T \mathbf{x}$, where $\mathbf{x} \in \mathbf{R}^n$ is understood as a column vector. Conversely, every $m \times n$ matrix A gives rise to a linear transformation $T_A : \mathbf{R}^n \rightarrow \mathbf{R}^m$ by defining $T_A(\mathbf{x}) = A\mathbf{x}$.*

For example, the linear transformation $T : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ given by

$$T(x, y, z) = (2x + y - z, x + 3z)$$

can be expressed as $T(\mathbf{x}) = A_T \mathbf{x}$, where

$$A_T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}.$$

The proof of the first, main, statement of this theorem will emerge in the course of this section. The second statement is a consequence of the basic properties of matrix multiplication.

The most important of several basic features of the correspondence between matrices and linear transformations is that matrix multiplication corresponds to composition of linear transformations:

$$[S \circ T] = [S][T],$$

and as a result, if $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a linear isomorphism, then

$$[T^{-1}] = [T]^{-1}.$$

We also note from the outset that the matrix representation of a linear transformation is not unique; it will be seen to depend on a choice of basis in both the domain and codomain. We return to this point later in the section.

The second theorem on its face gives a far more general method for constructing linear transformations, in the sense that it applies to the setting of linear transformations between arbitrary finite-dimensional vector spaces, not just between Euclidean spaces. It says that a linear transformation is uniquely defined by its action on a basis. The reader should compare this theorem to Theorem 2.5.5.

Theorem 2.6.2. *Let V be a finite-dimensional vector space with basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Let W be a vector space, and let $\mathbf{w}_1, \dots, \mathbf{w}_n$ be any n vectors in W , not necessarily distinct. Then there is a unique linear transformation $T : V \rightarrow W$ such that $T(\mathbf{e}_i) = \mathbf{w}_i$ for $i = 1, \dots, n$.*

If the set $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ is in fact a basis for W , then T is a linear isomorphism.

Proof. By Theorem 2.4.9, every element $\mathbf{v} \in V$ can be uniquely written as a linear combination of elements of the basis B , which is to say there exist unique scalars $v_1, \dots, v_n$ such that $\mathbf{v} = v_1\mathbf{e}_1 + \dots + v_n\mathbf{e}_n$. Then define $T(\mathbf{v}) = v_1\mathbf{w}_1 + \dots + v_n\mathbf{w}_n$; the reader may check that T so defined is in fact a linear transformation.

If $B' = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ is a basis, then T so defined is one-to-one and onto. Both statements follow from the fact that if $\mathbf{w} \in W$ is written according to B' as $\mathbf{w} = s_1\mathbf{w}_1 + \dots + s_n\mathbf{w}_n$, then the vector $\mathbf{v} = s_1\mathbf{e}_1 + \dots + s_n\mathbf{e}_n$ can be shown to be the unique vector such that $T(\mathbf{v}) = \mathbf{w}$. $\square$

Example 2.6.3. Consider the basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathbf{R}^2$, where $\mathbf{e}_1 = (-1, 1)$ and $\mathbf{e}_2 = (2, 1)$. Define a linear transformation $T : \mathbf{R}^2 \rightarrow \mathbf{R}^4$ in the manner of Theorem 2.6.2 by setting $T(\mathbf{e}_1) = (1, 2, 3, 4)$ and $T(\mathbf{e}_2) = (-2, -4, -6, -8)$. More explicitly, let $\mathbf{v} = (v_1, v_2)$ be an arbitrary vector in $\mathbf{R}^2$. Writing $\mathbf{v} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$ uniquely as a linear combination of $\mathbf{e}_1, \mathbf{e}_2$ amounts to solving the system

$$\begin{cases} c_1(-1) + c_2(2) = v_1, \\ c_1(1) + c_2(1) = v_2, \end{cases}$$

to obtain $c_1 = \frac{1}{3}(-v_1 + 2v_2)$ and $c_2 = \frac{1}{3}(v_1 + v_2)$. Hence

$$\begin{aligned} T(\mathbf{v}) &= c_1T(\mathbf{e}_1) + c_2T(\mathbf{e}_2) \\ &= \frac{1}{3}(-v_1 + 2v_2)(1, 2, 3, 4) + \frac{1}{3}(v_1 + v_2)(-2, -4, -6, -8) \\ &= (-v_1, -2v_1, -3v_1, -4v_1). \end{aligned}$$

As a matrix, $T(\mathbf{v}) = A_T \mathbf{v}$, where

$$A_T = \begin{bmatrix} -1 & 0 \\ -2 & 0 \\ -3 & 0 \\ -4 & 0 \end{bmatrix}.$$

Note that the method of Theorem 2.6.2 illustrated in Example 2.6.3 gives rise to a general method of representing linear transformations between finite-dimensional vector spaces as matrices, as we did in the case of Euclidean spaces in Theorem 2.6.1.

Suppose we are given a linear transformation $T : V \rightarrow W$ as well as a basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V and a basis $B' = \{\mathbf{e}'_1, \dots, \mathbf{e}'_m\}$ for W . Each of the vectors $T(\mathbf{e}_i)$ can be written uniquely as a linear combination of elements of B' :

$$\begin{aligned} T(\mathbf{e}_1) &= a_{11}\mathbf{e}'_1 + \dots + a_{1m}\mathbf{e}'_m, \\ &\vdots \\ T(\mathbf{e}_n) &= a_{n1}\mathbf{e}'_1 + \dots + a_{nm}\mathbf{e}'_m. \end{aligned} \tag{2.1}$$

It is a straightforward exercise to verify that if $\mathbf{x} \in V$, where $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n$, and if $\mathbf{y} = T(\mathbf{x}) = y_1\mathbf{e}'_1 + \dots + y_m\mathbf{e}'_m$, then $\mathbf{y} = A\mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_m)$, and $A = [a_{ij}]$ with entries a_{ij} given in (2.1) above. Then A is called *the matrix of T relative to the bases B, B'* and will be denoted by $A = [T]_{B', B}$. The reader may verify that if $T : V \rightarrow W$ is a linear isomorphism, then $[T^{-1}]_{B, B'} = [T]_{B', B}^{-1}$.

Example 2.6.4. For the linear transformation $T : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ defined by $T(x, y, z) = (2x + y - z, x + 3z)$, let $B = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, where $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (1, 1, 0)$, and $\mathbf{e}_3 = (1, 1, 1)$, and let $B' = \{\mathbf{e}'_1, \mathbf{e}'_2\}$, where $\mathbf{e}'_1 = (-1, 1)$ and $\mathbf{e}'_2 = (2, 1)$. It is an exercise to check that B is a basis for $\mathbf{R}^3$ and B' is a basis for $\mathbf{R}^2$. We now compute $[T]_{B', B}$.

Note that for a general vector $\mathbf{w} = (w_1, w_2) \in \mathbf{R}^2$, writing $\mathbf{w}$ as a linear combination of B' , $\mathbf{w} = c_1\mathbf{e}'_1 + c_2\mathbf{e}'_2$ amounts to solving the system

$$\begin{cases} c_1(-1) + c_2(2) = w_1, \\ c_1(1) + c_2(1) = w_2. \end{cases}$$

This is precisely the calculation we performed in Example 2.6.3. However, to illustrate an efficient general method for finding the matrix representation of a linear transformation, let us solve this system simultaneously for $T(\mathbf{e}_1) = (2, 1)$, $T(\mathbf{e}_2) = (3, 1)$, and $T(\mathbf{e}_3) = (2, 4)$ by Gaussian elimination of the matrix

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

yielding

$$\begin{bmatrix} 1 & 0 & \vdots & 0 & -1/3 & 2 \\ 0 & 1 & \vdots & 1 & 4/3 & 2 \end{bmatrix}.$$

In other words, $T(\mathbf{e}_1) = 0\mathbf{e}'_1 + 1\mathbf{e}'_2$, $T(\mathbf{e}_2) = (-1/3)\mathbf{e}'_1 + (4/3)\mathbf{e}'_2$, and $T(\mathbf{e}_3) = 2\mathbf{e}'_1 + 2\mathbf{e}'_2$. Hence the matrix for T relative to the bases B, B' is

$$[T]_{B',B} = \begin{bmatrix} 0 & -1/3 & 2 \\ 1 & 4/3 & 2 \end{bmatrix}.$$

A number of conclusions can be drawn from this example. First, comparing the matrix for T in Example 2.6.4 with the matrix for the same T given following Theorem 2.6.1 illustrates the dependence of the matrix for T on the bases involved. In particular, it illustrates the comment immediately following Theorem 2.6.1, that the matrix representation of a linear transformation is not unique.

Second, Theorem 2.6.2 in fact provides a proof for Theorem 2.6.1. The standard matrix representation of a linear transformation $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is obtained by applying Theorem 2.6.2 using the standard bases for $\mathbf{R}^n$ and $\mathbf{R}^m$.

Recall that Theorem 2.5.5 shows that if two vector spaces are isomorphic, then they have the same dimension. Theorem 2.6.2 shows that the converse is also true, again only for finite-dimensional vector spaces.

Corollary 2.6.5. *Let V and W be vector spaces with the same finite dimension n . Then V and W are isomorphic.*

The above theorem justifies the statement following Example 2.1.3: Every n -dimensional vector space is isomorphic to the familiar example $\mathbf{R}^n$.

We remind the reader of the following basic result from matrix algebra, expressed in these new terms.

Theorem 2.6.6. *Let $T : V \rightarrow W$ be a linear transformation between vector spaces of the same finite dimension. Then T is a linear isomorphism if and only if $\det(A) \neq 0$, where $A = [T]_{B',B}$ is the matrix representation of T relative to any bases B of V and B' of W .*

Finally, we recall that for linear transformations $T : V \rightarrow V$, the determinant of T is independent of the basis in the following sense.

Theorem 2.6.7. *Let V be a finite-dimensional vector space, and let $T : V \rightarrow V$ be a linear transformation. Then for any two bases B_1, B_2 of V , we have*

$$\det [T]_{B_1,B_1} = \det [T]_{B_2,B_2}.$$

Proof. The result is a consequence of the fact that

$$[T]_{B_2, B_2} = [\text{Id}]_{B_2, B_1} [T]_{B_1, B_1} [\text{Id}]_{B_1, B_2},$$

and that $[\text{Id}]_{B_2, B_1} = [\text{Id}]_{B_1, B_2}^{-1}$, where $\text{Id} : V \rightarrow V$ is the identity transformation. $\square$

For this reason, we refer to the *determinant of the linear transformation* $T : V \rightarrow V$ and write $\det(T)$ to be the value of $\det(A)$, where $A = [T]_{B, B}$ for any basis B of V .

2.7 Constructing Subspaces II: Subspaces and Linear Transformations

There are several subspaces naturally associated with a linear transformation $T : V \rightarrow W$.

Definition 2.7.1. The *kernel* of a linear transformation of $T : V \rightarrow W$, denoted by $\ker(T)$, is defined to be the set

$$\ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\} \subset V.$$

Definition 2.7.2. The *range* of a linear transformation $T : V \rightarrow W$, denoted by $R(T)$, is defined to be the set

$$R(T) = \{\mathbf{w} \in W \mid \text{there is } \mathbf{v} \in V \text{ such that } T(\mathbf{v}) = \mathbf{w}\} \subset W.$$

Theorem 2.7.3. Let $T : V \rightarrow W$ be a linear transformation. Then $\ker(T)$ and $R(T)$ are subspaces of V and W respectively.

Proof. Exercise. $\square$

It is a standard exercise in a first course in linear algebra to find a basis for the kernel of a given linear transformation.

Example 2.7.4. Let $T : \mathbf{R}^3 \rightarrow \mathbf{R}$ be given by $T(x, y, z) = ax + by + cz$, where a, b, c are not all zero. Then

$$\ker(T) = \{(x, y, z) \mid ax + by + cz = 0\},$$

the subspace we encountered in Example 2.2.6. Suppose that $a \neq 0$. Then we can also write

$$\ker(T) = \{(-bs - ct, as, at) \mid s, t \in \mathbf{R}\}.$$

The set $S = \{\mathbf{b}_1, \mathbf{b}_2\}$, where $\mathbf{b}_1 = (-b, a, 0)$ and $\mathbf{b}_2 = (-c, 0, a)$, is a basis for $\ker(T)$, and so $\dim(\ker(T)) = 2$.

For a linear transformation $T : V \rightarrow W$, the subspaces $\ker(T)$ and $R(T)$ are closely related to basic properties of T as a function. For example, by definition, T is onto if $R(T) = W$.

The following example highlights what might be thought of as the prototypical onto and one-to-one linear transformations.

Example 2.7.5. Consider Euclidean spaces $\mathbf{R}^n, \mathbf{R}^m$ with $m < n$.

The projection map $\text{Pr} : \mathbf{R}^n \rightarrow \mathbf{R}^m$, given by

$$\text{Pr}(x_1, \dots, x_n) = (x_1, \dots, x_m),$$

is a linear transformation that is onto but not one-to-one.

The inclusion map $\text{In} : \mathbf{R}^m \rightarrow \mathbf{R}^n$ given by

$$\text{In}(x_1, \dots, x_m) = (x_1, \dots, x_m, 0, \dots, 0)$$

is a linear transformation that is one-to-one but not onto.

We illustrate a powerful characterization of one-to-one linear transformations that has no parallel for general functions.

Theorem 2.7.6. *A linear transformation $T : V \rightarrow W$ is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$.*

Proof. Exercise. □

There is an important relationship between the dimensions of the kernel and range of a given linear transformation.

Theorem 2.7.7. *Let V be a finite-dimensional vector space, W another vector space, and $T : V \rightarrow W$ a linear transformation. Then*

$$\dim(R(T)) + \dim(\ker(T)) = \dim(V).$$

Proof. The proof involves a standard technique in linear algebra known as *completing a basis*. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . Then $\{T(\mathbf{e}_1), \dots, T(\mathbf{e}_n)\}$ spans $R(T)$, and so $\dim(R(T)) = r \leq n$. We will assume for the remainder of the proof that $1 \leq r < n$, and leave the special cases $r = 0, n$ to the reader.

Let $\{\mathbf{f}'_1, \dots, \mathbf{f}'_r\}$ be a basis for $R(T)$. There is a corresponding set

$$E = \{\mathbf{e}'_1, \dots, \mathbf{e}'_r\} \subset V$$

such that $T(\mathbf{e}'_i) = \mathbf{f}'_i$ for each $i = 1, \dots, r$. The reader may check that E must be linearly independent.

Since $r < n$, we now “complete” E by adding $(n - r)$ elements $\mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n$ to E in such a way that first, the new set

$$E' = \{\mathbf{e}'_1, \dots, \mathbf{e}'_r, \mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n\}$$

forms a basis for V , and second, that the set $\{\mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n\}$ forms a basis for $\ker(T)$. We illustrate the first step of this process. Choose $\mathbf{b}_{r+1} \notin \text{Span}\{\mathbf{e}'_1, \dots, \mathbf{e}'_r\}$. Since $\{\mathbf{f}'_1, \dots, \mathbf{f}'_r\}$ is a basis for $R(T)$, write $T(\mathbf{b}_{r+1}) = \sum a_i \mathbf{f}'_i$ and define $\mathbf{e}'_{r+1} = \mathbf{b}_{r+1} - \sum a_i \mathbf{e}'_i$. Then the reader can verify that $\mathbf{e}'_{r+1}$ is still independent of E and that $T(\mathbf{e}'_{r+1}) = \mathbf{0}$, so $\mathbf{e}'_{r+1} \in \ker(T)$. Form the new set $E \cup \{\mathbf{e}'_{r+1}\}$. Repeated application of this process yields E' . We leave to the reader the verification that $\{\mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n\}$ forms a basis for $\ker(T)$. $\square$

We will frequently refer to the dimension of the range of a linear transformation.

Definition 2.7.8. The *rank of a linear transformation* $T : V \rightarrow W$ is the dimension of $R(T)$.

The reader can verify that this definition of rank matches exactly that of the rank of any matrix representative of T relative to bases for V and W .

The following example illustrates both the statement of Theorem 2.7.7 and the notion of completing a basis used in the theorem's proof.

Example 2.7.9. Let V be a vector space with dimension n and let W be a subspace of V with dimension r , with $1 \leq r < n$. Let $B' = \{\mathbf{e}_1, \dots, \mathbf{e}_r\}$ be a basis for W . Complete this basis to a basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$ for V .

We define a linear transformation $\text{Pr}_{B',B} : V \rightarrow V$ as follows: For every vector $\mathbf{v} \in V$, there are unique scalars $v_1, \dots, v_n$ such that $\mathbf{v} = v_1 \mathbf{e}_1 + \dots + v_n \mathbf{e}_n$. Define

$$\text{Pr}_{B',B}(\mathbf{v}) = v_1 \mathbf{e}_1 + \dots + v_r \mathbf{e}_r.$$

We leave it as an exercise to show that $\text{Pr}_{B',B}$ is a linear transformation. Clearly $W = R(\text{Pr}_{B',B})$, and so $\dim(R(\text{Pr}_{B',B})) = r$. Theorem 2.7.7 then implies that $\dim(\ker(\text{Pr}_{B',B})) = n - r$, a fact that is also seen by noting that $\{\mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$ is a basis for $\ker(\text{Pr}_{B',B})$.

As the notation implies, the map $\text{Pr}_{B',B}$ depends on the choices of bases B' and B , not just on the subspace W .

Note that this example generalizes the projection defined in Example 2.7.5 above.

Theorem 2.7.7 has a number of important corollaries for finite-dimensional vector spaces. We leave the proofs to the reader.

Corollary 2.7.10. Let $T : V \rightarrow W$ be a linear transformation between finite-dimensional vector spaces. If T is one-to-one, then $\dim(V) \leq \dim(W)$. If T is onto, then $\dim(V) \geq \dim(W)$.

Note that this corollary gives another proof of Theorem 2.5.5.

As an application of the above results, we make note of the following corollary, which has no parallel in the nonlinear context.

Corollary 2.7.11. Let $T : V \rightarrow W$ be a linear transformation between vector spaces of the same finite dimension. Then T is one-to-one if and only if T is onto.

2.8 The Dual of a Vector Space, Forms, and Pullbacks

This section, while fundamental to linear algebra, is not generally presented in a first course on linear algebra. However, it is the algebraic foundation for the basic objects of differential geometry, differential forms, and tensors. For that reason, we will be more explicit with our proofs and explanations.

Starting with a vector space V , we will construct a new vector space V^* . Further, given vector spaces V and W along with a linear transformation $\Psi : V \rightarrow W$, we will construct a new linear transformation $\Psi^* : W^* \rightarrow V^*$ associated to Ψ .

Let V be a vector space. Define the set V^* to be the set of all linear transformations from V to $\mathbf{R}$:

$$V^* = \{T : V \rightarrow \mathbf{R} \mid T \text{ is a linear transformation}\}.$$

Note that an element $T \in V^*$ is a *function*. Define the operations of addition and scalar multiplication on V^* pointwise in the manner of Example 2.1.4. In other words, for $T_1, T_2 \in V^*$, define $T_1 + T_2 \in V^*$ by $(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$ for all $\mathbf{v} \in V$, and for $s \in \mathbf{R}$ and $T \in V^*$, define $sT \in V^*$ by $(sT)(\mathbf{v}) = sT(\mathbf{v})$ for all $\mathbf{v} \in V$.

Theorem 2.8.1. *The set V^* , equipped with the operations of pointwise addition and scalar multiplication, is a vector space.*

Proof. The main item requiring proof is to demonstrate the closure axioms. Suppose $T_1, T_2 \in V^*$. Then for every $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have

$$\begin{aligned} (T_1 + T_2)(\mathbf{v}_1 + \mathbf{v}_2) &= T_1(\mathbf{v}_1 + \mathbf{v}_2) + T_2(\mathbf{v}_1 + \mathbf{v}_2) \\ &= (T_1(\mathbf{v}_1) + T_1(\mathbf{v}_2)) + (T_2(\mathbf{v}_1) + T_2(\mathbf{v}_2)) \\ &= (T_1 + T_2)(\mathbf{v}_1) + (T_1 + T_2)(\mathbf{v}_2). \end{aligned}$$

We have relied on the linearity of T_1 and T_2 in the second equality. The proof that $(T_1 + T_2)(c\mathbf{v}) = c(T_1 + T_2)(\mathbf{v})$ for every $c \in \mathbf{R}$ and $\mathbf{v} \in V$ is identical. Hence $T_1 + T_2 \in V^*$.

The fact that sT_1 is also linear for every $s \in \mathbf{R}$ is proved similarly. Note that the zero “vector” $O \in V^*$ is defined by $O(\mathbf{v}) = 0$ for all $\mathbf{v} \in V$. $\square$

The space V^* is called the *dual vector space* to V . Elements of V^* are variously called *dual vectors*, *linear one-forms*, or *covectors*.

The proof of the following theorem, important in its own right, includes a construction that we will rely on often: the basis dual to a given basis.

Theorem 2.8.2. *Suppose that V is a finite-dimensional vector space. Then*

$$\dim(V) = \dim(V^*).$$

Proof. Let $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . We will construct a basis of V^* having n covectors.

For $i = 1, \dots, n$, define covectors $\varepsilon_i \in V^*$ by how they act on the basis B according to Theorem 2.6.2: $\varepsilon_i(\mathbf{e}_i) = 1$ and $\varepsilon_i(\mathbf{e}_j) = 0$ for $j \neq i$. In other words, for $\mathbf{v} = v_1\mathbf{e}_1 + \dots + v_n\mathbf{e}_n$,

$$\varepsilon_i(\mathbf{v}) = v_i.$$

We show that $B^* = \{\varepsilon_1, \dots, \varepsilon_n\}$ is a basis for V^* . To show that B^* is linearly independent, suppose that $c_1\varepsilon_1 + \dots + c_n\varepsilon_n = O$ (an equality of linear transformations). This means that for all $\mathbf{v} \in V$,

$$c_1\varepsilon_1(\mathbf{v}) + \dots + c_n\varepsilon_n(\mathbf{v}) = O(\mathbf{v}) = 0.$$

In particular, for each $i = 1, \dots, n$, setting $\mathbf{v} = \mathbf{e}_i$ gives

$$\begin{aligned} 0 &= c_1\varepsilon_1(\mathbf{e}_i) + \dots + c_n\varepsilon_n(\mathbf{e}_i) \\ &= c_i. \end{aligned}$$

Hence B^* is a linearly independent set.

To show that B^* spans V^* , choose an arbitrary $T \in V^*$, i.e., $T : V \rightarrow \mathbf{R}$ is a linear transformation. We need to find scalars $c_1, \dots, c_n$ such that $T = c_1\varepsilon_1 + \dots + c_n\varepsilon_n$. Following the idea of the preceding argument for linear independence, define $c_i = T(\mathbf{e}_i)$.

We need to show that for all $\mathbf{v} \in V$,

$$T(\mathbf{v}) = (c_1\varepsilon_1 + \dots + c_n\varepsilon_n)(\mathbf{v}).$$

Let $\mathbf{v} = v_1\mathbf{e}_1 + \dots + v_n\mathbf{e}_n$. On the one hand,

$$\begin{aligned} T(\mathbf{v}) &= T(v_1\mathbf{e}_1 + \dots + v_n\mathbf{e}_i) \\ &= v_1T(\mathbf{e}_1) + \dots + v_nT(\mathbf{e}_n) \\ &= v_1c_1 + \dots + v_nc_n. \end{aligned}$$

On the other hand,

$$\begin{aligned} (c_1\varepsilon_1 + \dots + c_n\varepsilon_n)(\mathbf{v}) &= c_1\varepsilon_1(\mathbf{v}) + \dots + c_n\varepsilon_n(\mathbf{v}) \\ &= c_1v_1 + \dots + c_nv_n. \end{aligned}$$

Hence $T = c_1\varepsilon_1 + \dots + c_n\varepsilon_n$, and B^* spans V^* . □

Definition 2.8.3. Let $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis for V . The basis $B^* = \{\varepsilon_1, \dots, \varepsilon_n\}$ for V^* , where $\varepsilon_i : V \rightarrow \mathbf{R}$ are linear transformations defined by their action on the basis vectors as

$$\varepsilon_i(\mathbf{e}_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

is called the *basis of V^* dual to the basis B* .

Example 2.8.4. Let $B_0 = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be the standard basis for $\mathbf{R}^n$, i.e.,

$$\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0),$$

with 1 in the i th component (see Example 2.4.6). The basis $B_0^* = \{\varepsilon_1, \dots, \varepsilon_n\}$ dual to B_0 is known as the *standard basis for $(\mathbf{R}^n)^*$* . Note that if $\mathbf{v} = (v_1, \dots, v_n)$, then $\varepsilon_i(\mathbf{v}) = v_i$. In other words, in the language of Example 2.7.5, ε_i is the projection onto the i th component.

We note that Theorem 2.6.1 gives a standard method of writing a linear transformation $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ as an $m \times n$ matrix. Linear one-forms $T \in (\mathbf{R}^n)^*$, $T : \mathbf{R}^n \rightarrow \mathbf{R}$, are no exception. In this way, elements of $(\mathbf{R}^n)^*$ can be thought of as $1 \times n$ matrices, i.e., as row vectors. For example, the standard basis B_0^* in this notation would appear as

$$[\varepsilon_1] = [1 \ 0 \ \cdots \ 0],$$

$$\vdots$$

$$[\varepsilon_n] = [0 \ 0 \ \cdots \ 1].$$

We now apply the “dual” construction to linear transformations between vector spaces V and W . For a linear transformation $\Psi : V \rightarrow W$, we will construct a new linear transformation

$$\Psi^* : W^* \rightarrow V^*.$$

(Note that this construction “reverses the arrow” of the transformation Ψ .)

Take an element of the domain $T \in W^*$, i.e., $T : W \rightarrow \mathbf{R}$ is a linear transformation. We wish to assign to T a linear transformation $S = \Psi^*(T) \in V^*$. In other words, given $T \in W^*$, we want to be able to describe a map $S : V \rightarrow \mathbf{R}$, $S(\mathbf{v}) = (\Psi^*(T))(\mathbf{v})$ for $\mathbf{v} \in V$, in such a way that S has the properties of a linear transformation.

Theorem 2.8.5. *Let $\Psi : V \rightarrow W$ be a linear transformation and let $\Psi^* : W^* \rightarrow V^*$ be given by*

$$(\Psi^*(T))(\mathbf{v}) = T(\Psi(\mathbf{v}))$$

for all $T \in W^*$ and $\mathbf{v} \in V$. Then Ψ^* is a linear transformation.

The transformation $\Psi^* : W^* \rightarrow V^*$ so defined is called the *pullback map* induced by Ψ , and $\Psi^*(T)$ is called the *pullback of T by Ψ* .

Proof. The first point to be verified is that for a fixed $T \in W^*$, we have in fact $\Psi^*(T) \in V^*$. In other words, we need to show that if $T : W \rightarrow \mathbf{R}$ is a linear

transformation, then $\Psi^*(T) : V \rightarrow \mathbf{R}$ is a linear transformation. For $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have

$$\begin{aligned} (\Psi^*(T))(\mathbf{v}_1 + \mathbf{v}_2) &= T(\Psi(\mathbf{v}_1 + \mathbf{v}_2)) \\ &= T(\Psi(\mathbf{v}_1) + \Psi(\mathbf{v}_2)) \quad \text{since } \Psi \text{ is linear} \\ &= T(\Psi(\mathbf{v}_1)) + T(\Psi(\mathbf{v}_2)) \quad \text{since } T \text{ is linear} \\ &= (\Psi^*(T))(\mathbf{v}_1) + (\Psi^*(T))(\mathbf{v}_2). \end{aligned}$$

The proof that $(\Psi^*(T))(s\mathbf{v}) = s(\Psi^*(T))(\mathbf{v})$ for a fixed T and for any vector $\mathbf{v} \in V$ and scalar $s \in \mathbf{R}$ is similar.

To prove linearity of Ψ^* itself, suppose that $s \in \mathbf{R}$ and $T \in W^*$. Then for all $\mathbf{v} \in V$,

$$\begin{aligned} (\Psi^*(sT))(\mathbf{v}) &= (sT)(\Psi(\mathbf{v})) \\ &= sT(\Psi(\mathbf{v})) \\ &= s((\Psi^*(T))(\mathbf{v})), \end{aligned}$$

and so $\Psi^*(sT) = s\Psi^*(T)$.

We leave as an exercise the verification that for all $T_1, T_2 \in W^*$, $\Psi^*(T_1 + T_2) = \Psi^*(T_1) + \Psi^*(T_2)$. $\square$

Note that $\Psi^*(T) = T \circ \Psi$. It is worth mentioning that the definition of the pullback in Theorem 2.8.5 is the sort of “canonical” construction typical of abstract algebra. It can be expressed by the diagram

$$\begin{array}{ccc} V & \xrightarrow{\Psi} & W \\ & \searrow \Psi^*T & \downarrow T \\ & & \mathbf{R} \end{array}$$

Example 2.8.6 (The matrix form of a pullback). Let $\Psi : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ be given by $\Psi(x, y, z) = (2x + y - z, x + 3z)$ and let $T \in (\mathbf{R}^2)^*$ be given by $T(u, v) = u - 5v$. Then $\Psi^*(T) \in (\mathbf{R}^3)^*$ is given by

$$\begin{aligned} (\Psi^*T)(x, y, z) &= T(\Psi(x, y, z)) \\ &= T(2x + y - z, x + 3z) \\ &= (2x + y - z) - 5(x + 3z) \\ &= -3x + y - 16z. \end{aligned}$$

In the standard matrix representation of Theorem 2.6.1, we have

$$[\Psi] = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix}, [T] = \begin{bmatrix} 1 & -5 \end{bmatrix} \text{ and } [\Psi^*T] = \begin{bmatrix} -3 & 1 & -16 \end{bmatrix} = [T][\Psi].$$

Thus the pullback operation by Ψ on linear one-forms corresponds to matrix multiplication of a given row vector *on the right* by the matrix of Ψ .

This fact may seem strange to the reader who has become accustomed to linear transformations represented as matrices acting by multiplication on the left. It reflects the fact that all the calculations in the preceding paragraph were carried out by relying on the standard bases in $\mathbf{R}^n$ and $\mathbf{R}^m$ as opposed to the *dual bases* for $(\mathbf{R}^n)^*$ and $(\mathbf{R}^m)^*$.

Let us reconsider these calculations, this time using the dual basis from Example 2.8.4 and the more general matrix representation from the method following Theorem 2.6.2. Using the standard bases $B_0 = \{\varepsilon_1, \varepsilon_2\}$ for $(\mathbf{R}^2)^*$ and $B'_0 = \{\varepsilon'_1, \varepsilon'_2, \varepsilon'_3\}$ for $(\mathbf{R}^3)^*$, where $\varepsilon_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $\varepsilon_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $\varepsilon'_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\varepsilon'_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, and $\varepsilon'_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, we note that $\Psi^*(\varepsilon_1) = 2\varepsilon'_1 + \varepsilon'_2 - \varepsilon'_3$ and $\Psi^*(\varepsilon_2) = \varepsilon'_1 + 3\varepsilon'_3$. Hence

$$[\Psi^*]_{B'_0, B_0} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ -1 & 3 \end{bmatrix} = [\Psi]^T.$$

Now, when the calculations for the pullback of $T = \varepsilon_1 - 5\varepsilon_2$ by Ψ are written using the *column vector* $[T]_{B_0} = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, we see that

$$\begin{aligned} [\Psi^*(T)]_{B'_0} &= [\Psi^*]_{B'_0, B_0} [T]_{B_0} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -5 \end{bmatrix} \\ &= \begin{bmatrix} -3 \\ 1 \\ -16 \end{bmatrix}. \end{aligned}$$

Since the pullback of a linear transformation is related to the matrix transpose, as the example illustrates, the following property is not surprising in light of the familiar property $(AB)^T = B^T A^T$.

Proposition 2.8.7. *Let $\Psi_1 : V_1 \rightarrow V_2$ and $\Psi_2 : V_2 \rightarrow V_3$ be linear transformations. Then*

$$(\Psi_2 \circ \Psi_1)^* = \Psi_1^* \circ \Psi_2^*.$$

Proof. Let $T \in V_3^*$ and choose $\mathbf{v} \in V_1$. Then on the one hand,

$$\begin{aligned}
 (\Psi_2 \circ \Psi_1)^*(T)(\mathbf{v}) &= T((\Psi_2 \circ \Psi_1)(\mathbf{v})) \\
 &= T(\Psi_2(\Psi_1(\mathbf{v}))),
 \end{aligned}$$

while on the other hand,

$$\begin{aligned}
 (\Psi_1^* \circ \Psi_2^*)(T)(\mathbf{v}) &= (\Psi_1^*(\Psi_2^*(T)))(\mathbf{v}) \\
 &= (\Psi_1^*(T \circ \Psi_2))(\mathbf{v}) \\
 &= ((T \circ \Psi_2) \circ \Psi_1)(\mathbf{v}) \\
 &= T(\Psi_2(\Psi_1(\mathbf{v}))). \quad \square
 \end{aligned}$$

The construction of the dual space V^* is a special case of a more general construction. Suppose we are given several vector spaces $V_1, \dots, V_k$. Recall (see Sect. 1.1) that the *Cartesian product of $V_1, \dots, V_k$* is the set of ordered k -tuples of vectors

$$V_1 \times \cdots \times V_k = \{(\mathbf{v}_1, \dots, \mathbf{v}_k) \mid \mathbf{v}_i \in V_i \text{ for all } i = 1, \dots, k\}.$$

The set $V = V_1 \times \cdots \times V_k$ can be given the structure of a vector space by defining vector addition and scalar multiplication componentwise.

Definition 2.8.8. Let $V_1, \dots, V_k$ and W be vector spaces. A function

$$T : V_1 \times \cdots \times V_k \rightarrow W$$

is *multilinear* if it is linear in each component:

$$\begin{aligned}
 T(\mathbf{x}_1 + \mathbf{y}, \mathbf{x}_2, \dots, \mathbf{x}_k) &= T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) + T(\mathbf{y}, \mathbf{x}_2, \dots, \mathbf{x}_k), \\
 &\vdots \\
 T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}, \mathbf{x}_k + \mathbf{y}) &= T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}, \mathbf{x}_k) + T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}, \mathbf{y}),
 \end{aligned}$$

and

$$\begin{aligned}
 T(s\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) &= sT(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k), \\
 &\vdots \\
 T(\mathbf{x}_1, \mathbf{x}_2, \dots, s\mathbf{x}_k) &= sT(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k).
 \end{aligned}$$

In the special case that all the V_i are the same and $W = \mathbf{R}$, then a multilinear function $T : V \times \cdots \times V \rightarrow \mathbf{R}$ is called a *multilinear k -form on V* .

Example 2.8.9 (The zero k -form on V). The trivial example of a k -form on a vector space V is the zero form. Define $O(\mathbf{v}_1, \dots, \mathbf{v}_k) = 0$ for all $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$. We leave it to the reader to show that O is multilinear.

Example 2.8.10 (The determinant as an n -form on $\mathbf{R}^n$). Define the map $\Omega : \mathbf{R}^n \times \cdots \times \mathbf{R}^n \rightarrow \mathbf{R}$ by

$$\Omega(\mathbf{a}_1, \dots, \mathbf{a}_n) = \det A,$$

where A is the matrix whose columns are given by the vectors $\mathbf{a}_i \in \mathbf{R}^n$ relative to the standard basis: $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$. The fact that Ω is an n -form follows from properties of the determinant of matrices.

In the work that follows, we will see several important examples of *bilinear* forms (i.e., 2-forms) on $\mathbf{R}^n$.

Example 2.8.11. Let $G_0 : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}$ be the function defined by

$$G_0(\mathbf{x}, \mathbf{y}) = x_1y_1 + \cdots + x_ny_n,$$

where $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$. Then G_0 is a bilinear form. (Readers should recognize G_0 as the familiar “dot product” of vectors in $\mathbf{R}^n$.) We leave it as an exercise to verify the linearity of G_0 in each component. Note that $G_0(\mathbf{x}, \mathbf{y}) = G_0(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$.

Example 2.8.12. Let A be an $n \times n$ matrix and let G_0 be the bilinear form on $\mathbf{R}^n$ defined in the previous example. Then define $G_A : \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}$ by $G_A(\mathbf{x}, \mathbf{y}) = G_0(A\mathbf{x}, A\mathbf{y})$. Bilinearity of G_A is a consequence of the bilinearity of G_0 and the linearity of matrix multiplication:

$$\begin{aligned} G_A(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) &= G_0(A(\mathbf{x}_1 + \mathbf{x}_2), A\mathbf{y}) \\ &= G_0(A\mathbf{x}_1 + A\mathbf{x}_2, A\mathbf{y}) \\ &= G_0(A\mathbf{x}_1, A\mathbf{y}) + G_0(A\mathbf{x}_2, A\mathbf{y}) \\ &= G_A(\mathbf{x}_1, \mathbf{y}) + G_A(\mathbf{x}_2, \mathbf{y}), \end{aligned}$$

and

$$\begin{aligned} G_A(s\mathbf{x}, \mathbf{y}) &= G_0(A(s\mathbf{x}), A\mathbf{y}) \\ &= G_0(sA\mathbf{x}, A\mathbf{y}) \\ &= sG_0(A\mathbf{x}, A\mathbf{y}) \\ &= sG_A(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Linearity in the second component can be shown in the same way, or the reader may note that $G_A(\mathbf{x}, \mathbf{y}) = G_A(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$.

Example 2.8.13. Define $S : \mathbf{R}^2 \times \mathbf{R}^2 \rightarrow \mathbf{R}$ by $S(\mathbf{x}, \mathbf{y}) = x_1y_2 - x_2y_1$, where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$. For $\mathbf{z} = (z_1, z_2)$, we have

$$\begin{aligned} S(\mathbf{x} + \mathbf{z}, \mathbf{y}) &= (x_1 + z_1)y_2 - (x_2 + z_2)y_1 \\ &= (x_1y_2 - x_2y_1) + (z_1y_2 - z_2y_1) \\ &= S(\mathbf{x}, \mathbf{y}) + S(\mathbf{z}, \mathbf{y}). \end{aligned}$$

Similarly, for every $c \in \mathbf{R}$, $S(c\mathbf{x}, \mathbf{y}) = cS(\mathbf{x}, \mathbf{y})$. Hence S is linear in the first component. Linearity in the second component then follows from the fact that $S(\mathbf{y}, \mathbf{x}) = -S(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^2$. This shows that S is a bilinear form.

Let V be a vector space of dimension n , and let $b : V \times V \rightarrow \mathbf{R}$ be a bilinear form. There is a standard way to represent b by means of an $n \times n$ matrix B , assuming that a basis is specified.

Proposition 2.8.14. *Let V be a vector space with basis $\mathcal{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and let $b : V \times V \rightarrow \mathbf{R}$ be a bilinear form. Let $B = [b_{ij}]$, where $b_{ij} = b(\mathbf{e}_i, \mathbf{e}_j)$. Then for every $\mathbf{v}, \mathbf{w} \in V$, we have*

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{w}^T B \mathbf{v},$$

where $\mathbf{v}$ and $\mathbf{w}$ are written as column vectors relative to the basis $\mathcal{E}$.

Proof. On each side, write $\mathbf{v}$ and $\mathbf{w}$ as linear combinations of the basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$. The result follows from the bilinearity of b and the linearity of matrix multiplication. $\square$

This proposition allows us to study properties of the bilinear form b by means of properties of its matrix representation B , a fact that we will use in the future. Note that the matrix representation for G_A in Example 2.8.12 relative to the standard basis for $\mathbf{R}^n$ is $A^T A$.

Finally, the pullback operation can be extended to multilinear forms. We illustrate this in the case of bilinear forms, although we will return to this topic in more generality in Chap. 4.

Definition 2.8.15. Suppose $T : V \rightarrow W$ is a linear transformation between vector spaces V and W . Let $B : W \times W \rightarrow \mathbf{R}$ be a bilinear form on W . Then the *pullback* of B by T is the bilinear form $T^*B : V \times V \rightarrow \mathbf{R}$ defined by

$$(T^*B)(\mathbf{v}_1, \mathbf{v}_2) = B(T(\mathbf{v}_1), T(\mathbf{v}_2))$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in V$.

The reader may check that T^*B so defined is in fact a bilinear form.

Proposition 2.8.16. *Let U , V , and W be vector spaces and let $T_1 : U \rightarrow V$ and $T_2 : V \rightarrow W$ be linear transformations. Let $B : W \times W \rightarrow \mathbf{R}$ be a bilinear form on W . Then*

$$(T_2 \circ T_1)^*B = T_1^*(T_2^*B).$$

Proof. The proof is a minor adaptation of the proof of Proposition 2.8.7. $\square$

2.9 Geometric Structures I: Inner Products

There are relatively few traditional geometric concepts that can be defined strictly within the axiomatic structure of vector spaces and linear transformations as presented above. One that we might define, for example, is the notion of two vectors

being parallel: For two vectors $\mathbf{v}, \mathbf{w}$ in a vector space V , we could say that $\mathbf{v}$ is *parallel* to $\mathbf{w}$ if there is a scalar $s \in \mathbf{R}$ such that $\mathbf{w} = s\mathbf{v}$.

The notion of vectors being *perpendicular* might also be defined in a crude way by means of the projection map of Example 2.7.9. Namely, a vector $\mathbf{v} \in V$ could be defined to be perpendicular to a nonzero vector $\mathbf{w} \in V$ if $\mathbf{v} \in \ker \text{Pr}_{B', B}$, where $B' = \{\mathbf{w}\}$ and B is chosen to be a basis for V whose first vector is $\mathbf{w}$. This has the distinct disadvantage of being completely dependent on the choice of basis B , in the sense that a vector might be perpendicular to another vector with respect to one basis but not with respect to another.

The reader should note that both these attempts at definitions of basic geometric notions are somewhat stilted, since we are defining the terms parallel and perpendicular without reference to the notion of angle. In fact, we have already noted that in the entire presentation of linear algebra up to this point, two notions traditionally associated with vectors—magnitude and direction—have not been defined at all. These notions do not have a natural description using the vector space axioms alone.

The notions of magnitude and direction can be described easily, however, by means of an additional mathematical structure that generalizes the familiar “dot product” (Example 2.8.11).

Definition 2.9.1. An *inner product* on a vector space V is a function $G : V \times V \rightarrow \mathbf{R}$ with the following properties:

- (I1) G is a bilinear form;
- (I2) G is symmetric: For all $\mathbf{v}, \mathbf{w} \in V$, $G(\mathbf{v}, \mathbf{w}) = G(\mathbf{w}, \mathbf{v})$;
- (I3) G is positive definite: For all $\mathbf{v} \in V$, $G(\mathbf{v}, \mathbf{v}) \geq 0$, with $G(\mathbf{v}, \mathbf{v}) = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

The pair (V, G) is called an *inner product space*.

We mention that the conditions (I1)–(I3) imply that the matrix A corresponding to the bilinear form G according to Proposition 2.8.14 must be symmetric ($A^T = A$) and positive definite ($\mathbf{x}^T A \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbf{R}^n$, with equality only when $\mathbf{x} = \mathbf{0}$).

Example 2.9.2 (The dot product). On the vector space $\mathbf{R}^n$, define $G_0(\mathbf{v}, \mathbf{w}) = v_1 w_1 + \cdots + v_n w_n$, where $\mathbf{v} = (v_1, \dots, v_n)$ and $\mathbf{w} = (w_1, \dots, w_n)$. We saw in Example 2.8.11 that G_0 is a bilinear form on $\mathbf{R}^n$. The reader may verify property (I2). To see property (I3), note that $G_0(\mathbf{v}, \mathbf{v}) = v_1^2 + \cdots + v_n^2$ is a quantity that is always nonnegative and is zero exactly when $v_1 = \cdots = v_n = 0$, i.e., when $\mathbf{v} = \mathbf{0}$.

Note that Example 2.9.2 can be generalized to any finite-dimensional vector space V . Starting with any basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for V , define $G_B(\mathbf{v}, \mathbf{w}) = v_1 w_1 + \cdots + v_n w_n$, where $\mathbf{v} = v_1 \mathbf{e}_1 + \cdots + v_n \mathbf{e}_n$ and $\mathbf{w} = w_1 \mathbf{e}_1 + \cdots + w_n \mathbf{e}_n$. This function G_B is well defined because of the unique representation of $\mathbf{v}$ and $\mathbf{w}$ in the basis B . This observation proves the following:

Theorem 2.9.3. Every finite-dimensional vector space carries an inner product structure.

Of course, there is no *unique* inner product structure on a given vector space. The geometry of an inner product space will be determined by the choice of inner product.

It is easy to construct new inner products from a given inner product structure. We illustrate one method in $\mathbf{R}^n$, starting from the standard inner product in Example 2.9.2 above.

Example 2.9.4. Let A be any invertible $n \times n$ matrix. Define a bilinear form G_A on $\mathbf{R}^n$ as in Example 2.8.12: $G_A(\mathbf{v}, \mathbf{w}) = G_0(A\mathbf{v}, A\mathbf{w})$, where G_0 is the standard inner product from Example 2.9.2. Then G_A is symmetric, since G_0 is symmetric. Similarly, $G_A(\mathbf{v}, \mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$ because of the corresponding property of G_0 . Now suppose that $G_A(\mathbf{v}, \mathbf{v}) = 0$. Since $0 = G_A(\mathbf{v}, \mathbf{v}) = G_0(A\mathbf{v}, A\mathbf{v})$, we have $A\mathbf{v} = \mathbf{0}$ by property (I3) for G_0 . Since A is invertible, $\mathbf{v} = \mathbf{0}$. This completes the verification that G_A is an inner product on $\mathbf{R}^n$.

To illustrate this construction with a simple example in $\mathbf{R}^2$, consider the matrix $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$. Then if $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$, we have

$$\begin{aligned} G_A(\mathbf{v}, \mathbf{w}) &= G_0(A\mathbf{v}, A\mathbf{w}) \\ &= G_0((2v_1 - v_2, v_1), (2w_1 - w_2, w_1)) \\ &= (2v_1 - v_2)(2w_1 - w_2) + v_1w_1 \\ &= 5v_1w_1 - 2v_1w_2 - 2v_2w_1 + v_2w_2. \end{aligned}$$

Note that the matrix representation for G_A as described in Proposition 2.8.14 is given by $A^T A$.

An inner product allows us to define geometric notions such as length, distance, magnitude, angle, and direction.

Definition 2.9.5. Let (V, G) be an inner product space. The *magnitude* (also called the *length* or the *norm*) of a vector $\mathbf{v} \in V$ is given by

$$\|\mathbf{v}\| = G(\mathbf{v}, \mathbf{v})^{1/2}.$$

The *distance* between vectors $\mathbf{v}$ and $\mathbf{w}$ is given by

$$d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\| = G(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w})^{1/2}.$$

To define the notion of direction, or angle between vectors, we first state a fundamental property of inner products.

Theorem 2.9.6 (Cauchy–Schwarz). Let (V, G) be an inner product space. Then for all $\mathbf{v}, \mathbf{w} \in V$,

$$|G(\mathbf{v}, \mathbf{w})| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|.$$

The standard proof of the Cauchy–Schwarz inequality relies on the non-intuitive observation that the discriminant of the quadratic expression (in t) $G(t\mathbf{v} + \mathbf{w}, t\mathbf{v} + \mathbf{w})$ must be nonpositive by property (I3).

Definition 2.9.7. Let (V, G) be an inner product space. For every two nonzero vectors $\mathbf{v}, \mathbf{w} \in V$, the *angle* $\angle_G(\mathbf{v}, \mathbf{w})$ between $\mathbf{v}$ and $\mathbf{w}$ is defined to be

$$\angle_G(\mathbf{v}, \mathbf{w}) = \cos^{-1} \left(\frac{G(\mathbf{v}, \mathbf{w})}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|} \right).$$

Note that this definition of angle is well defined as a result of Theorem 2.9.6.

As a consequence of this definition of angle, it is possible to define a notion of orthogonality: Two vectors $\mathbf{v}, \mathbf{w} \in V$ are *orthogonal* if $G(\mathbf{v}, \mathbf{w}) = 0$, since then $\angle_G(\mathbf{v}, \mathbf{w}) = \pi/2$. The notion of orthogonality, in turn, distinguishes “special” bases for V and provides a further method for producing new subspaces of V from a given set of vectors in V .

Theorem 2.9.8. Let (V, G) be an inner product space with $\dim(V) = n > 0$. There exists a basis $B = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ satisfying the following two properties:

- (O1) For each $i = 1, \dots, n$, $G(\mathbf{u}_i, \mathbf{u}_i) = 1$;
- (O2) For each $i \neq j$, $G(\mathbf{u}_i, \mathbf{u}_j) = 0$.

Such a basis is known as an *orthonormal basis*.

The reader is encouraged to review the proof of this theorem, which can be found in any elementary linear algebra text. It relies on an important procedure, similar in spirit to the proof of Theorem 2.7.7, known as *Gram–Schmidt orthonormalization*. Beginning with any given basis, the procedure gives a way of constructing a new basis satisfying (O1) and (O2). In the next section, we will carry out the details of an analogous procedure in the symplectic setting.

As with bases in general, there is no unique orthonormal basis for a given inner product space (V, G) .

We state without proof a kind of converse to Theorem 2.9.8. This theorem is actually a restatement of the comment following Example 2.9.2.

Theorem 2.9.9. Let V be a vector space and let $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a finite basis for V . Define a function G_B by requiring that

$$G_B(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

and extending linearly in both components in the manner of Theorem 2.6.2. Then G_B is an inner product.

For any vector $\mathbf{v}$ in an inner product space (V, G) , the set W of all vectors orthogonal to $\mathbf{v}$ can be seen to be a subspace of V . One could appeal directly to Theorem 2.2.2 (since $\mathbf{0} \in W$), or one could note that W is the kernel of the linear

transformation $i_{\mathbf{v}} : V \rightarrow \mathbf{R}$ given by $i_{\mathbf{v}}(\mathbf{w}) = G(\mathbf{v}, \mathbf{w})$. More generally, we have the following:

Theorem 2.9.10. *Let S be any nonempty set of vectors in an inner product space (V, G) . The set $S^\perp$ defined by*

$$S^\perp = \{\mathbf{w} \in V \mid \text{For all } \mathbf{v} \in S, G(\mathbf{v}, \mathbf{w}) = 0\}$$

is a subspace of V .

Proof. Exercise. □

The set $S^\perp$ is called the *orthogonal complement* to S .

Example 2.9.11. Let $S = \{\mathbf{v}\} \subset \mathbf{R}^3$, where $\mathbf{v} = (a, b, c) \neq \mathbf{0}$. Let G_0 be the standard inner product on $\mathbf{R}^3$ (see Example 2.9.2). Then

$$S^\perp = \{(x, y, z) \mid ax + by + cz = 0\}.$$

See Example 2.2.4.

Example 2.9.12. Let $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$ and let G_A be the inner product defined on $\mathbf{R}^2$ according to Example 2.9.4. Let $\mathbf{v} = (1, 0)$, and let $S = \{\mathbf{v}\}$. Then the reader may verify that

$$S^\perp = \{(w_1, w_2) \mid 5w_1 - 2w_2 = 0\},$$

which is spanned by the set $\{(2, 5)\}$. See Fig. 2.3.

Theorem 2.9.13. *Let (V, G) be an inner product space. Let S be a finite subset of V and let $W = \text{Span}(S)$. Then $W^\perp = S^\perp$.*

Proof. Let $S = \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$. Take a vector $\mathbf{v} \in W^\perp$, so that $G(\mathbf{w}, \mathbf{v}) = 0$ for all $\mathbf{w} \in W$. In particular, since $S \subset W$, for each $\mathbf{w}_i \in S$, $G(\mathbf{w}_i, \mathbf{v}) = 0$. So $\mathbf{v} \in S^\perp$ and $W^\perp \subset S^\perp$.

Now take a vector $\mathbf{v} \in S^\perp$. Let $\mathbf{w} \in W$, so that there are scalars $c_1, \dots, c_k$ such that $\mathbf{w} = c_1\mathbf{w}_1 + \dots + c_k\mathbf{w}_k$. Relying on the linearity of G in the first component, we obtain

$$\begin{aligned} G(\mathbf{w}, \mathbf{v}) &= G(c_1\mathbf{w}_1 + \dots + c_k\mathbf{w}_k, \mathbf{v}) \\ &= G(c_1\mathbf{w}_1, \mathbf{v}) + \dots + G(c_k\mathbf{w}_k, \mathbf{v}) \\ &= c_1G(\mathbf{w}_1, \mathbf{v}) + \dots + c_kG(\mathbf{w}_k, \mathbf{v}) \\ &= 0, \quad \text{since } \mathbf{v} \in S^\perp. \end{aligned}$$

Hence $\mathbf{v} \in W^\perp$, and so $S^\perp \subset W^\perp$.

Together, these two statements show that $W^\perp = S^\perp$. □

Corollary 2.9.14. *Let B be a basis for a subspace $W \subset V$. Then $W^\perp = B^\perp$.*

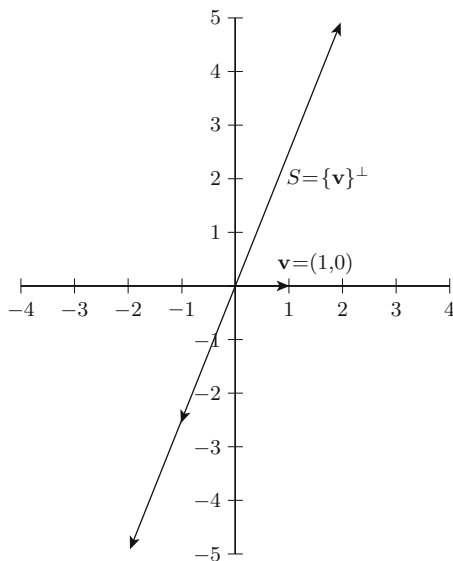


Fig. 2.3 The orthogonal complement to $\{(1, 0)\}$ relative to G in Example 2.9.12.

The following theorems discuss the relationship between a vector subspace W and its orthogonal complement $W^\perp$.

Theorem 2.9.15. *Let W be a subspace of an inner product space (V, G) . Then $W \cap W^\perp = \{0\}$.*

Proof. Exercise. □

Theorem 2.9.16. *Let W be a subspace of a finite-dimensional inner product space (V, G) . Then*

$$\dim(W) + \dim(W^\perp) = \dim(V).$$

Proof. Consider the map $S : V \rightarrow W^*$ given by $S(\mathbf{v}) = i_{\mathbf{v}}$, where for every $\mathbf{w} \in W$, $i_{\mathbf{v}}(\mathbf{w}) = G(\mathbf{v}, \mathbf{w})$. The reader can check that S is a linear transformation, that $\ker S = W^\perp$, and that S is onto. The result then follows from Theorem 2.7.7. □

We now turn our attention to linear transformations of an inner product space that preserve the additional inner product structure.

Definition 2.9.17. Let (V_1, G_1) be an inner product space. A linear transformation $T : V \rightarrow V$ is a *linear isometry* if for all $\mathbf{v}, \mathbf{w} \in V$, $G(T(\mathbf{v}), T(\mathbf{w})) = G(\mathbf{v}, \mathbf{w})$. Stated in the language of Sect. 2.8, T is a linear isometry if $T^*G = G$.

Note that a linear isometry preserves all quantities defined in terms of the inner product: distance, magnitude, the angle between vectors, and orthogonality.

The reader may verify the following property of linear isometries.

Proposition 2.9.18. *Let (V, G) be an inner product space. If T_1, T_2 are linear isometries of V , then $T_2 \circ T_1$ is also a linear isometry.*

The following theorem, which we state without proof, gives a matrix characterization of linear isometries.

Theorem 2.9.19. *Let (V, G) be a finite-dimensional inner product space with $\dim(V) = n > 0$, and let $T : V \rightarrow V$ be a linear isometry. Then the matrix representation $A = [T]$ of T relative to any orthonormal basis of V satisfies $A^T A = I_n$, where I_n is the $n \times n$ identity matrix.*

A matrix with the property that $A^T A = I_n$ is called an *orthogonal matrix*.

Corollary 2.9.20. *Let $T : V \rightarrow V$ be a linear isometry of a finite-dimensional inner product space (V, G) . Then $\det(T) = \pm 1$. In particular, T is invertible.*

Proposition 2.9.21. *Let (V, G) be a finite-dimensional inner product space, and let $T : V \rightarrow V$ be a linear isometry. Then its inverse T^{-1} is also a linear isometry.*

Proof. Assuming T is a linear isometry, apply Proposition 2.8.16 to $G = (\text{Id})^* G = (T \circ T^{-1})^* G$ and use the assumption that $T^* G = G$. $\square$

We conclude this section with an important technical theorem, a consequence of the positive definite property of inner products. Recall that V and V^* have the same dimension by Theorem 2.8.2, and so by Corollary 2.6.5, the two vector spaces are isomorphic. A choice of an inner product on V , however, induces a distinguished isomorphism between them.

Theorem 2.9.22. *Let G be an inner product on a finite-dimensional vector space V . For every $\mathbf{v} \in V$, define $i_{\mathbf{v}} \in V^*$ by $i_{\mathbf{v}}(\mathbf{w}) = G(\mathbf{v}, \mathbf{w})$ for $\mathbf{w} \in V$. Then the function*

$$\Phi : V \rightarrow V^*$$

defined by $\Phi(\mathbf{v}) = i_{\mathbf{v}}$ is a linear isomorphism.

Proof. The fact that Φ is linear is a consequence of the fact that G is bilinear. For example, for $\mathbf{v} \in V$ and $s \in \mathbf{R}$, $\Phi(s\mathbf{v}) = i_{s\mathbf{v}}$, and so for all $\mathbf{w} \in V$, $i_{s\mathbf{v}}(\mathbf{w}) = G(s\mathbf{v}, \mathbf{w}) = sG(\mathbf{v}, \mathbf{w}) = si_{\mathbf{v}}(\mathbf{w})$. Hence $i_{s\mathbf{v}} = si_{\mathbf{v}}$, and so $\Phi(s\mathbf{v}) = s\Phi(\mathbf{v})$. Likewise, $\Phi(\mathbf{v} + \mathbf{w}) = \Phi(\mathbf{v}) + \Phi(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$.

To show that Φ is one-to-one, we show that $\ker(\Phi) = \{\mathbf{0}\}$. Let $\mathbf{v} \in \ker(\Phi)$. Then $\Phi(\mathbf{v}) = 0$, i.e., $G(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in V$. In particular, $G(\mathbf{v}, \mathbf{v}) = 0$, and so $\mathbf{v} = \mathbf{0}$ by positive definiteness. Hence $\ker(\Phi) = \{\mathbf{0}\}$, and so by Theorem 2.7.6, Φ is one-to-one.

The fact that Φ is onto now follows from the fact that a one-to-one linear map between vector spaces of the same dimension must be onto (Corollary 2.7.11). However, we will show directly that Φ is onto in order to exhibit the inverse transformation $\Phi^{-1} : V^* \rightarrow V$.

Let $T \in V^*$. We need to find $\mathbf{v}_T \in V$ such that $\Phi(\mathbf{v}_T) = T$. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be an orthonormal basis for (V, G) , as guaranteed by Theorem 2.9.8. Define c_i by

$c_i = T(\mathbf{u}_i)$, and define $\mathbf{v}_T = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$. By the linearity of G in the first component, we have $\Phi(\mathbf{v}_T) = T$, or, what is the same, $\mathbf{v}_T = \Phi^{-1}(T)$. Hence Φ is onto. $\square$

The reader should notice the similarity between the construction of Φ^{-1} and the procedure outlined in the proof of Theorem 2.8.2.

The fact that the map Φ in Theorem 2.9.22 is one-to-one can be rephrased by saying that the inner product G is *nondegenerate*: If $G(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in V$, then $\mathbf{v} = \mathbf{0}$. We will encounter this condition again shortly in the symplectic setting.

2.10 Geometric Structures II: Linear Symplectic Forms

In this section, we outline the essentials of linear symplectic geometry, which will be the starting point for one of the main differential-geometric structures that we will present later in the text. The presentation here will parallel the development of inner product structures in Sect. 2.9 in order to emphasize the similarities and differences between the two structures, both of which are defined by bilinear forms. We will discuss more about the background of symplectic geometry in Chap. 7.

Unlike most of the material in this chapter so far, what follows is not generally presented in a first course in linear algebra. As in Sect. 2.8, we will be more detailed in the presentation and proof of the statements in this section.

Definition 2.10.1. A *linear symplectic form* on a vector space V is a function $\omega : V \times V \rightarrow \mathbf{R}$ satisfying the following properties:

- (S1) ω is a bilinear form on V ;
- (S2) ω is skew-symmetric: For all $\mathbf{v}, \mathbf{w} \in V$, $\omega(\mathbf{w}, \mathbf{v}) = -\omega(\mathbf{v}, \mathbf{w})$;
- (S3) ω is nondegenerate: If $\mathbf{v} \in V$ has the property that $\omega(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in V$, then $\mathbf{v} = \mathbf{0}$.

The pair (V, ω) is called a *symplectic vector space*.

Note that the main difference between (S1)–(S3) and (I1)–(I3) in Definition 2.9.1 is that a linear symplectic form is skew-symmetric, in contrast to the symmetric inner product. We can summarize properties (S1) and (S2) by saying that ω is an *alternating bilinear form* on V . We will discuss the nondegeneracy condition (S3) in more detail below. Note that in sharp contrast to inner products, $\omega(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v} \in V$ as a consequence of (S2).

Example 2.10.2. On the vector space $\mathbf{R}^2$, define $\omega_0(\mathbf{v}, \mathbf{w}) = v_1w_2 - v_2w_1$, where $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. The reader may recognize this as the determinant of the matrix whose column vectors are $\mathbf{v}, \mathbf{w}$. That observation, or direct verification, will confirm properties (S1) and (S2). To verify (S3), suppose $\mathbf{v} = (v_1, v_2)$ is such that $\omega_0(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in \mathbf{R}^2$. In particular, $0 = \omega_0(\mathbf{v}, (1, 0)) = (v_1)(0) - (1)(v_2) = -v_2$, and so $v_2 = 0$. Likewise,

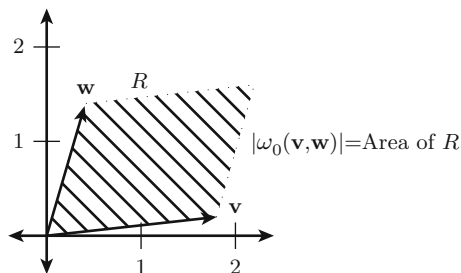


Fig. 2.4 The standard symplectic form on $\mathbf{R}^2$.

$0 = \omega_0(\mathbf{v}, (0, 1)) = v_1$. Together, these show that $\mathbf{v} = \mathbf{0}$, and so (S3) is satisfied. In this case, ω_0 measures the *oriented area* of the parallelogram defined by two vectors. See Fig. 2.4.

Example 2.10.3. Generalizing Example 2.10.2, consider the Euclidean vector space $\mathbf{R}^{2n}$. Define the function

$$\omega_0(\mathbf{v}, \mathbf{w}) = (a_1 t_1 - b_1 s_1) + \cdots + (a_n t_n - b_n s_n),$$

where $\mathbf{v} = (a_1, b_1, \dots, a_n, b_n)$ and $\mathbf{w} = (s_1, t_1, \dots, s_n, t_n)$. The verification that ω_0 is a symplectic form proceeds exactly as in Example 2.10.2; it will be called the *standard symplectic form* on $\mathbf{R}^{2n}$. Similarly, the pair $(\mathbf{R}^{2n}, \omega_0)$ will be called the *standard symplectic vector space*.

Before proceeding to more examples, we immediately prove in detail a consequence of the existence of a linear symplectic form on a finite-dimensional vector space: *The dimension of a symplectic vector space must be even.* In other words, there can be no linear symplectic form on an odd-dimensional vector space. This is the first significant difference between symmetric and skew-symmetric bilinear forms.

Theorem 2.10.4 below should be read alongside Theorems 2.9.3 and 2.9.8, which show that every finite-dimensional vector space carries an inner product, to which correspond distinguished orthonormal bases.

Theorem 2.10.4. *Let (V, ω) be a finite-dimensional symplectic vector space. Then V has a basis $\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$ with the following properties:*

(SO1) $\omega(\mathbf{e}_i, \mathbf{f}_i) = 1$ for all $i = 1, \dots, n$;

(SO2) $\omega(\mathbf{e}_i, \mathbf{e}_j) = 0$ for all $i, j = 1, \dots, n$;

(SO3) $\omega(\mathbf{f}_i, \mathbf{f}_j) = 0$ for all $i, j = 1, \dots, n$;

(SO4) $\omega(\mathbf{e}_i, \mathbf{f}_j) = 0$ for $i \neq j$.

In particular, the dimension of V is even.

Proof. The inductive process of constructing a basis with properties (SO1)–(SO4) is a modified version of the Gram–Schmidt orthonormalization process that is at the heart of the proof of Theorem 2.9.8. The reader should also compare the technique here with the technique in the proof of Theorem 2.7.7.

Choose any nonzero $\mathbf{v} \in V$ and define $\mathbf{e}_1 = \mathbf{v}$. Since $\mathbf{e}_1 \neq \mathbf{0}$, property (S3) of the symplectic form guarantees the existence of a vector $\mathbf{w}_1$ such that $\omega(\mathbf{e}_1, \mathbf{w}_1) = c_1 \neq 0$. Define $\mathbf{f}_1 = (1/c_1)\mathbf{w}_1$. By the bilinearity condition (S1), we have $\omega(\mathbf{e}_1, \mathbf{f}_1) = 1$. Note also that $\{\mathbf{e}_1, \mathbf{f}_1\}$ is a linearly independent set, since otherwise, by Theorem 2.4.3 we would have a scalar s such that $\mathbf{f}_1 = s\mathbf{e}_1$, and so by (S2), $\omega(\mathbf{e}_1, \mathbf{f}_1) = 0$, contradicting the definition of $\mathbf{f}_1$. Further, by (S2), we have $\omega(\mathbf{e}_1, \mathbf{e}_1) = \omega(\mathbf{f}_1, \mathbf{f}_1) = 0$, and so the set $\{\mathbf{e}_1, \mathbf{f}_1\}$ satisfies properties (SO1)–(SO4).

Now suppose we have constructed k pairs of linearly independent vectors

$$B_k = \{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_k, \mathbf{f}_k\}$$

satisfying (SO1)–(SO4). If $\text{Span}(B_k) = V$, then B_k is the desired basis and we are done.

If $\text{Span}(B_k) \subsetneq V$, then there is a nonzero vector $\mathbf{v}_{k+1} \notin \text{Span}(B_k)$. Define

$$\mathbf{e}_{k+1} = \mathbf{v}_{k+1} - \sum_{j=1}^k \omega(\mathbf{v}_{k+1}, \mathbf{f}_j) \mathbf{e}_j - \sum_{j=1}^k \omega(\mathbf{e}_j, \mathbf{v}_{k+1}) \mathbf{f}_j.$$

Note that $\mathbf{e}_{k+1} \neq \mathbf{0}$, since otherwise $\mathbf{v}_{k+1}$ would be a linear combination of vectors in B_k , contradicting the definition of $\mathbf{v}_{k+1}$. Since ω is bilinear, we have for each $i = 1, \dots, k$ that

$$\begin{aligned} \omega(\mathbf{e}_{k+1}, \mathbf{e}_i) &= \omega(\mathbf{v}_{k+1}, \mathbf{e}_i) - \sum_{j=1}^k \omega(\mathbf{v}_{k+1}, \mathbf{f}_j) \omega(\mathbf{e}_j, \mathbf{e}_i) - \sum_{j=1}^k \omega(\mathbf{e}_j, \mathbf{v}_{k+1}) \omega(\mathbf{f}_j, \mathbf{e}_i) \\ &= \omega(\mathbf{v}_{k+1}, \mathbf{e}_i) - \omega(\mathbf{e}_i, \mathbf{v}_{k+1}) \omega(\mathbf{f}_i, \mathbf{e}_i) \quad \text{by the inductive hypothesis} \\ &= \omega(\mathbf{v}_{k+1}, \mathbf{e}_i) - \omega(\mathbf{v}_{k+1}, \mathbf{e}_i) \quad \text{by the inductive hypothesis, (S2)} \\ &= 0, \end{aligned}$$

and similarly,

$$\begin{aligned} \omega(\mathbf{e}_{k+1}, \mathbf{f}_i) &= \omega(\mathbf{v}_{k+1}, \mathbf{f}_i) - \sum_{j=1}^k \omega(\mathbf{v}_{k+1}, \mathbf{f}_j) \omega(\mathbf{e}_j, \mathbf{f}_i) - \sum_{j=1}^k \omega(\mathbf{e}_j, \mathbf{v}_{k+1}) \omega(\mathbf{f}_j, \mathbf{f}_i) \\ &= \omega(\mathbf{v}_{k+1}, \mathbf{f}_i) - \omega(\mathbf{v}_{k+1}, \mathbf{f}_i) \\ &= 0. \end{aligned}$$

Now, by property (S3), there is a vector $\mathbf{w}_{k+1}$ such that $\omega(\mathbf{e}_{k+1}, \mathbf{w}_{k+1}) = c_{k+1} \neq 0$. Note that $\mathbf{w}_{k+1}$ cannot be in $\text{Span}(B_k)$, since otherwise, the previous calculations would imply that $\omega(\mathbf{e}_{k+1}, \mathbf{w}_{k+1}) = 0$. Define further

$$\mathbf{u}_{k+1} = \mathbf{w}_{k+1} - \sum_{j=1}^k \omega(\mathbf{w}_{k+1}, \mathbf{f}_j) \mathbf{e}_j - \sum_{j=1}^k \omega(\mathbf{e}_j, \mathbf{w}_{k+1}) \mathbf{f}_j.$$

Again we have $\mathbf{u}_{k+1} \neq \mathbf{0}$, since if $\mathbf{u}_{k+1} = \mathbf{0}$, then $\mathbf{w}_{k+1} \in \text{Span}(B_k)$.

Let $\mathbf{f}_{k+1} = (1/c_{k+1})\mathbf{u}_{k+1}$, which makes $\omega(\mathbf{e}_{k+1}, \mathbf{f}_{k+1}) = 1$. As in the preceding argument, $\mathbf{f}_{k+1} \notin \text{Span}(B_k \cup \{\mathbf{e}_{k+1}\})$. Properties (SO3) and (SO4) about $\mathbf{f}_{k+1}$ are proved in the same way as the analogous properties were derived above for $\mathbf{e}_{k+1}$. Hence, defining $B_{k+1} = B_k \cup \{\mathbf{e}_{k+1}, \mathbf{f}_{k+1}\}$, we have a linearly independent set satisfying properties (SO1)–(SO4). But since V is finite-dimensional, there must be an n such that $V = \text{Span}(B_n)$, which completes the proof. $\square$

Note that this proof relies in an essential way on the nondegeneracy condition (S3) of ω . We will see another proof of this result below in Theorem 2.10.24.

A basis of a symplectic vector space (V, ω) satisfying (SO1)–(SO4) is called a *symplectic basis* for V .

For example, for the standard symplectic space $(\mathbf{R}^{2n}, \omega_0)$, the set

$$B_0 = \{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\} \quad (2.2)$$

given by

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, 0, \dots, 0, 0), \\ \mathbf{f}_1 &= (0, 1, 0, 0, \dots, 0, 0), \\ &\vdots \\ \mathbf{e}_n &= (0, 0, \dots, 0, 0, 1, 0), \\ \mathbf{f}_n &= (0, 0, \dots, 0, 0, 0, 1), \end{aligned}$$

is a symplectic basis. We will call B_0 the *standard symplectic basis* for $(\mathbf{R}^{2n}, \omega_0)$.

As in the case of orthonormal bases for an inner product space, there is no unique symplectic basis for a given symplectic vector space (V, ω) .

The following is another indication of the importance of the nondegeneracy condition (S3) of Definition 2.10.1. It gives further evidence of why (S3) is the correct symplectic analogue to the inner product positive definite condition (I3) in Definition 2.9.1.

Theorem 2.10.5. *Let (V, ω) be a finite-dimensional symplectic vector space. The map $\Psi : V \rightarrow V^*$ defined by $\Psi(\mathbf{v}) = i_{\mathbf{v}} \in V^*$ for all $\mathbf{v} \in V$, where $i_{\mathbf{v}}(\mathbf{w}) = \omega(\mathbf{v}, \mathbf{w})$ for $\mathbf{w} \in V$, is a linear isomorphism.*

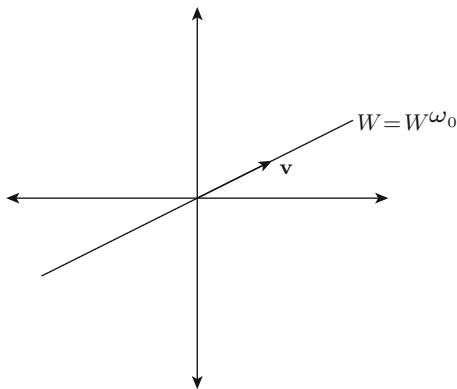


Fig. 2.5 Symplectic orthogonality in $(\mathbf{R}^2, \omega_0)$; see Example 2.10.8.

The proof exactly parallels the proof of Theorem 2.9.22.

The geometric consequences of the existence of a symplectic structure are quite different from those of an inner product structure. There is no sense, for example, of the length of a vector or of the angle between two vectors; it is enough to recall again that $\omega(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v} \in V$. There is, however, a notion corresponding to the inner product notion of orthogonality.

Definition 2.10.6. Let (V, ω) be a symplectic vector space. Two vectors $\mathbf{v}, \mathbf{w} \in V$ are called ω -orthogonal (or skew-orthogonal) if $\omega(\mathbf{v}, \mathbf{w}) = 0$.

In contrast to inner product orthogonality, as noted earlier, every vector is ω -orthogonal to itself. This consequence of the skew-symmetry of ω is in marked contrast to the symmetric case; compare, for example, to Theorem 2.9.10.

Following the idea of Sect. 2.9, we have the following:

Theorem 2.10.7. Let S be a set of vectors in a symplectic vector space (V, ω) . The set

$$S^\omega = \{\mathbf{w} \in V \mid \text{for all } \mathbf{v} \in S, \omega(\mathbf{v}, \mathbf{w}) = 0\}$$

is a subspace of V .

Proof. Exercise. □

The set S^ω is called the ω -orthogonal complement to S .

Example 2.10.8. Let $S = \{\mathbf{v}\} \subset \mathbf{R}^2$, where $\mathbf{v} = (a, b) \neq \mathbf{0}$. Let ω_0 be the standard linear symplectic form on $\mathbf{R}^2$. Then $S^{\omega_0} = \{(sa, sb) \mid s \in \mathbf{R}\}$. Indeed, if $(x, y) \in S^{\omega_0}$, then $\omega_0((a, b), (x, y)) = ay - bx = 0$. Hence $ay = bx$, and since either a or b is not 0, we can write (for example if $a \neq 0$) $y = (b/a)x$, and so $(x, y) = (x, (b/a)x) = (x/a)(a, b)$. In fact, if $W = \text{Span}(S)$, then $W = W^{\omega_0}$. See Fig. 2.5.

Example 2.10.9. Let $S = \{\mathbf{v}\} \subset (\mathbf{R}^4, \omega_0)$, where $\mathbf{v} = (1, 2, 3, 4)$. Then

$$\begin{aligned} S^{\omega_0} &= \{(x_1, y_1, x_2, y_2) \mid \omega_0((1, 2, 3, 4), (x_1, y_1, x_2, y_2)) = 0\} \\ &= \{(x_1, y_1, x_2, y_2) \mid -2x_1 + y_1 - 4x_2 + 3y_2 = 0\}. \end{aligned}$$

We can also write

$$S^{\omega_0} = \{(s, 2s + 4t - 3u, t, u) \mid s, t, u \in \mathbf{R}\},$$

so the set $\{(1, 2, 0, 0), (0, 4, 1, 0), (0, -3, 0, 1)\}$ is a basis for S^{ω_0} .

The following theorems parallel the corresponding results for inner product spaces.

Theorem 2.10.10. *Let S be a finite set of vectors in a symplectic vector space (V, ω) and let $W = \text{Span}(S)$. Then $W^\omega = S^\omega$.*

Proof. The proof follows mutatis mutandis that of Theorem 2.9.13. $\square$

Corollary 2.10.11. *Let B be a basis for a subspace $W \subset V$. Then $W^\omega = B^\omega$.*

Despite the significant differences between the notions of orthogonality and ω -orthogonality, Theorem 2.9.16 concerning the dimension of the orthogonal complement has a direct parallel in the symplectic setting.

Theorem 2.10.12. *Let W be a subspace of a finite-dimensional symplectic vector space (V, ω) . Then*

$$\dim(W) + \dim(W^\omega) = \dim(V).$$

Proof. Let $k = \dim(W)$. Note that if $k = 0$, then $W^\omega = V$, and the conclusion of the theorem trivially holds.

If $k \geq 1$, we rely on the isomorphism $\Psi : V \rightarrow V^*$ given by $\Psi(\mathbf{v}) = i_{\mathbf{v}}$ of Theorem 2.10.5, where $i_{\mathbf{v}} : V \rightarrow \mathbf{R}$ is defined to be $T_{\mathbf{v}}(\mathbf{w}) = \omega(\mathbf{v}, \mathbf{w})$. Consider the map $S : V \rightarrow W^*$ given by $S(\mathbf{v}) = \Psi(\mathbf{v})|_W$. On the one hand, S is onto. To see this, let $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_{2n}\}$ be a basis for V such that $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ is a basis for W . For any $\alpha \in W^*$, define $\tilde{\alpha} \in V^*$ to be $\tilde{\alpha}(\mathbf{v}) = c_1\alpha(\mathbf{w}_1) + \dots + c_k\alpha(\mathbf{w}_k)$, where we are writing $\mathbf{v} = c_1\mathbf{w}_1 + \dots + c_{2n}\mathbf{w}_{2n}$ according to the basis $\mathcal{B}$. The reader can check that $\tilde{\alpha} \in V^*$; in fact, $\tilde{\alpha} = \alpha \circ \text{Pr}$, where $\text{Pr} : V \rightarrow W$ is the projection onto W relative to the bases chosen above, defined in Example 2.7.9. Let $\mathbf{v}_\alpha \in V$ be such that $\Psi(\mathbf{v}_\alpha) = \tilde{\alpha}$. Then $S(\mathbf{v}_\alpha) = \alpha$, and so S is onto.

In addition, practically by definition,

$$\begin{aligned} \ker S &= \{\mathbf{v} \in V \mid (\Psi(\mathbf{v}))(\mathbf{w}) = 0 \text{ for all } \mathbf{w} \in W\} \\ &= \{\mathbf{v} \in V \mid \omega(\mathbf{v}, \mathbf{w}) = 0 \text{ for all } \mathbf{w} \in W\} \\ &= W^\omega. \end{aligned}$$

We thus rely on Theorem 2.7.7:

$$\dim(V) = \dim \ker S + \dim R(S) = \dim W^\omega + \dim W^* = \dim W^\omega + \dim W,$$

the last equality due to Theorem 2.8.2. $\square$

The analogue of Theorem 2.9.15 about the trivial intersection of W with $W^\perp$ does not always hold in a symplectic vector space. In fact, we can identify a number of possible relationships between a subspace W and its ω -orthogonal complement W^ω .

Definition 2.10.13. Let W be a subspace of a symplectic vector space (V, ω) . Then W is called

- *isotropic* if $W \subset W^\omega$;
- *coisotropic* if $W^\omega \subset W$;
- *Lagrangian* if $W = W^\omega$;
- *symplectic* if $W \cap W^\omega = \{\mathbf{0}\}$.

Proposition 2.10.14. Let W be a subspace of the symplectic vector space (V, ω) with $\dim V = 2n$. Then

- If W is isotropic, then $\dim W \leq n$.
- If W is coisotropic, then $\dim W \geq n$.
- If W is Lagrangian, then $\dim W = n$.
- If W is symplectic, then $\dim W = 2m$ for some $m \leq n$.

Proof. The first three statements are corollaries of Theorem 2.10.12. For example, if W is isotropic, then $\dim W \leq \dim W^\omega$, and so if $\dim W > n$, then $\dim V > 2n$, a contradiction.

To prove the last statement, note that the symplectic condition amounts to saying that ω is nondegenerate on W : If not, then there is a $\mathbf{w}_0 \in W$ having the property that $\mathbf{w}_0 \neq \mathbf{0}$ and $\omega(\mathbf{w}_0, \mathbf{w}) = 0$ for all $\mathbf{w} \in W$. But this means that $\mathbf{w}_0 \in W^\omega$, contradicting the assumption that W is symplectic.

Because ω is nondegenerate on W , we can apply the argument of the proof of Theorem 2.10.4 to construct a symplectic basis for W that necessarily has an even number of elements, as claimed. $\square$

Example 2.10.15 (Examples of ω -orthogonal complements). The subspace $W_1 = \text{Span}(S) \subset \mathbf{R}^2$, where S is the set in Example 2.10.8, is Lagrangian; note that this means it is both isotropic and coisotropic.

The subspace $W_2 = \text{Span}(S) \subset \mathbf{R}^4$, where S is the set in Example 2.10.9, is isotropic.

If (V, ω) is a symplectic vector space and $\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$ is a symplectic basis for V , then every subspace $W_3 = \text{Span}(\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_k, \mathbf{f}_k\})$ where $1 \leq k \leq n$ is a symplectic subspace of V .

As with inner product spaces, a linear symplectic form on a vector space (V, ω) distinguishes special linear transformations on V , namely those that preserve the symplectic structure.

Definition 2.10.16. Let (V, ω) be a symplectic vector space. A linear transformation $T : V \rightarrow V$ is a *linear symplectomorphism* (or a *linear symplectic transformation*) if for all $\mathbf{v}, \mathbf{w} \in V$ we have $\omega(T(\mathbf{v}), T(\mathbf{w})) = \omega(\mathbf{v}, \mathbf{w})$. In the language of Sect. 2.8, T is a linear symplectomorphism if $T^*\omega = \omega$.

We list here some basic properties of linear symplectomorphisms.

Proposition 2.10.17. *Let (V, ω) be a finite-dimensional symplectic vector space. Then:*

- *If $T_1, T_2 : V \rightarrow V$ are linear symplectomorphisms, then $T_2 \circ T_1$ is also a linear symplectomorphism.*
- *If $T : V \rightarrow V$ is a linear symplectomorphism, then T has an inverse T^{-1} . Moreover, T^{-1} is a linear symplectomorphism.*

Proof. The reader should refer to Proposition 2.8.16 on the pullback of the composition of linear maps.

The first statement follows immediately from the fact that

$$(T_2 \circ T_1)^*\omega = T_1^*(T_2^*\omega).$$

To prove the second statement, we will show that $\ker(T) = \{\mathbf{0}\}$; by Theorem 2.7.6, this means that T is one-to-one, and so by Corollary 2.7.11, T is also onto.

To do this, suppose $\mathbf{v} \in \ker(T)$ and assume $\mathbf{v} \neq \mathbf{0}$. By the nondegeneracy of ω , there exists a vector $\mathbf{w}$ such that $\omega(\mathbf{v}, \mathbf{w}) \neq 0$. But since T is a linear symplectomorphism,

$$\begin{aligned} \omega(\mathbf{v}, \mathbf{w}) &= (T^*\omega)(\mathbf{v}, \mathbf{w}) \\ &= \omega(T(\mathbf{v}), T(\mathbf{w})) \\ &= 0 \quad \text{since } \mathbf{v} \in \ker T, \end{aligned}$$

a contradiction. Hence $\ker(T) = \{\mathbf{0}\}$, proving the claim. So T is invertible with inverse T^{-1} . That T^{-1} is itself a linear symplectomorphism follows from the fact that

$$\omega = (\text{Id})^*\omega = (T \circ T^{-1})^*\omega = (T^{-1})^*(T^*\omega) = (T^{-1})^*\omega,$$

assuming that T is a linear symplectomorphism. $\square$

Linear symplectomorphisms can be characterized in terms of the concept of a symplectic basis.

Theorem 2.10.18. *Let (V, ω) be a symplectic vector space with dimension $\dim V = 2n$ and symplectic basis $\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$. If $T : V \rightarrow V$ is a linear symplectomorphism, then*

$$\{T(\mathbf{e}_1), T(\mathbf{f}_1), \dots, T(\mathbf{e}_n), T(\mathbf{f}_n)\}$$

is also a symplectic basis for V .

Conversely, suppose that

$$\mathcal{B} = \{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$$

and

$$\mathcal{B}' = \{\mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{u}_n, \mathbf{v}_n\}$$

are two symplectic bases for (V, ω) and $T : V \rightarrow V$ is the linear isomorphism defined (according to Theorem 2.6.2) by $T(\mathbf{e}_i) = \mathbf{u}_i$ and $T(\mathbf{f}_i) = \mathbf{v}_i$. Then T is a linear symplectomorphism.

Proof. The first statement follows from the fact that

$$\omega(T(\mathbf{v}), T(\mathbf{w})) = (T^*\omega)(\mathbf{v}, \mathbf{w}) = \omega(\mathbf{v}, \mathbf{w}),$$

assuming that T is a linear symplectomorphism.

To prove the converse, note that if vectors $\mathbf{v}$ and $\mathbf{w}$ are written according to the symplectic basis $\mathcal{B}$, i.e.,

$$\mathbf{v} = \sum (s_i \mathbf{e}_i + t_i \mathbf{f}_i), \quad \mathbf{w} = \sum (a_i \mathbf{e}_i + b_i \mathbf{f}_i),$$

then a calculation shows that

$$\omega(\mathbf{v}, \mathbf{w}) = \sum (s_i b_i - t_i a_i).$$

This calculation holds for any symplectic basis. In particular, we have

$$(T^*\omega)(\mathbf{v}, \mathbf{w}) = \omega(T(\mathbf{v}), T(\mathbf{w})) = \sum (s_i b_i - t_i a_i) = \omega(\mathbf{v}, \mathbf{w}). \quad \square$$

Proposition 2.10.19. *If T is a linear symplectomorphism of (V, ω) and $\mathbf{v}, \mathbf{w}$ are ω -orthogonal, then so are $T(\mathbf{v}), T(\mathbf{w})$.*

Proof. Exercise. $\square$

We turn now to the matrix representation of the standard symplectic form ω_0 on $\mathbf{R}^{2n}$. This case in fact covers the matrix representation for any linear symplectic form on a symplectic vector space (V, ω) , as long as vectors are represented in components relative to a symplectic basis.

Recall that the standard symplectic basis B_0 for $(\mathbf{R}^{2n}, \omega_0)$ was described above, following Theorem 2.10.4. Writing vectors $\mathbf{v}, \mathbf{w}$ as column vectors relative to B_0 , the reader can verify using Proposition 2.8.14 that

$$\omega_0(\mathbf{v}, \mathbf{w}) = \mathbf{w}^T J \mathbf{v},$$

where, using block matrix notation,

$$J = \begin{bmatrix} J_0 & O & \cdots & O \\ O & J_0 & \cdots & O \\ O & O & \ddots & O \\ O & \cdots & O & J_0 \end{bmatrix}, \quad J_0 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The matrix J , representing the standard symplectic form, also allows a matrix characterization of a linear symplectomorphism.

Theorem 2.10.20. *Let $T : \mathbf{R}^{2n} \rightarrow \mathbf{R}^{2n}$ be a linear transformation. Then T is a linear symplectomorphism of $(\mathbf{R}^{2n}, \omega_0)$ if and only if its matrix representation $A = [T]$ relative to the standard symplectic basis satisfies*

$$A^T J A = J.$$

Proof. The condition that $T^* \omega_0 = \omega_0$ means that for all $\mathbf{v}, \mathbf{w} \in \mathbf{R}^{2n}$, $\omega_0(T(\mathbf{v}), T(\mathbf{w})) = \omega_0(\mathbf{v}, \mathbf{w})$. But, in matrix notation,

$$\omega_0(T(\mathbf{v}), T(\mathbf{w})) = (A\mathbf{w})^T J (A\mathbf{v}) = \mathbf{w}^T (A^T J A) \mathbf{v},$$

while $\omega_0(\mathbf{v}, \mathbf{w}) = \mathbf{w}^T J \mathbf{v}$. Hence $T^* \omega_0 = \omega_0$ is equivalent to the matrix equation $A^T J A = J$. $\square$

A $2n \times 2n$ matrix satisfying the condition that $A^T J A = J$ will be called a *symplectic matrix*. We write $\text{Sp}(2n)$ to denote the set of all $2n \times 2n$ symplectic matrices. A number of properties of symplectic matrices will be explored in the exercises. The following theorem indicates only the most important properties.

Theorem 2.10.21. *Let $A \in \text{Sp}(2n)$. Then:*

1. A is invertible;
2. $A^T \in \text{Sp}(2n)$;
3. $A^{-1} \in \text{Sp}(2n)$.

Proof. In light of Theorem 2.10.20, statements (1) and (3) are in fact corollaries of Proposition 2.10.17. However, we prove the statements here using matrix techniques.

Suppose $A \in \text{Sp}(2n)$, i.e., $A^T J A = J$. Then since $\det J = 1$, we have

$$\begin{aligned} 1 &= \det J \\ &= \det(A^T J A) \\ &= (\det A)^2, \end{aligned}$$

and so $\det A = \pm 1 \neq 0$. Hence A is invertible.

Since $J^{-1} = -J$ and $J^2 = -I$, and using the fact that $A^T J A = J$, we have

$$J A^T J A = J^2 = -I,$$

which shows that

$$-JA^T = (JA)^{-1} = A^{-1}J^{-1} = -A^{-1}J,$$

and hence $AJAT = (A^T)^T J(A^T) = J$. So $A^T \in \text{Sp}(2n)$.

We leave the proof of (3) as an exercise. $\square$

We saw in the context of the preceding proof that the determinant of a symplectic matrix is ± 1 . In fact, a stronger results holds.

Theorem 2.10.22. *If $A \in \text{Sp}(2n)$, then $\det(A) = 1$.*

We will defer the proof, however, to Chap. 7. We will ultimately rely on the tools of exterior algebra that we present in Chap. 3.

The following statement concerns the eigenvalues of a symplectic matrix.

Theorem 2.10.23. *Suppose λ is an eigenvalue of the symplectic matrix $A \in \text{Sp}(2n)$ with multiplicity k . Then $1/\lambda$, $\bar{\lambda}$, and $1/\bar{\lambda}$ are also eigenvalues of A with multiplicity k . (Here $\bar{\lambda}$ is the complex conjugate of λ .)*

Proof. Consider the characteristic polynomial $p(x) = \det(A - xI)$; note that 0 cannot be a root, since then A would not be invertible. It is always the case that $\bar{\lambda}$ is a root of p if λ is, since p is a real polynomial, and that the multiplicities of λ and $\bar{\lambda}$ are the same.

For every nonzero x , we have the following:

$$\begin{aligned} p(x) &= \det(A - xI) \\ &= \det(J(A - xI)J^{-1}) \\ &= \det(JAJ^{-1} - xI) \\ &= \det((A^{-1})^T - xI) \quad \text{since } A^T JA = J \\ &= \det((A^{-1} - xI)^T) \\ &= \det(A^{-1} - xI) \\ &= \det(A^{-1}(I - xA)) \\ &= \det(A^{-1}) \det(I - xA) \\ &= \det(I - xA) \quad \text{by Theorem 2.10.22} \\ &= x^{2n} \det\left(\frac{1}{x}I - A\right) \\ &= x^{2n} p\left(\frac{1}{x}\right). \end{aligned}$$

This shows that if λ is a root of p , then so is $1/\lambda$ (and hence $1/\bar{\lambda}$ also).

Now assume that λ is a root of the characteristic polynomial p with multiplicity k , so that

$$p(x) = (x - \lambda)^k q(x)$$

for some polynomial q satisfying $q(\lambda) \neq 0$. But then for $x \neq 0$ we have

$$\begin{aligned} p(x) &= x^{2n} p\left(\frac{1}{x}\right) \quad \text{by the above calculation} \\ &= x^{2n} \left(\frac{1}{x} - \lambda\right)^k q\left(\frac{1}{x}\right) \\ &= \lambda^k x^{2n-k} \left(\frac{1}{\lambda} - x\right)^k q\left(\frac{1}{x}\right). \end{aligned}$$

Hence, since $q(\lambda) \neq 0$, we have that $1/\lambda$ is a root of p with multiplicity k . $\square$

We will have occasion to consider the case of a vector space V that has both a symplectic linear form and an inner product. Unfortunately, the Gram–Schmidt methods of Theorems 2.9.8 and 2.10.4 are not compatible, in the sense that they cannot produce a basis that is simultaneously symplectic and orthogonal. Nevertheless, it is possible to construct such a basis by resorting to techniques particular to complex vector spaces—vector spaces whose scalars are complex numbers.

For basic results about complex vector spaces, the reader may consult any textbook in linear algebra, for example [2]. In the proof of the following theorem, *Hermitian* matrices will play a prominent role. A Hermitian matrix A is a square $(n \times n)$ matrix with complex entries having the property that

$$A = (\overline{A})^T,$$

where the bar represents componentwise complex conjugation. The most important property of Hermitian matrices for our purposes is that they have n linearly independent (over $\mathbf{C}$) eigenvectors that are orthonormal with respect to the standard Hermitian product $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\mathbf{x}}^T \mathbf{y}$ and whose corresponding eigenvalues are real.

Theorem 2.10.24. *Let (V, ω) be a symplectic vector space with $\dim V = 2n$. Suppose that G is an inner product on V . Then there is a symplectic basis*

$$\mathcal{B} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{u}}_n, \tilde{\mathbf{v}}_n\}$$

that is also G -orthogonal, i.e.,

$$G(\tilde{\mathbf{u}}_j, \tilde{\mathbf{v}}_k) = 0 \quad \text{for all } j, k = 1, \dots, n; \quad G(\tilde{\mathbf{u}}_j, \tilde{\mathbf{u}}_k) = G(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_k) = 0 \quad \text{for } j \neq k.$$

Moreover, the basis can be chosen so that $G(\tilde{\mathbf{u}}_j, \tilde{\mathbf{u}}_j) = G(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)$ for all $j = 1, \dots, n$.

Proof. We begin with an orthonormal basis

$$\mathcal{B}' = \{\mathbf{e}_1, \dots, \mathbf{e}_{2n}\}$$

of V relative to G , which exists according to Theorem 2.9.8. Let A be the $2n \times 2n$ matrix defined by the symplectic form ω relative to $\mathcal{B}'$ as follows:

$$\omega(\mathbf{v}, \mathbf{w}) = G(\mathbf{v}, A\mathbf{w})$$

for all $\mathbf{v}, \mathbf{w} \in V$. We write $A = [a_{jk}]$, where $a_{jk} = \omega(\mathbf{e}_j, \mathbf{e}_k)$. Due to the skew-symmetry of ω , the matrix A is skew-symmetric: $A^T = -A$.

Throughout this proof, we will consider vectors $\mathbf{v}, \mathbf{w}$ to be column vectors written using components relative to the basis $\mathcal{B}'$. In particular,

$$G(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T \mathbf{w},$$

$$\text{and } \omega(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T A \mathbf{w}.$$

The skew-symmetry of A implies that the $2n \times 2n$ matrix iA with purely imaginary entries ia_{jk} is Hermitian:

$$\begin{aligned} (\overline{iA})^T &= (-iA)^T && \text{since the entries of } iA \text{ are purely imaginary} \\ &= -iA^T \\ &= -i(-A) && \text{since } A \text{ is skew-symmetric} \\ &= iA. \end{aligned}$$

Since ω is nondegenerate, A must be invertible, and hence the eigenvalues of A are nonzero. By the property of Hermitian matrices mentioned above, there are in fact $2n$ linearly independent eigenvectors of iA that are orthonormal with respect to the Hermitian product and with real corresponding eigenvalues. In fact, the reader can verify that the eigenvectors of iA occur in pairs

$$\mathbf{y}_1, \overline{\mathbf{y}}_1, \dots, \mathbf{y}_n, \overline{\mathbf{y}}_n$$

(these are vectors with complex components, and the bar represents component-wise complex conjugation). The corresponding eigenvalues will be denoted by

$$\mu_1, -\mu_1, \dots, \mu_n, -\mu_n.$$

The orthonormality is expressed in matrix notation as

$$\overline{\mathbf{y}}_j^T \mathbf{y}_k = \delta_k^j = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

Note that since $(iA)\mathbf{y}_j = \mu_j\mathbf{y}_j$, we have $A\mathbf{y}_j = (-i\mu_j)\mathbf{y}_j$; in other words, the eigenvalues of A are $\pm i\mu_j$. For each $j = 1, \dots, n$, we choose pairs λ_j and $\mathbf{x}_j$ as follows: From each pair of eigenvectors $\mathbf{y}_j, \bar{\mathbf{y}}_j$ with corresponding nonzero eigenvalues $\mu_j, -\mu_j$, choose $\lambda_j = \pm\mu_j$ so that $\lambda_j > 0$, and then if $\lambda_j = \mu_j$, choose $\mathbf{x}_j = \bar{\mathbf{y}}_j$, while if $\lambda_j = -\mu_j$, choose $\mathbf{x}_j = \mathbf{y}_j$. In this way, we have $A\mathbf{x}_j = i\lambda_j\mathbf{x}_j$ with $\lambda_j > 0$.

Write

$$\mathbf{x}_j = \mathbf{u}_j + i\mathbf{v}_j$$

with vectors $\mathbf{u}_j$ and $\mathbf{v}_j$ having real components. We claim that the set $\mathcal{B}'' = \{\mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{u}_n, \mathbf{v}_n\}$ is a G -orthogonal basis for V . The fact that $\mathcal{B}''$ is a basis for V is a consequence of the above-mentioned property of the eigenvectors of a Hermitian matrix. To show that $\mathcal{B}''$ is G -orthogonal, we note that the Hermitian orthonormality condition $\bar{\mathbf{x}}_j^T \mathbf{x}_k = \delta_k^j$ can be expressed as

$$\mathbf{u}_j^T \mathbf{u}_k + \mathbf{v}_j^T \mathbf{v}_k = \delta_k^j, \quad \mathbf{u}_j^T \mathbf{v}_k - \mathbf{v}_j^T \mathbf{u}_k = 0.$$

Also, the fact that $A\mathbf{x}_j = i\lambda_j\mathbf{x}_j$ means that

$$A\mathbf{u}_j = -\lambda_j\mathbf{v}_j, \quad A\mathbf{v}_j = \lambda_j\mathbf{u}_j.$$

Hence, for $j \neq k$, we have

$$\begin{aligned} \mathbf{u}_j^T \mathbf{u}_k &= \mathbf{u}_j^T \left(\frac{1}{\lambda_k} A\mathbf{v}_k \right) \\ &= \frac{1}{\lambda_k} (\mathbf{u}_j^T A\mathbf{v}_k)^T \quad \text{since the quantity in parentheses is a scalar} \\ &= \frac{1}{\lambda_k} (\mathbf{v}_k^T A^T \mathbf{u}_j) \\ &= -\frac{1}{\lambda_k} (\mathbf{v}_k^T A\mathbf{u}_j) \quad \text{since } A \text{ is skew-symmetric} \\ &= \frac{\lambda_j}{\lambda_k} (\mathbf{v}_k^T \mathbf{v}_j) \quad \text{since } A\mathbf{u}_j = -\lambda_j\mathbf{v}_j \\ &= -\frac{\lambda_j}{\lambda_k} \mathbf{u}_k^T \mathbf{u}_j \quad \text{since } \mathbf{u}_j^T \mathbf{u}_k + \mathbf{v}_j^T \mathbf{v}_k = 0 \text{ for } j \neq k \\ &= -\frac{\lambda_j}{\lambda_k} (\mathbf{u}_k^T \mathbf{u}_j)^T \quad \text{since the quantity in parentheses is a scalar} \\ &= -\frac{\lambda_j}{\lambda_k} \mathbf{u}_j^T \mathbf{u}_k, \end{aligned}$$

which implies, since $\lambda_j, \lambda_k > 0$, that $\mathbf{u}_j^T \mathbf{u}_k = 0$.

In the same way,

$$\mathbf{v}_j^T \mathbf{v}_k = 0 \quad \text{for } j \neq k.$$

We leave it to the reader to show, in a similar way, that for all j and k ,

$$\mathbf{u}_j^T \mathbf{v}_k = \mathbf{v}_j^T \mathbf{u}_k = 0.$$

All this shows that $\mathcal{B}''$ is G -orthogonal, since $G(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T \mathbf{w}$.

Note that

$$\begin{aligned} \omega(\mathbf{u}_j, \mathbf{v}_j) &= \mathbf{u}_j^T A \mathbf{v}_j \\ &= \lambda_j \mathbf{u}_j^T \mathbf{u}_j \\ &= \lambda_j |\mathbf{u}_j|^2 \\ &> 0. \end{aligned}$$

We leave it as an exercise for the reader to find scalars c_j and d_j such that for $\tilde{\mathbf{u}}_j = c_j \mathbf{u}_j$ and $\tilde{\mathbf{v}}_j = d_j \mathbf{v}_j$,

$$\omega(\tilde{\mathbf{u}}_j, \tilde{\mathbf{v}}_j) = 1 \quad \text{and} \quad G(\tilde{\mathbf{u}}_j, \tilde{\mathbf{u}}_j) = G(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)$$

for all $j = 1, \dots, n$. The set

$$\mathcal{B} = \{\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{u}}_n, \tilde{\mathbf{v}}_n\}$$

is the desired basis. □

We will see that for the standard symplectic vector space $(\mathbf{R}^{2n}, \omega_0)$, ellipsoids play an important role in measuring linear symplectomorphisms. By an *ellipsoid*, we mean a set $E \subset \mathbf{R}^{2n}$ defined by a positive definite symmetric matrix A in the following way:

$$E = \{\mathbf{x} \in \mathbf{R}^{2n} \mid \mathbf{x}^T A \mathbf{x} \leq 1\}.$$

An important fact about ellipsoids is that they can be brought into a “normal form” by means of linear symplectomorphisms.

Theorem 2.10.25. *Let $E \subset \mathbf{R}^{2n}$ be an ellipsoid defined by the positive definite symmetric matrix A . Then there are positive constants $r_1, \dots, r_n$ and a linear symplectomorphism $\Phi : (\mathbf{R}^{2n}, \omega_0) \rightarrow (\mathbf{R}^{2n}, \omega_0)$ such that $\Phi(E(r_1, \dots, r_n)) = E$, where*

$$E(r_1, \dots, r_n) = \left\{ (x_1, y_1, \dots, x_n, y_n) \mid \sum \left(\frac{x_i^2 + y_i^2}{r_i^2} \right) \leq 1 \right\}.$$

The constants are uniquely determined when ordered $0 < r_1 \leq \dots \leq r_n$.

Proof. Since A is a positive definite symmetric matrix, it defines an inner product G by $G(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}$. The ellipsoid E is then characterized as

$$E = \{\mathbf{b} \in \mathbf{R}^{2n} \mid G(\mathbf{b}, \mathbf{b}) \leq 1\}.$$

According to Theorem 2.10.24, there is a basis

$$\{\mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{u}_n, \mathbf{v}_n\}$$

that is both symplectic relative to ω_0 and G -orthogonal, with $G(\mathbf{u}_i, \mathbf{u}_i) = G(\mathbf{v}_i, \mathbf{v}_i)$ for all $i = 1, \dots, n$. So define the positive constants r_i by

$$\frac{1}{r_i^2} = G(\mathbf{u}_i, \mathbf{u}_i).$$

Let $\Phi : \mathbf{R}^{2n} \rightarrow \mathbf{R}^{2n}$ be the linear symplectomorphism defined by its action on the standard symplectic basis $\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$ for $(\mathbf{R}^{2n}, \omega_0)$:

$$\Phi(\mathbf{e}_i) = \mathbf{u}_i, \quad \Phi(\mathbf{f}_i) = \mathbf{v}_i.$$

More explicitly, since

$$(x_1, y_1, \dots, x_n, y_n) = x_1 \mathbf{e}_1 + y_1 \mathbf{f}_1 + \dots + x_n \mathbf{e}_n + y_n \mathbf{f}_n,$$

we have

$$\Phi(x_1, y_1, \dots, x_n, y_n) = x_1 \mathbf{u}_1 + y_1 \mathbf{v}_1 + \dots + x_n \mathbf{u}_n + y_n \mathbf{v}_n.$$

We will show that $\Phi(E(r_1, \dots, r_n)) = E$. On the one hand, suppose $\mathbf{b} \in \Phi(E(r_1, \dots, r_n))$. In other words, there is $\mathbf{a} \in E(r_1, \dots, r_n)$ such that $\Phi(\mathbf{a}) = \mathbf{b}$. Writing

$$\mathbf{a} = (x_1, y_1, \dots, x_n, y_n) = x_1 \mathbf{e}_1 + y_1 \mathbf{f}_1 + \dots + x_n \mathbf{e}_n + y_n \mathbf{f}_n,$$

we then have

$$\mathbf{b} = \Phi(\mathbf{a}) = x_1 \mathbf{u}_1 + y_1 \mathbf{v}_1 + \dots + x_n \mathbf{u}_n + y_n \mathbf{v}_n,$$

and so

$$\begin{aligned} G(\mathbf{b}, \mathbf{b}) &= \sum (x_i^2 G(\mathbf{u}_i, \mathbf{u}_i) + y_i^2 G(\mathbf{v}_i, \mathbf{v}_i)) \\ &= \sum \left(x_i^2 \left(\frac{1}{r_i^2} \right) + y_i^2 \left(\frac{1}{r_i^2} \right) \right) \\ &\leq 1 \quad \text{since } \mathbf{a} \in E(r_1, \dots, r_n). \end{aligned}$$

Hence $\mathbf{b} \in E$, and so $\Phi(E(r_1, \dots, r_n)) \subset E$.

On the other hand, suppose that $\mathbf{b} \in E$, so that $G(\mathbf{b}, \mathbf{b}) \leq 1$. There is $\mathbf{a} \in \mathbf{R}^{2n}$ such that $\Phi(\mathbf{a}) = \mathbf{b}$, since Φ is a linear isomorphism. Writing $\mathbf{b}$ according to the basis above, we obtain

$$\mathbf{b} = \tilde{x}_1 \mathbf{u}_1 + \tilde{y}_1 \mathbf{v}_1 + \cdots + \tilde{x}_n \mathbf{u}_n + \tilde{y}_n \mathbf{v}_n,$$

so

$$\mathbf{a} = (\tilde{x}_1, \tilde{y}_1, \dots, \tilde{x}_n, \tilde{y}_n).$$

But

$$\sum \left(\frac{\tilde{x}_i^2 + \tilde{y}_i^2}{r_i^2} \right) = G(\mathbf{b}, \mathbf{b}) \leq 1,$$

and so $\mathbf{a} \in E(r_1, \dots, r_n)$ and $E \subset \Phi(E(r_1, \dots, r_n))$.

All this shows that $\Phi(E(r_1, \dots, r_n)) = E$.

To show that the constants r_i are uniquely determined up to ordering, suppose that there are linear symplectomorphisms $\Phi_1, \Phi_2 : \mathbf{R}^{2n} \rightarrow \mathbf{R}^{2n}$ and n -tuples $(r_1, \dots, r_n), (r'_1, \dots, r'_n)$ with $0 < r_1 \leq \cdots \leq r_n$ and $0 < r'_1 \leq \cdots \leq r'_n$ such that

$$\Phi_1(E(r_1, \dots, r_n)) = E, \quad \Phi_2(E(r'_1, \dots, r'_n)) = E.$$

Then, writing $\Phi = \Phi_1^{-1} \circ \Phi_2$, we have

$$\Phi(E(r'_1, \dots, r'_n)) = E(r_1, \dots, r_n).$$

In matrix notation, this says that $\mathbf{x}^T D' \mathbf{x} \leq 1$ if and only if $(\Phi \mathbf{x})^T D (\Phi \mathbf{x}) = \mathbf{x}^T (\Phi^T D \Phi) \mathbf{x} \leq 1$, where $\mathbf{x} = (x_1, y_1, \dots, x_n, y_n)$ is a column vector, D is the diagonal matrix $D = \text{diag} [1/(r_1)^2, 1/(r_1)^2, \dots, 1/(r_n)^2, 1/(r_n)^2]$, and D' is the diagonal matrix $D' = \text{diag} [1/(r'_1)^2, 1/(r'_1)^2, \dots, 1/(r'_n)^2, 1/(r'_n)^2]$. This implies that

$$\Phi^T D \Phi = D'.$$

The facts that Φ satisfies $\Phi^T J \Phi = J$ and $J^{-1} = -J$ together imply $\Phi^T = -J \Phi^{-1} J$, and so

$$\Phi^{-1} J D \Phi = J D'.$$

This shows that $J D$ is similar to $J D'$, and so the two matrices have the same eigenvalues. The reader may verify that the eigenvalues of $J D$ are $\pm i r_j$ and those of $J D'$ are $\pm i r'_j$. Since the r_i and r'_i are ordered from least to greatest, we must have $r_j = r'_j$ for all $j = 1, \dots, n$. $\square$

Theorem 2.10.25 prompts the following definition.

Definition 2.10.26. Let $E \subset \mathbf{R}^{2n}$ be an ellipsoid in the standard symplectic space $(\mathbf{R}^{2n}, \omega_0)$. The *symplectic spectrum* of E is the unique n -tuple $\sigma(E) = (r_1, \dots, r_n)$, $0 < r_1 \leq \cdots \leq r_n$, such that there is a linear symplectomorphism Φ with $\Phi(E(r_1, \dots, r_n)) = E$ (Fig. 2.6).

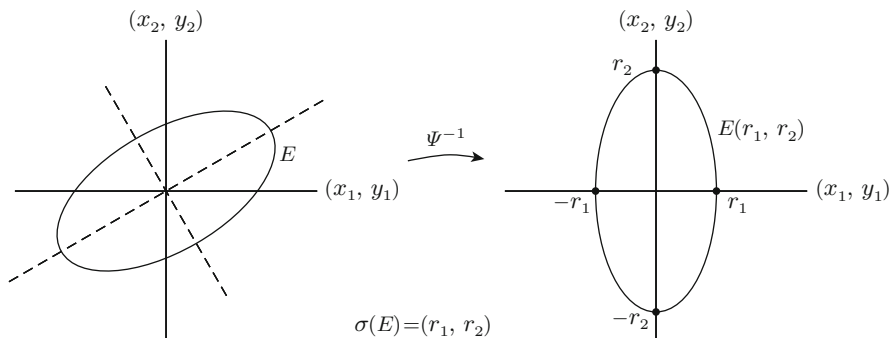


Fig. 2.6 Linear symplectomorphisms and the symplectic spectrum.

We will continue to develop some topics in linear symplectic geometry in Sect. 7.7 as motivation for a key concept in (nonlinear) symplectic geometry, the symplectic capacity.

2.11 For Further Reading

With the exception of Sects. 2.8 and 2.10, much of the material in this chapter can be found in any textbook on linear algebra. The notation here generally follows that in [2].

While many linear algebra textbooks have detailed presentations of inner product spaces, symplectic vector spaces are usually presented only as introductory matter in the context of specialized texts. We refer to A. Banyaga's summary in [4, Chap. 1] or to [31].

2.12 Exercises

The exercises in this chapter emphasize topics not usually presented in a first elementary linear algebra course.

2.1. Prove Theorem 2.4.3.

2.2. Prove Theorem 2.4.10.

2.3. Prove Theorem 2.4.13.

2.4. Let $T : V \rightarrow W$ be a linear isomorphism between vector spaces V and W , and let $T^{-1} : W \rightarrow V$ be the inverse of T . Show that T^{-1} is a *linear* transformation.

2.5. Consider the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ of $\mathbf{R}^3$, where

$$\mathbf{b}_1 = (1, 0, 1), \quad \mathbf{b}_2 = (1, 1, 0), \quad \mathbf{b}_3 = (0, 2, 1).$$

- (a) Write the components of $\mathbf{w} = (2, 3, 5)$ relative to the basis $\mathcal{B}$.
 (b) Let $\{\beta_1, \beta_2, \beta_3\}$ be the basis of $(\mathbf{R}^3)^*$ dual to $\mathcal{B}$. Compute $\beta_i(\mathbf{w})$ for each $i = 1, 2, 3$, where $\mathbf{w}$ is the vector given in part (a).
 (c) For each $i = 1, 2, 3$, compute $\beta_i(\mathbf{v})$, where $\mathbf{v} = (v_1, v_2, v_3)$ is an arbitrary vector in $\mathbf{R}^3$.

2.6. For each of the linear transformations Ψ and linear one-forms T below, compute Ψ^*T :

(a) $\Psi : \mathbf{R}^3 \rightarrow \mathbf{R}^3, \Psi(u, v, w) = (2u, 3u - v - w, u + 2w),$

$$T(x, y, z) = 3x + y - z.$$

(b) $\Psi : \mathbf{R}^3 \rightarrow \mathbf{R}^2, \Psi(u, v, w) = (v, 2u - w),$

$$T(x, y) = x + 3y.$$

(c) $\Psi : \mathbf{R}^4 \rightarrow \mathbf{R}^3, \Psi(x, y, z, w) = (x + y - z - 2w, w - 4x - z, y + 3z),$

$$T(x, y, z) = x - 2y + 3z.$$

2.7. Let $\alpha \in (\mathbf{R}^3)^*$ be given by $\alpha(x, y, z) = 4y + z$.

- (a) Find $\ker \alpha$.
 (b) Find all linear transformations $\Psi : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ with the property that $\Psi^*\alpha = \alpha$.

2.8. Consider the linear transformation $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ given by $T(x_1, x_2) = (2x_1 - x_2, x_1 + 3x_2)$.

- (a) Compute T^*G_0 , where G_0 is the standard inner product defined in Example 2.8.11.
 (b) Compute T^*S , where S is the bilinear form in Example 2.8.13.

2.9. Let $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be a linear transformation described in matrix form $[T]$ relative to the standard basis for $\mathbf{R}^n$. Show that for every $n \times n$ matrix A , one has

$$T^*G_A = G_{A[T]},$$

where G_A and $G_{A[T]}$ are defined according to Example 2.8.12.

2.10. Prove the following converse to Proposition 2.8.14: Let B be an $n \times n$ matrix and let $\mathcal{E}$ be a basis for the n -dimensional vector space V . Then the function $b : V \times V \rightarrow \mathbf{R}$ defined by

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{w}^T B \mathbf{v},$$

where $\mathbf{v}$ and $\mathbf{w}$ are written as column vectors relative to the basis $\mathcal{E}$, is a bilinear form.

2.11. Use Exercise 2.10 to give five examples of bilinear forms on $\mathbf{R}^3$ and five examples of bilinear forms on $\mathbf{R}^4$.

2.12. Let b , B , and $\mathcal{E}$ be as given in Exercise 2.10, and let $T : V \rightarrow V$ be a linear transformation. Show that $T^*b = \tilde{b}$, where $\tilde{b}$ is the bilinear form corresponding to the matrix $A^T B A$ for $A = [T]_{\mathcal{E}, \mathcal{E}}$, the matrix representation of T relative to the basis $\mathcal{E}$.

2.13. For each of the following 2×2 matrices, write the coordinate expression for the inner product G_A relative to the standard basis as in Example 2.9.4. For each, compute $G_A(\mathbf{e}_1, \mathbf{e}_1)$, $G_A(\mathbf{e}_1, \mathbf{e}_2)$, and $G_A(\mathbf{e}_2, \mathbf{e}_2)$ along with $\angle(\mathbf{e}_1, \mathbf{e}_2)$, where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$:

(a) $\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix};$

(b) $\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix};$

(c) $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}.$

2.14. Show that the function $G(\mathbf{v}, \mathbf{w}) = v_1 w_1 + 2v_1 w_2 + 2v_2 w_1 + 5v_2 w_2$ is an inner product on $\mathbf{R}^2$, where $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. Find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for $\mathbf{R}^2$ relative to G .

2.15. Let $\{\mathbf{u}_1, \mathbf{u}_2\}$ be the basis for $\mathbf{R}^2$ given by $\mathbf{u}_1 = (3, 2)$ and $\mathbf{u}_2 = (1, 1)$. Let G be the inner product on $\mathbf{R}^2$ such that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is orthonormal (see Theorem 2.9.9). Find $G(\mathbf{v}, \mathbf{w})$, where $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. Find $\angle((1, 0), (0, 1))$.

2.16. Prove Theorem 2.9.10.

2.17. For the following subspaces W of $\mathbf{R}^n$, find a basis for $W^\perp$, the orthogonal complement of W relative to the standard inner product on $\mathbf{R}^n$

(a) $W = \text{Span}\{(1, 2)\} \subset \mathbf{R}^2;$

(b) $W = \text{Span}\{(1, 2, 3)\} \subset \mathbf{R}^3;$

(c) $W = \text{Span}\{(1, 0, 1), (-1, 1, 0)\} \subset \mathbf{R}^3;$

(d) $W = \text{Span}\{(1, -2, 2, 1), (0, 1, 1, -3)\} \subset \mathbf{R}^4.$

2.18. Provide the details for the proof of Theorem 2.9.16.

2.19. Let (V, G) be an inner product space and let W be a subset of V .

(a) Show that $W \subset (W^\perp)^\perp$.

(b) Show that if V is finite-dimensional, then there is an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ of V such that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a basis for W and $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is a basis for $W^\perp$. (See Theorem 2.9.16.)

(c) Show that if V is finite-dimensional, then $(W^\perp)^\perp \subset W$, and so by (a), $W = (W^\perp)^\perp$.

2.20. Prove Proposition 2.9.18.

2.21. Prove Proposition 2.9.21.

2.22. Let (V, G) be a finite-dimensional inner product space. Show that a linear transformation $T : V \rightarrow V$ is a linear isometry if and only if for every orthonormal basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of V , the set $\{T(\mathbf{e}_1), \dots, T(\mathbf{e}_n)\}$ is also an orthonormal basis for V .

2.23. Give three examples of linear symplectic forms on $\mathbf{R}^4$.

2.24. Suppose $\mathcal{B} = \{\mathbf{a}_1, \mathbf{b}_1, \dots, \mathbf{a}_n, \mathbf{b}_n\}$ is a basis for $\mathbf{R}^{2n}$. Define the alternating bilinear form $\omega_{\mathcal{B}}$ by specifying its action on the basis vectors,

$$\begin{aligned}\omega_{\mathcal{B}}(\mathbf{a}_i, \mathbf{a}_j) &= \omega_{\mathcal{B}}(\mathbf{b}_i, \mathbf{b}_j) = 0 \quad \text{for all } i, j, \\ \omega_{\mathcal{B}}(\mathbf{a}_i, \mathbf{b}_j) &= 0 \quad \text{for } i \neq j, \\ \omega_{\mathcal{B}}(\mathbf{a}_i, \mathbf{b}_i) &= 1,\end{aligned}$$

and extending bilinearly. Show that $\omega_{\mathcal{B}}$ is a linear symplectic form.

2.25. Define a bilinear form S on $\mathbf{R}^4$ by $S(\mathbf{v}, \mathbf{w}) = \mathbf{w}^T A \mathbf{v}$, where

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & 2 & 0 \\ -1 & -2 & 0 & 3 \\ -1 & 0 & -3 & 0 \end{bmatrix}.$$

(a) Show that S is a linear symplectic form.

(b) Use the process outlined in Theorem 2.10.4 to find a symplectic basis $\{\mathbf{e}_1, \mathbf{f}_1, \mathbf{e}_2, \mathbf{f}_2\}$ for $\mathbf{R}^4$ relative to S .

2.26. Use the procedure in Theorem 2.10.4 to construct three different symplectic bases for $\mathbf{R}^4$ by making appropriate choices at different stages of the process.

2.27. Consider $(\mathbf{R}^4, \omega_0)$, where ω_0 is the standard linear symplectic form on $\mathbf{R}^4$. Decide whether the following subspaces of $\mathbf{R}^4$ are isotropic, coisotropic, Lagrangian, or symplectic:

- (a) $W_1 = \text{Span}\{(1, 0, -1, 3)\}$;
- (b) $W_2 = \text{Span}\{(3, 1, 0, -1), (2, 1, 2, 1)\}$;
- (c) $W_3 = \text{Span}\{(1, 0, 2, -1), (0, 1, 1, -1)\}$;
- (d) $W_4 = \text{Span}\{(1, 1, 1, 0), (2, -1, 0, 1), (0, 2, 0, -1)\}$;
- (e) $W_5 = \ker T$, where $T : \mathbf{R}^4 \rightarrow \mathbf{R}^2$ is given by

$$T(x_1, y_1, x_2, y_2) = (2x_2 - y_1, x_1 + x_2 + y_1 + y_2).$$

2.28. Prove Theorem 2.10.7.

2.29. Prove Theorem 2.10.12.

2.30. Let W_1 and W_2 be subspaces of a symplectic vector space (V, ω) . Show that if $W_1 \subset W_2$, then $(W_2)^\omega \subset (W_1)^\omega$.

2.31. Show that if W is a subspace of a finite-dimensional symplectic vector space (V, ω) , then $(W^\omega)^\omega = W$. (See Exercise 2.19.)

2.32. Is it possible for a 2-dimensional subspace of a 4-dimensional symplectic vector space to be neither symplectic nor Lagrangian? If so, find necessary conditions for this to occur. If not, state and prove the corresponding result. To what extent can this question be generalized to higher dimensions?

2.33. Prove Proposition 2.10.19.

2.34. Prove Theorem 2.10.5.

2.35. For each of the examples in Exercise 2.23, write the isomorphism Ψ described in Theorem 2.10.5 explicitly in terms of the standard bases of $\mathbf{R}^4$ and $(\mathbf{R}^4)^*$.

2.36. Let W be a subspace of a finite-dimensional symplectic vector space (V, ω) . Let $\Psi : V \rightarrow V^*$ be the isomorphism described in Theorem 2.10.5.

(a) Let

$$W^0 = \{\alpha \in V^* \mid \alpha(\mathbf{w}) = 0 \text{ for all } \mathbf{w} \in W\}.$$

Show that W^0 is a subspace of V^* .

(b) Show that $\Psi(W^\omega) = W^0$.

(c) Show that $\Psi(W) = (W^\omega)^0$.

2.37. Provide the details for the proof of Theorem 2.10.24. In particular:

(a) Show that the set $\mathcal{B}''$ is a basis for V .

(b) Show that $\mathbf{u}_j^T \mathbf{v}_k = \mathbf{v}_j^T \mathbf{u}_k = 0$.

(c) Find scalars c_j and d_j such that for $\tilde{\mathbf{u}}_j = c_j \mathbf{u}_j$ and $\tilde{\mathbf{v}}_j = d_j \mathbf{v}_j$,

$$\omega(\tilde{\mathbf{u}}_j, \tilde{\mathbf{v}}_j) = 1 \quad \text{and} \quad G(\tilde{\mathbf{u}}_j, \tilde{\mathbf{u}}_j) = G(\tilde{\mathbf{v}}_j, \tilde{\mathbf{v}}_j)$$

for all $j = 1, \dots, n$.

2.38. Verify directly that the matrix

$$A = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 2 & -1 & 1 & -1 \end{bmatrix}$$

is a symplectic matrix, i.e., that $A^T J A = J$.

2.39. Let $\mathcal{B} = \{\mathbf{a}_1, \mathbf{b}_1, \dots, \mathbf{a}_n, \mathbf{b}_n\}$ be a symplectic basis for the standard symplectic space $(\mathbf{R}^{2n}, \omega_0)$. Show that the matrix

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{b}_1 & \cdots & \mathbf{a}_n & \mathbf{b}_n \end{bmatrix}$$

is a symplectic matrix.

2.40. Show that if $A \in \text{Sp}(2n)$, then $A^{-1} \in \text{Sp}(2n)$.

2.41. Show that if $A \in \text{Sp}(2n)$, then $A^{-1} = -JA^T J$.

<http://www.springer.com/978-1-4614-7731-0>

First Steps in Differential Geometry

Riemannian, Contact, Symplectic

McInerney, A.

2013, XIII, 410 p. 54 illus., 25 illus. in color., Hardcover

ISBN: 978-1-4614-7731-0