First Steps in Differential Geometry: Riemannian, Contact, Symplectic

Andrew McInerney

Contents

Introduction

Differential geometry can be described as the application of the tools of calculus to questions of geometry. Beginning with the verification of age-old geometrical measurements like the circumference and area of a circle, the new techniques of calculus showed their power by quickly dispensing with questions that had long engaged thinkers from antiquity through the 17th and 18th centuries.

In the 19th century, Gauss displayed the extent to which calculus (in particular, the first two derivatives) determines basic properties of curves and surfaces, at least locally. These results can properly be called the beginning of "classical" differential geometry.

Gauss's student Riemann introduced the notion of a manifold, which brought differential geometry into its own. In particular, the first notions of the tangent space would allow the techniques of linearization and ultimately the tools of linear algebra to be brought to bear on geometric questions.

Perhaps most prominently, Einstein used the notion of a manifold to frame the general theory of relativity. Another conceptual milestone accomplished with the notion of a manifold concept was allowing a common framework for the new, non-Euclidean geometries that were struggling to gain currency in the 19th century. Both of these developments occurred by generalizing the notion of distance and imposing more general metric structures on a manifold. This is what is today known as *riemannian geometry*.

To lesser fanfare, the notion of a manifold allowed a whole new way of generalizing geometry. Sophus Lie, under the influence of Felix Klein, introduced the notion of transformation groups in the course of investigating differential equations. Later, into the early 20th century through the continued work of F. Engel and E. Cartan, these notions led to what are today known as *symplectic* and *contact* structures on manifolds, which like the metric structures of riemannian geometry, are defined by referring to the tangent space.

It is impossible to overstate the impact of this latter development for the field of differential geometry. In fact, today it is possible to describe differential geometry as "the study of smoothly-varying structures on the tangent space." It is the aim of this text to develop this point of view.

For that reason, the development here is somewhat different than the classical introductory texts in differential geometry, for example the works of Struik [**41**], do Carmo [**12**], O'Neill [**34**] or Kühnel [**26**]. Those texts are aimed at introducing riemannian geometry, and especially the metric tensor and its derived concepts like curvature.

Here, our goal is to develop the architecture necessary to introduce contact and symplectic geometry alongside their riemannian cousin. After presenting some preliminary material needed from linear algebra, we spend more time than usual in presenting the definition of the tangent space and notions immediately connected to it, like vector fields. We then present a chapter on differential forms and tensors, which are the "structures" on the tangent space referred to above.

The first three chapters are really a prelude to the core of the book, which is an exposition of the differential geometry of a symmetric, positive-definite 2-tensor (riemannian geometry), a nondegenerate one-form (contact geometry), and a closed, nondegenerate two-form (symplectic geometry). There will be no attempt to give an exhaustive treatment of any of these vast areas of current mathematical research. Rather, the goal is to introduce students early in their mathematical careers to this broader view of geometry.

It is rare to present these three geometric structures side by side as we do here. We do so to emphasize one of the text's major themes: differential geometry as the study of tensor structures on the tangent space. In each case, we will show how a tensor structure determines not only certain key geometric objects, but also singles out special functions or transformations that "preserve the structure."

Differential geometry offers a smooth transition from the standard university mathematics sequence of the first four semesters—calculus through differential equations and linear algebra—to the higher levels of abstraction and proof encountered at the upper division by mathematics majors. Topics introduced or hinted at in calculus and linear algebra are used concretely, but in a new setting. Granted, the simplicity and *ab initio* nature of first courses in abstract algebra or number theory make them ideal settings for students to learn the practice of proofs. Elementary differential geometry sacrifices these in favor of the familiar ground of derivatives and linear transformations, emphasizing instead the importance of proper definition and generality in mathematics.

Indeed, here lies another main goal of this book: to bring the student who has finished two years with a solid foundation in the standard mathematics curriculum into contact with the beauty of "higher" mathematics. In particular, the presentation here emphasizes the consequences of a definition and the careful use of examples and constructions in order to explore those consequences.

This goal places certain limitations on the presentation. The notion of a manifold, which is the basic setting for modern differential geometry, implies a significant role for topology. A manifold is "locally euclidean," in the same sense that the surface of the earth looks flat to the typical human observer on earth.

This text, however, will steer clear of topology as much as possible in order to give center stage to the role of calculus. For the more advanced reader, this will mean that virtually the entire text is "local." Important theorems about global riemannian geometry like the Gauss-Bonnet theorem are thus missing from this presentation. This is an even more severe limitation in the cases of contact and symplectic geometry, as we will discuss in their respective chapters.

For this reason, we will avoid the use of the term "manifold" altogether. This will have some unfortunate consequences in terminology. For example, we will refer to submanifolds as "geometric sets."

As a text aimed at a "transitional" audience, students who have completed the traditional calculus and linear algebra sequence but who have not necessarily been exposed to the more abstract formulations of pure mathematics, we should say a word about the role of proofs. The body of the text is structured to include proofs, or ideas of proof, of most of the main statements and results. This is done not just for the sake of mathematical rigor. Rather, it is premised on the perspective that proofs provide more than deductive logical justifications; they also provide a showcase where the main techniques and concepts can be put on display. In addition, in some cases the reader will be asked to supply proofs or details of a proof as a way to exercise this vital mathematical skill.

Most exercises, however, will be designed to present and explore examples of the mathematical constructions involved. This is based on the point of view that mathematics, and geometry in particular, is not merely a deductive undertaking. There is a rich content to mathematics, the appreciation of which requires intuition and familiarity.

CHAPTER 0

Basic objects and notation

Most of modern mathematics is expressed using the language of sets and functions. This can be a significant hurdle for the student whose mathematical experience, possibly through the entire calculus sequence, has not included any emphasis on sets or set operations. For that reason, we review these basic ideas in this chapter with the goal of both establishing the notation and serving as a quick reference when proceeding through the main part of the text.

0.1. Sets

The basic concept that the notion of a set is meant to capture is that of inclusion or exclusion. Unfortunately, there are inherent logical difficulties in writing a formal definition of a set. We can resort to one standard "definition:" "A set is a collection of objects." This gives a sense of both the distinction between the objects under consideration and the collection of objects, as well as the sense of being "included" in the collection or "not included." Unfortunately, it leaves undefined what is meant exactly by the terms "collection" and "object," and so leaves much to be desired from the perspective of a mathematical definition.

For that reason, we will not try to be too precise in defining a set. Rather, we will call the objects under consideration *elements*, and we think of a *set* as "something that contains elements."

We will generally write a set using an upper-case letter: A, B, S, etc. We will write elements using small-case letters: a, b, x, y, etc. We express the relation that "x is an element of A" by writing

$$x \in A.$$

If, on the contrary, an element y is not an element of the set A, we write $y \notin A$.

The basic assumption that we make about sets is that they are *well-defined*: For every element x and every set A,

$$\text{Either } x \in A \text{ or } x \notin A,$$

and the statement "$x \in A$ and $x \notin A$" is false for all x.

In any particular problem, the context will imply a *universal set*, which is the set of all objects under consideration. For example, a team of statisticians might be concerned with a data set of measurements of heights in a given population of people. A team of geometers might be concerned with properties of the set of points in three-dimensional space. When the statisticians ask whether an element is in a particular set, they will only consider elements in their "universe," and so in particular will not even ask the question of whether the geometers' points are in the statisticians' sets.

While the universal set for a given discussion or problem may or may not be explicitly stated, it should always be able to be established from the context. From a logical point of view, in fact, the universal set *is* the context.

There are several standard ways of describing sets. The most basic way is by listing the elements, written using brackets to enclose the elements of the set. For example, a set A with three elements a, b and c is written

$$A = \{a, b, c\}.$$

The statement $b \in A$ is read, "b is an element of A," or just "b is in A." Note that the order in which the elements are listed is not important, so that for example $\{a, b, c\}$ is the same set as $\{b, a, c\}$.

Describing a set by means of a list is also possible in the case of (countably) infinite sets, at least when there is a pattern involved. For example, the set of even natural numbers can be expressed as

$$E = \{2, 4, 6, \ldots\}.$$

Here, the ellipsis $(\ldots)$ expresses a pattern which should be obvious to the reader in context.

Most often, however, sets will be described in a form known as *set-builder notation*. In this notation, a set is described as all elements (of the universal set) having a certain property or properties. These properties are generally given in the form of a logical statement about an element x, which we can write as $P(x)$. In other words, $P(x)$ is true if x has property P and $P(x)$ is false if x does not have property P. Hence we write

$$\{x \in X \mid P(x)\}$$

to represent the set of all x in the universal set X for which the statement $P(x)$ is true. When the universal set is clear from the context, we often simply write $\{x \mid P(x)\}$. In this notation, for example, the set of even natural numbers can be written

$$E = \{n \mid \text{There exists a natural number } k \text{ such that } n = 2k\}.$$

Here the property $P(n)$ is the statement, "There exists a natural number k such that $n = 2k$." $P(4)$ is a true statement since $4 = 2(2)$, and so $4 \in E$. On the other hand, $P(5)$ is false since there is no natural number k such that $5 = 2k$, and so $5 \notin E$.

We will also encounter sets of elements decribed by several properties. For example, the set

$$\{x \mid P_i(x),\ i = 1, \ldots, r\}$$

means the set of all elements for which the r distinct statements

$$P_1(x), P_2(x), \ldots, P_r(x)$$

are all true.

Set builder notation has a number of advantages. First, it gives a way to effectively describe very large or infinite sets without having to resort to lists or cumbersome patterns. For example, the set of rational numbers can be described as

$$\left\{x \;\middle|\; \text{There are integers } p \text{ and } q \text{ such that } x = \tfrac{p}{q}\right\}.$$

More importantly, the notation makes explicit the logical structure that is implicit in the language of sets. Since this structure underlies the entire development of mathematics as statements which can be proven according to the rigors of logic, we will emphasize this here.

One special set deserves mention. The *empty set*, denoted $\emptyset$, is the set with no elements. There is only one such set, although it may appear in many forms. For example, if for some set X, a statement $P(x)$ is *false* for all $x \in X$, then

$$\emptyset = \{x \mid P(x)\}\,.$$

We now consider relations and operations among sets. For this purpose, we suppose that we are given two sets A and B with the universal set X. We will suppose that both A and B are described in set builder notation

$$A = \{x \mid P_A(x)\}\,, \quad B = \{x \mid P_B(x)\}\,,$$

where P_A and P_B are properties describing the sets A and B respectively.

We will say that A *is a subset of* B, written $A \subset B$, if for all $a \in A$, we also have $a \in B$. Hence,

$$\boxed{\text{To prove } A \subset B\text{, show that if } P_A(x) \text{ is true, then } P_B(x) \text{ is true.}}$$

In fact, there are two standard ways of proving that $A \subset B$. In the direct method, one assumes that $P_A(x)$ is true and then deduces, by definitions and previously-proved statements, that $P_B(x)$ must also be true. In the indirect method, by contrast, one assumes that $P_A(x)$ is true and that $P_B(x)$ is false, and then attempts to derive a contradiction. This is what is known as "proof by contradiction."

It is a consequence of formal logic that the empty set is a subset of all sets: For any set X, $\emptyset \subset X$.

There are several basic set operations corresponding to the basic logical connectives "and," "or," and "not." The *intersection of* A *and* B, written $A \cap B$, is the set

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}\,.$$

We have:

$$\boxed{\text{To prove } x \in A \cap B\text{, show that } P_A(x) \text{ is true AND } P_B(x) \text{ is true.}}$$

Also, the *union of* A *and* B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}\,,$$

and

$$\boxed{\text{To prove } x \in A \cup B\text{, show that either } P_A(x) \text{ is true OR } P_B(x) \text{ is true.}}$$

The *difference of* A *and* B is the set

$$A \backslash B = \{x \mid x \in A \text{ and } x \notin B\}\,,$$

so

$$\boxed{\text{To prove } x \in A \backslash B\text{, show that } P_A(x) \text{ is true AND } P_B(x) \text{ is false.}}$$

In particular, the *complement of* A is the set $A^c = X \backslash A$. So

$$\boxed{\text{To prove } x \in A^c\text{, show that } P_A(x) \text{ is false.}}$$

There is yet another set operation which is of a somewhat different nature than the previous operations. The *Cartesian product of sets* A *and* B is a set $A \times B$ whose elements are ordered pairs of the form (a, b). More precisely,

$$A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\}\,, \text{ i.e.}$$

$$\boxed{\text{To prove } (x,y) \in A \times B, \text{ show that } P_A(x) \text{ is true AND } P_B(y) \text{ is true.}}$$

In this case, the separate entries in the ordered pair are referred to as *components*, and if there is need for specificity, the A-component and the B-component.

Notice the distinction between the logical statements corresponding to $A \cap B$, where statements $P_A(x)$ and $P_B(x)$ must be true about the *same* element x, whereas for $A \times B$, statements $P_A(x)$ and $P_B(y)$ must be true of *different* elements, one in A and the other in B.

Finally, we list several common sets of numbers along with the standard notation:

N, the set of natural numbers,

Z, the set of integers,

Q, the set of rational numbers,

R, the set of real numbers, and

C, the set of complex numbers.

We will be especially concerned with the set $\mathbf{R}^n = \mathbf{R} \times \cdots \times \mathbf{R}$:

$$\mathbf{R}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbf{R} \text{ for } i = 1, \ldots, n\}.$$

Many times we will encounter special subsets of the real numbers which are defined by the order relations $<$, $\leq$, $>$ and $\geq$. These are the *intervals*, and we use the special notation

$$\begin{aligned}
[a,b] &= \{r \in \mathbf{R} \mid a \leq r \leq b\}, \\
(a,b) &= \{r \in \mathbf{R} \mid a < r < b\}, \\
(-\infty,b) &= \{r \in \mathbf{R} \mid r < b\}, \\
(-\infty,b] &= \{r \in \mathbf{R} \mid r \leq b\}, \\
(a,\infty) &= \{r \in \mathbf{R} \mid r > a\}, \\
[a,\infty) &= \{r \in \mathbf{R} \mid r \geq a\},
\end{aligned}$$

where in all cases $a < b$. We can similarly define the half-open intervals $(a, b]$, etc.

0.2. Functions

Most readers will recall the definition of a function that is typically presented, for example, in a precalculus course. A function is defined to be a rule assigning to each element of one set (the domain) exactly one element of another set (the range). The advantage of this definition is that it emphasizes the relationship established between elements of the domain and those of the range by means of the rule. It has the disadvantage, however, of lacking mathematical precision, especially by relying on the imprecise term "rule."

In order to be more precise, mathematicians in the 1920s established the following definition of a function. Given two sets A and B, a function f is defined to be a subset $f \subset A \times B$ with the following two properties: first, for all $a \in A$ there is $b \in B$ such that $(a, b) \in f$ and second, if $(a_1, b_1) \in f$ and $(a_2, b_2) \in f$ with $a_1 = a_2$, then $b_1 = b_2$. The logical structure of this definition implies a clear distinction between the sets A and B, with a statement about elements of A ($a_1 = a_2$) implying a statement about elements of B ($b_1 = b_2$). The set A is called the *domain* and the

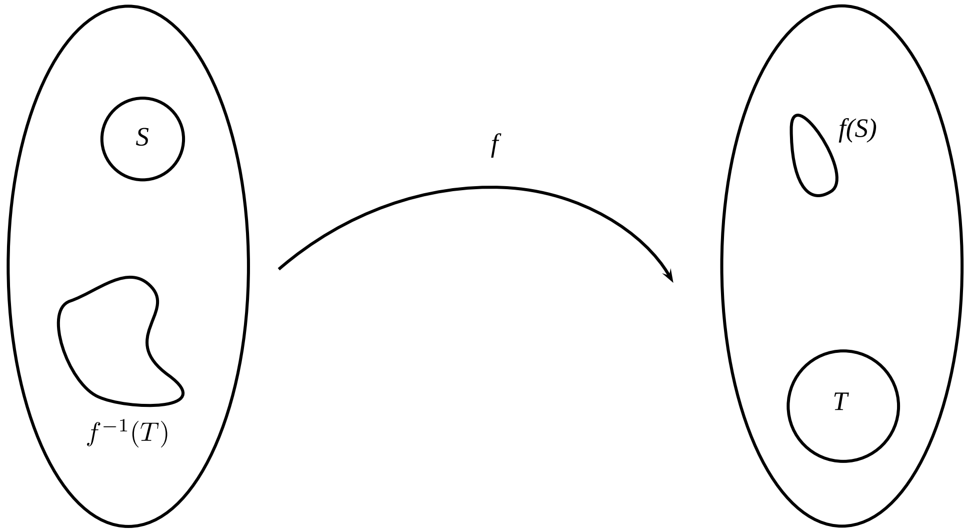

FIGURE 0.1. A representation of the image and preimage of a function f

set B is called the *codomain*. We will reserve the term "range" for a special subset of B defined below.

Despite this formal definition, throughout this text we will rely on the standard notation for functions. In particular, for a function f with domain A and codomain B, we use the notation

$$f : A \to B$$

to represent the function; the arrow points from the domain to the codomain. We write $f(a)$ to represent the unique element of B such that $(a, f(a)) \in f$. We will occasionally use the notation

$$a \mapsto f(a).$$

As a central concept in mathematics, a number of different terms have emerged to describe a function. We will use the words *function*, *map* and *transformation* interchangeably.

Given a function $f : A \to B$ and a set $S \subset A$, the *image of S (under f)* is defined to be the set

$$f(S) = \{y \in B \mid \text{There is } x \in S \text{ such that } f(x) = y\}.$$

The *range of f* is defined to be the set $f(A)$. For a subset $T \subset B$, the *preimage of T (under f)* is the (possibly empty) set

$$f^{-1}(T) = \{x \in A \mid f(x) \in T\}.$$

A function $f : A \to B$ is *onto* is $f(A) = B$, i.e. if the range of f coincides with the codomain of f. To demonstrate that a function is onto, it is necessary then to show that $B \subset f(A)$ (since $f(A) \subset B$ by definition). In other words, for any element $b \in B$, it is necessary to produce $a \in A$ such that $f(a) = b$.

A function $f : A \to B$ is *one-to-one* if for any two elements $a_1, a_2 \in A$, the condition $f(a_1) = f(a_2)$ implies that $a_1 = a_2$. There are two basic approaches to show that a function is one-to-one. The direct method is to assume that there are two elements a_1, a_2 with the property that $f(a_1) = f(a_2)$ and to show that this implies that $a_1 = a_2$. The indirect method is to suppose that there are two different elements $a_1, a_2 \in A$ such that $a_1 \neq a_2$ and to show that this implies that $f(a_1) \neq f(a_2)$.

Given any set A, there is always a distinguished function $\mathrm{Id}_A : A \to A$ defined by $\mathrm{Id}_A(a) = a$ for all $a \in A$. This function is called the *identity map of* A.

Suppose there are two functions $f : A \to B$ and $g : C \to D$ with the property that $f(A) \subset C$. Then it is possible to define a new function $g \circ f : A \to D$ by $(g \circ f)(a) = g(f(a))$. This function is called the *composition* of g with f. This operation on functions is associative, i.e.

$$(h \circ g) \circ f = h \circ (g \circ f),$$

assuming that all compositions are defined. A function $f : A \to B$ is one-to-one if and only if it has an *inverse* defined on the image of A, i.e. there is a function $g : f(A) \to A$ such that $f \circ g = \mathrm{Id}_{f(A)}$ and $g \circ f = \mathrm{Id}_A$. We normally write f^{-1} to denote the inverse of f.

In addition, for *real-valued* functions $f : A \to \mathbf{R}$, i.e. functions whose codomain is a set of real numbers, there are a number of operations inherited from the normal operations on real numbers. For example, if f and g are real-valued functions, then $f+g$, $f-g$, $f \cdot g$ and f/g are defined *pointwise*. For example, $f+g$ is the function whose domain is defined to be the intersection of the domains of f and g and whose value $(f+g)(a)$ is given by $f(a)+g(a)$.

CHAPTER 1

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 $f : \mathbf{R} \to \mathbf{R}$ 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 which seem to yield little new information about linear functions. Linear algebra is generally introduced after the basic calculus sequence has been 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.

1.1. Vector spaces

Modern mathematics can be described as the study of sets with 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 1.1.1.

A (real) vector space *consists of a set V together with two operations, addition and scalar multiplication. Scalars are understood here as real numbers. Elements of V are called* vectors *and will often be written in bold type $\mathbf{v}$. Addition is written using the conventional symbol $\mathbf{v}+\mathbf{w}$. Scalar multiplication is denoted as $s\mathbf{v}$ 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, *denoted* $\mathbf{0}$, *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 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 1.1.2. *Let* V *be a vector space. Then:*

(a) *The identity element* $\mathbf{0}$ *is unique.*
(b) *For all* $\mathbf{v} \in V$, *the additive inverse* $-\mathbf{v}$ *of* $\mathbf{v}$ *is unique.*
(c) *For all* $\mathbf{v} \in V$, $0 \cdot \mathbf{v} = \mathbf{0}$.
(d) *For all* $\mathbf{v} \in V$, $(-1) \cdot \mathbf{v} = -\mathbf{v}$.

Idea of proof. The proofs of (a) and (b) follow the standard method of proof of uniqueness, namely assuming to the contrary that there are two elements with the given properties and then showing that the two elements must in fact be equal. The proof of (c) follows by writing a scalar s as $s+0$ and applying axioms (V8) and (V4) in combination with the uniqueness results proved in (a). Finally, the proof of (d) is a consequence of writing $0 = 1 + (-1)$, axioms (V8) and (V10), along with the previous statements in this theorem.

□

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 these concepts arise naturally in the context of *inner product spaces*, which we treat in Section 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 1.1.3. *(*Euclidean space*) For any positive integer* n, *define the set* $\mathbf{R}^n$ *to be the set of all* n*-tuples of real numbers:*

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

Define vector addition componentwise as

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

and likewise define scalar multiplication

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

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

$\mathbf{R}^n$ can be thought of as the "model" finite-dimensional vector space in at least two senses. First, it is the most familiar example, 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 all finite-dimensional vector spaces are "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 1.1.4. *(*Vector spaces of functions*) Given any set* X*, let* $\mathcal{F}(X)$ *be the set of all real-valued functions with domain* X*:*

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

For any 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* $O : X \to \mathbf{R}$ *which 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.

1.2. 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 which are themselves vector spaces under the same operations of vector addition and scalar multiplication as in the larger set.

DEFINITION 1.2.1. *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 Figure 1.1.

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

THEOREM 1.2.2. *Suppose* $W \subset V$ *is a 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*.*

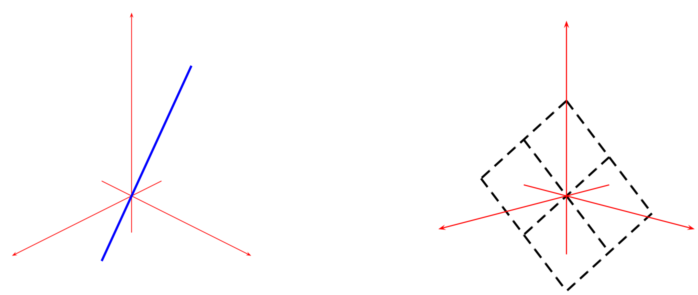

FIGURE 1.1. Subspaces in $\mathbf{R}^3$

IDEA OF PROOF. The reader will note that the closure axioms (W1)–(W2) correspond to the vector space axioms (V1) and (V6); these two properties are not consequences of the fact that W is a subset of V. Many of the rest of the vector space axioms, however, are simply consequences of considering the elements of W as vectors in V and relying on the fact that V is assumed to be a vector space. The only two axioms which require special attention are the existence axioms (V4) and (V5), verifying that $\mathbf{0}$ and $-\mathbf{w}$ are vectors *in* W. The existence of these special elements is a consequence of the fact that V is a vector space. To show that they are in fact elements of W requires Theorem 1.1.2. □

We note that for any 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 $\mathbf{R}^3$.

EXAMPLE 1.2.3. *Let $W_1 = \{(s,0,0) \mid s \in \mathbf{R}\}$. Then W_1 is a subspace of $\mathbf{R}^3$.*

EXAMPLE 1.2.4. *Let $\mathbf{v} = (a,b,c) \neq \mathbf{0}$ and let*

$$W_2 = \{s\mathbf{v} \mid s \in \mathbf{R}\}.$$

To see that W_2 is a subspace of $\mathbf{R}^3$, we rely on Theorem 1.2.2. Take $\mathbf{w}_1, \mathbf{w}_2 \in W_2$. That is, there are scalars $s_1, s_2 \in \mathbf{R}$ such that $\mathbf{w}_1 = s_1\mathbf{v}$ and $\mathbf{w}_2 = s_2\mathbf{v}$. To show that $\mathbf{w}_1 + \mathbf{w}_2 \in W_2$, it is necessary to find a scalar s such that $\mathbf{w}_1 + \mathbf{w}_2 = s\mathbf{v}$. The reader can check that $s = s_1 + s_2$ is the required scalar.

To show that W_2 is closed under scalar multiplication, choose $\mathbf{w} \in W_2$ and $s \in \mathbf{R}$. To show that $s\mathbf{w} \in W_2$ requires finding a scalar $u \in \mathbf{R}$ such that $s\mathbf{w} = u\mathbf{v}$. But since $\mathbf{w} \in W_2$, there is a scalar t such that $\mathbf{w} = t\mathbf{v}$. So choose $u = st$.

Note that Example 1.2.3 is a special case of this example when $\mathbf{v} = (1,0,0)$.

EXAMPLE 1.2.5. *Let $W_3 = \{(s,t,0) \mid s,t \in \mathbf{R}\}$. Then W_3 is a subspace of $\mathbf{R}^3$.*

EXAMPLE 1.2.6. *As in Example 1.2.4, let $\mathbf{v} = (a,b,c) \neq \mathbf{0}$. Relying on the usual "dot product" in $\mathbf{R}^3$, define*

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

Choose $\mathbf{x}_1, \mathbf{x}_2 \in W_4$*; that means that* $\mathbf{v} \cdot \mathbf{x}_1 = 0$ *and* $\mathbf{v} \cdot \mathbf{x}_2 = 0$*. Choose* $s \in \mathbf{R}$*. To show that* $\mathbf{x}_1 + \mathbf{x}_2$ *and* $s\mathbf{x}_1$ *are in* W_4*, we need to show that* $\mathbf{v} \cdot (\mathbf{x}_1 + \mathbf{x}_2) = 0$ *and that* $\mathbf{v} \cdot (s\mathbf{x}_1) = 0$*. Both of these follow from the properties of the dot product and the fact that* $\mathbf{x}_1, \mathbf{x}_2 \in W_4$*.*

Note that Example 1.2.5 is a special case of this example when $\mathbf{v} = (0, 0, 1)$*.*

We will show at the end of Section 4 that all proper, nontrivial subspaces of $\mathbf{R}^3$ can be realized either in the form of W_2 or W_4.

Example 1.2.7 (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* $f : \mathbf{R} \to \mathbf{R}$*. 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 which 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 following section will be to exhibit a method for constructing vector subspaces of any vector space V.

1.3. 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.

Definition 1.3.1. *Suppose* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *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* $c_1, \ldots, c_n$ *such that*

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

A basic question in a first course in linear algebra is: Given a vector $\mathbf{w}$ and a set S as in Definition 1.3.1, decide whether or not $\mathbf{w}$ is a linear combination of S. In practice, this can be answered using the tools of matrix algebra.

Example 1.3.2. *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 this definition to obtain a method for constructing subspaces.

DEFINITION 1.3.3. *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* $\mathrm{Span}(S)$, *is defined to be the set of all linear combinations of* S*:*

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

We note immediately the utility of this construction.

THEOREM 1.3.4. *Let* $S \subset V$ *be a finite set of vectors. Then* $W = \mathrm{Span}(S)$ *is a subspace of* V.

PROOF. The proof is an immediate application of Theorem 1.2.2. □

EXAMPLE 1.3.5. *Let* $S = \{\mathbf{0}\}$. *Then* $\mathrm{Span}(S) = S$ *is the trivial subspace.*

EXAMPLE 1.3.6. *Let*

$$S = \{\mathbf{v}_1\} \subset \mathbf{R}^3,$$

where $\mathbf{v}_1 = (1, 0, 0)$. *Then* $\mathrm{Span}(S) = \{s(1, 0, 0) \mid s \in \mathbf{R}\} = \{(s, 0, 0) \mid s \in \mathbf{R}\}$. *Compare to Example 1.2.3.*

EXAMPLE 1.3.7. *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*

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

EXAMPLE 1.3.8. *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*

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

EXAMPLE 1.3.9. *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*

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

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

EXAMPLE 1.3.10. *Note that the set of four vectors S in $\mathbf{R}^3$ from Example 1.3.9* 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 $\mathbf{w} \in \mathrm{Span}(S)$, 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 any vector $\mathbf{w}$ such that $w_1 + w_2 - 2w_3 \neq 0$, the system is not consistent and $\mathbf{w} \notin \mathrm{Span}(S)$. For example $(1, 1, 2) \notin \mathrm{Span}(S)$.

Given a subspace W, a finite set S is said to be a *spanning set for W* if $\mathrm{Span}(S) = W$. 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 1.3.8, we exhibited a different spanning set for $\mathbf{R}^3$.

1.4. Linear independence, basis, and dimension

In the preceding section, we started from a finite set $S \subset V$ in order to generate a subspace $W = \mathrm{Span}(S)$ in V. This procedure prompts the following question: Given 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.

Given 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 seen the question of whether or not $\mathbf{w} \in \mathrm{Span}(S)$. Intuitively, we might say that $\mathbf{w}$ "depends linearly" on S if $\mathbf{w} \in \mathrm{Span}(S)$, 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 is $\mathbf{w}$ is not a scalar multiple of $\mathbf{v}$.

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

DEFINITION 1.4.1. *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*

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

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 $s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n = \mathbf{0}$ implies that

$$s_1 = \cdots = s_n = 0.$$

EXAMPLE 1.4.2. *We refer back to the set of four vectors $S \subset \mathbf{R}^3$ in Example 1.3.9; see also Example 1.3.10. We will show that the set S is linearly dependent: 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:

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

Gaussian elimination of the corresponding augmented matrix

$$\left[\begin{array}{cccc:c} 1 & -1 & 1 & -3 & 0 \\ 1 & 1 & 3 & 1 & 0 \\ 1 & 0 & 2 & -1 & 0 \end{array}\right]$$

yields

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

In other words, introducing free variables t, u, the 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 1.4.2 illustrates the fact that deciding whether a set is linearly dependent or independent amounts to deciding whether or not a corresponding *homogeneous* system of linear equations has nontrivial solutions or only the trivial solution.

The following facts are immediate from Definition 1.4.1. The reader is invited to supply proofs.

THEOREM 1.4.3. *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) *S is a linearly dependent set of nonzero vectors if and only if one vector in S 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 1.4.4. *Let S be a finite set of vectors in a vector space V. Let $\mathbf{w} \in \mathrm{Span}(S)$, and let $S' = S \cup \{\mathbf{w}\}$. Then $\mathrm{Span}(S') = \mathrm{Span}(S)$.*

PROOF. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and suppose $\mathbf{w} = s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n$. Showing that two sets are equal requires showing that each one is a subset of the other. In this case, we show that if $\mathbf{v} \in \mathrm{Span}(S')$ then $\mathbf{v} \in \mathrm{Span}(S)$ and conversely that if $\mathbf{v} \in \mathrm{Span}(S)$, then $\mathbf{v} \in \mathrm{Span}(S')$.

First assume that $\mathbf{v} \in \mathrm{Span}(S')$. Then there are scalars $c_1, \ldots, c_{n+1}$ such that

$$\begin{aligned}\mathbf{v} &= c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n + c_{n+1}\mathbf{w} \\ &= c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n + c_{n+1}(s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n) \\ &= (c_1 + c_{n+1}s_1)\mathbf{v}_1 + \cdots + (c_n + c_{n+1}s_n)\mathbf{v}_n\end{aligned}$$

and so $\mathbf{v} \in \mathrm{Span}(S)$.

Now assume that $\mathbf{v} \in \mathrm{Span}(S)$. Then there are scalars $c_1, \ldots, c_n$ such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots c_n\mathbf{v}_n$. Rewriting this as

$$\begin{aligned}\mathbf{v} &= (c_1 - s_1)\mathbf{v}_1 + \cdots + (c_n - s_n)\mathbf{v}_n + (s_1\mathbf{v}_1 + \cdots s_n\mathbf{v}_n) \\ &= (c_1 - s_1)\mathbf{v}_1 + \cdots + (c_n - s_n)\mathbf{v}_n + \mathbf{w}\end{aligned}$$

shows that $\mathbf{v} \in \mathrm{Span}(S')$.

Hence $\mathrm{Span}(S') = \mathrm{Span}(S)$.

□

Generating "larger" subspaces requires adding *linearly independent* vectors to the set of vectors from which we create the spanning set.

We return to a version of the question at the outset of this section: Given a subspace, what is the "smallest" subset which can serve as a spanning set for this subspace? This motivates the definition of a *basis*.

DEFINITION 1.4.5. *Let V be a vector space. A* basis *of V is a set $B \subset V$ such that*

- $\mathrm{Span}(B) = V$*; and*
- *B is a linearly independent set.*

EXAMPLE 1.4.6. *For the vector space $V = \mathbf{R}^n$, the set $B_0 = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, \ldots, 0, 1)$ is a basis for $\mathbf{R}^n$. B_0 is called the* standard basis *for $\mathbf{R}^n$. We leave it as an exercise to check that B_0 has the two properties of a basis.*

EXAMPLE 1.4.7. *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

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

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 only has 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* $\mathbf{R}^3$*.*

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 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 is a linearly independent set has an important consequence.

THEOREM 1.4.8. *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 $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, so that there are scalars $c_1, \ldots, c_n$ and $d_1, \ldots, d_n$ 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

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

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

□

The discussion in Example 1.4.7 hints at a powerful technique for determining whether a set of vectors in $\mathbf{R}^n$ forms a basis for $\mathbf{R}^n$.

THEOREM 1.4.9. *A set of* n *vectors* $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbf{R}^n$ *forms a basis for* $\mathbf{R}^n$ *if an only if* $\det(A) \neq 0$*, where* $A = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ *is the matrix formed by the column vectors* $\mathbf{v}_i$*.*

PROOF. See the argument in Example 1.4.7.

□

The preceding theorem, in turn, is a reflection of a more general fact, although the proof is quite similar to the discussion in Example 1.4.7.

THEOREM 1.4.10. *Let* V *be a vector space and let* B *be a basis for* V *which consists of* n *vectors. Then no set with less than* n *vectors spans* V*, and no set with more than* n *vectors is linearly independent.*

PROOF. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, and let $A = [a_{ij}]$, where, for $j = 1, \ldots, m$,

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

In other words, A is the matrix of components of the vectors $\mathbf{v}_i$ relative to the basis B. Then if $m < n$, the row echelon form of A has a row of zeros, so that there are vectors $\mathbf{w} \in \mathbf{R}^n$ such that the vector equation $A\mathbf{c} = \mathbf{w}$, where $\mathbf{c} = (c_1, \ldots, c_m)$, has no solution. Hence S cannot span V.

Likewise, if $m > n$, then the row echelon form of A must have at most n leading ones, and so the vector equation $A\mathbf{c} = \mathbf{0}$ has nontrivial solutions. Hence S cannot be linearly independent.

□

COROLLARY 1.4.11. *Let V be a vector space and let B be a basis for V which consists of n vectors. Then every other basis B' of V must also have n elements.*

DEFINITION 1.4.12. *Let V be a vector space. If there is no finite subset of V which 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 1.4.11, every basis has n vectors), then V is* finite-dimensional, *We call n the* dimension *of V and we write $\dim(V) = n$.*

We say that the trivial vector space $\{\mathbf{0}\}$ is zero-dimensional:

$$\dim(\{\mathbf{0}\}) = 0.$$

Most of the examples we consider here will be finite dimensional. However, of the vector spaces listed in Example 1.2.7, 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 1.4.12 makes sense in this context.

THEOREM 1.4.13. *Let V be a finite-dimensional vector space and let W be a subspace of V. Then W has finite dimension. Further, $\dim(W) \leq \dim(V)$, with $\dim(W) = \dim(V)$ if and only if $W = V$.*

PROOF. Exercise.

□

EXAMPLE 1.4.14. *Recall $W_2 \subset \mathbf{R}^3$ from Example 1.2.4:*

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

where $(a, b, c) \neq \mathbf{0}$. Clearly $W_2 = \operatorname{Span}(\{(a, b, c)\})$ and the set $\{(a, b, c)\}$ is linearly independent by Theorem 1.4.3, so $\dim(W_2) = 1$.

EXAMPLE 1.4.15. *Recall $W_4 \subset \mathbf{R}^3$ from Example 1.2.6:*

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

for some $(a, b, c) \neq \mathbf{0}$. Assume without loss of generality that $a \neq 0$. Then W_4 can be seen, for example by Gaussian elimination and the introduction of free variables, 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 1.4.16. *We now justify the statement at the end of Section 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 1.4.17. *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 1.3.9. We already showed that* S *cannot be a basis for* $\mathbf{R}^3$: *In Example 1.4.2 we showed that* S *is linearly dependent, and in Example 1.3.10 we showed that* S *does not span* $\mathbf{R}^3$. *A closer look at Example 1.4.2 shows that the rank of the matrix* $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix}$ *is two, since the row-echelon form of* A *after Gaussian elimination has two leading ones. A basis for* $W = \operatorname{Span}(S)$ *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* $\dim(W) = 2$.

1.5. Linear transformations

Given a a set along with some extra structure, the next notion to consider are functions between the sets that in some suitable sense "preserve 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 Section 0.2 on functions in order to review the terminology and relevant definitions.

DEFINITION 1.5.1. *A function* $T : V \to W$ *is a* linear transformation *if*

(1) *For all* $\mathbf{u}, \mathbf{v} \in V$, $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;
(2) *For all* $s \in \mathbf{R}$ *and* $\mathbf{v} \in V$, $T(s\mathbf{v}) = sT(\mathbf{v})$.

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."

EXAMPLE 1.5.2. *Consider the map* $T : \mathbf{R}^3 \to \mathbf{R}^2$ *given by*

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

To show that T *is a linear transformation, choose arbitrary vectors* $\mathbf{u}, \mathbf{v} \in \mathbf{R}^3$, *where* $\mathbf{u} = (x_1, y_1, z_1)$ *and* $\mathbf{v} = (x_2, y_2, z_2)$, *and let* $s \in \mathbf{R}$ *be an arbitrary scalar.*

We have

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T(x_1 + x_2, y_1 + y_2, z_1 + z_2) \\ &= \big(2x_1 + 2x_2 + y_1 + y_2 - z_1 - z_2, x_1 + x_2 + 3z_1 + 3z_2\big) \end{aligned}$$

and

$$\begin{aligned} T(\mathbf{u}) + T(\mathbf{v}) &= (2x_1 + y_1 - z_1, x_1 + 3z_1) + (2x_2 + y_2 - z_2, x_2 + 3z_2) \\ &= (2x_1 + y_1 - z_1 + 2x_2 + y_2 - z_2, x_1 + 3z_1 + x_2 + 3z_2) \end{aligned}$$

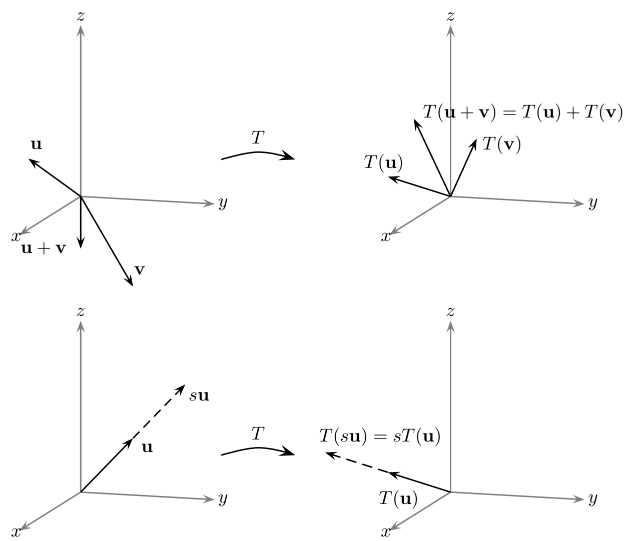

FIGURE 1.2. The two conditions defining a linear transformation.

Comparing components verifies property (1) of Definition 1.5.1.

Likewise,

$$\begin{aligned} T(s\mathbf{u}) &= T(sx_1, sy_1, sz_1) \\ &= \big(2(sx_1) + (sy_1) - (sz_1), (sx_1) + 3(sz_1)\big) \end{aligned}$$

and

$$\begin{aligned} sT(\mathbf{u}) &= s(2x_1 + y_1 - z_1, x_1 + 3z_1) \\ &= (2sx_1 + sy_1 - sz_1, sx_1 + 3sz_1) \end{aligned}$$

This verifies property (2) of Definition 1.5.1, which together with the prior calculation, shows that T is a linear transformation.

We recall some elementary properties of linear transformations which are consequences of Definition 1.5.1.

THEOREM 1.5.3. *Let V and W be vector spaces with corresponding zero vectors $\mathbf{0}_V$ and $\mathbf{0}_W$. Let $T : V \to W$ be a linear transformation. Then*

(a) $T(\mathbf{0}_V) = \mathbf{0}_W$*;*

(b) *For all $\mathbf{u} \in V$, $T(-\mathbf{u}) = -T(\mathbf{u})$.*

IDEA OF PROOF. Keeping in mind Theorem 1.1.2, both of these statements are consequences of the scalar multiplication condition in Definition 1.5.1, using $s = 0$ and $s = -1$ respectively.

□

It is worth stating the contrapositive of part (a) of Theorem 1.5.3 separately as a corollary.

COROLLARY 1.5.4. *If $T(\mathbf{0}) \neq \mathbf{0}$, then T is not a linear transformation.*

This shows in particular that the traditional definition of linear functions $f : \mathbf{R} \to \mathbf{R}$, i.e. polynomial functions of degree one, are not necessarily linear transformations in the sense of linear algebra, thinking of $\mathbf{R} = \mathbf{R}^1$ with the usual vector space structure. For example, $f(x) = 2x + 1$ is not a linear transformation since $f(0) = 1$. From the perspective of linear algebra, a real-valued function of one variable is a linear transformation if it has the form $f(x) = ax$. Graphs of linear transformations are lines through the origin. We will see this again in the following section.

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 1.5.5. *Suppose V and W are vector spaces. A linear transformation $T : V \to W$ is a* linear isomorphism *if it is one-to-one and onto.*

Theorem 1.5.6. *Suppose that V and W are finite-dimensional vector spaces, and suppose there is a linear isomorphism $T : V \to W$. Then* $\dim V = \dim W$.

Idea of 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. □

1.6. 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 generate 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 the basic correspondence between $(m \times n)$ matrices and linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$.

Theorem 1.6.1. *Every linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ 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 $\mathbf{x}$ is understood as a $n \times 1$ column vector. Conversely, every $m \times n$ matrix A gives rise to a linear transformation $T_A : \mathbf{R}^n \to \mathbf{R}^m$ by defining $T_A(\mathbf{x}) = A\mathbf{x}$.*

The proof of the first, main, statement of this theorem will emerge in the course of this section. The second statement, however, is obvious from the basic properties of matrix multipliciation.

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:

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

We also note from the outset that the matrix representation of a linear transformation is not unique, as 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.

Example 1.6.2. *The linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ given by*

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

(see Example 1.5.2) can be expressed as $T(\mathbf{x}) = A_T\mathbf{x}$*, where*

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

EXAMPLE 1.6.3. *The identity transformation* $\mathrm{Id} : \mathbf{R}^n \to \mathbf{R}^n$ *given by*

$$\mathrm{Id}(\mathbf{x}) = \mathbf{x}$$

is a linear transformation whose standard matrix representation is given by

$$A_{\mathrm{Id}} = I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

the $n \times n$ *identity matrix.*

EXAMPLE 1.6.4. *The zero transformation* $Z : \mathbf{R}^n \to \mathbf{R}^m$ *given by* $Z(\mathbf{x}) = \mathbf{0}$ *for all* $\mathbf{x} \in \mathbf{R}^n$ *is also a linear transformation. It can be expressed as* $Z(\mathbf{x}) = O\mathbf{x}$*, where* O *is the* $m \times n$ *matrix all of whose entries are* 0*.*

The second theorem on its face gives a far more general method for constructing linear transormations, in the sense that it applies to the setting of linear transformations between arbitrary 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 1.5.6.

THEOREM 1.6.5. *Let* V *be a finite-dimensional vector space with basis* $B = \{\mathbf{e}_i, \dots, \mathbf{e}_n\}$*. Let* W *be a vector space, and let* $B' = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ *be any set of* n *vectors in* W*, not necessarily distinct. Then there is a unique linear transformation* $T : V \to W$ *such that* $T(\mathbf{e}_i) = \mathbf{w}_i$ *for* $i = 1, \dots, n$*.*

If the set B' *is in fact a basis for* W*, then* T *is a linear isomorphism.*

IDEA OF PROOF. By Theorem 1.4.8, 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 $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$. Then define $T(\mathbf{v}) = v_1\mathbf{w}_1 + \cdots + v_n\mathbf{w}_n$ and check that it satifies the required properties.

If B' 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 $\mathbf{w} = s_1\mathbf{w}_1 + \cdots + s_n\mathbf{w}_n$, then the vector $\mathbf{v} = s_1\mathbf{e}_1 + \cdots + s_n\mathbf{e}_n$ can be shown to be the unique vector such that $T(\mathbf{v}) = \mathbf{w}$.

□

EXAMPLE 1.6.6. *The reader may confirm that* $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ *where* $\mathbf{e}_1 = (-1, 1)$ *and* $\mathbf{e}_2 = (2, 1)$ *is a basis for* $\mathbf{R}^2$*. Define a linear transformation* $T : \mathbf{R}^2 \to \mathbf{R}^4$ *in the manner of Theorem 1.6.5 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:*

$$\begin{aligned} T(\mathbf{v}) &= c_1T(\mathbf{e}_1) + c_2T(\mathbf{e}_2) \\ &= \frac{1}{3}\big(-v_1+2v_2\big)\Big(1,2,3,4\Big) + \frac{1}{3}(v_1+v_2)\Big(-2,-4,-6,-8\Big) \\ &= \frac{1}{3}\big(-3v_1,-6v_1,-9v_1,-12v_1\big) \\ &= (-v_1,-2v_1,-3v_1,-4v_1). \end{aligned}$$

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 1.6.5 illustrated in Example 1.6.6 gives rise to a general method of representing linear transformations between general vector spaces as matrices, as we did in the case of Euclidean spaces in Theorem 1.6.1.

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

$$\begin{aligned} T(\mathbf{e}_1) &= a_{11}\mathbf{e}'_1 + \cdots + a_{1m}\mathbf{e}'_m \\ &\;\;\vdots \\ T(\mathbf{e}_n) &= a_{n1}\mathbf{e}'_1 + \cdots + a_{nm}\mathbf{e}'_m \end{aligned} \tag{1.1}$$

It is a straightforward exercise to verify that if $\mathbf{x} \in V$ where $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and if $\mathbf{y} = T(\mathbf{x}) = y_1\mathbf{e}'_1 + \cdots + y_m\mathbf{e}'_m$, then $\mathbf{y} = A\mathbf{x}$, where $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, and $A = [a_{ij}]$ where a_{ij} are given in (1.1) above. A is called *the matrix of* T *relative to the bases* B, B' and will be denoted $A = [T]_{B',B}$.

Example 1.6.7. *Recall again the linear transformation* $T : \mathbf{R}^3 \to \mathbf{R}^2$ *from Example 1.5.2 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*

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

This is precisely the calculation we performed in Example 1.6.6. 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*

$$\left[\begin{array}{cc:ccc} -1 & 2 & 2 & 3 & 2 \\ 1 & 1 & 1 & 1 & 4 \end{array}\right],$$

yielding

$$\left[\begin{array}{cc:ccc} 1 & 0 & 0 & -1/3 & 2 \\ 0 & 1 & 1 & 4/3 & 2 \end{array}\right].$$

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 1.6.7 with the matrix for the same T in Example 1.6.2 illustrates the dependence of the matrix for T on the bases involved. In particular, it illustrates the comment immediately following Theorem 1.6.1, that the matrix representation of a linear transformation is not unique.

Second, Theorem 1.6.5 in fact provides the proof for Theorem 1.6.1. The standard matrix representation of a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ is obtained by applying Theorem 1.6.5 using the standard bases for $\mathbf{R}^n$ and $\mathbf{R}^m$.

1.7. Constructing subspaces II: Subspaces and linear transformations

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

DEFINITION 1.7.1. *The* kernel of T, $\ker(T) \subset V$, *is defined to be the set*

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

DEFINITION 1.7.2. *The* range of T, $R(T) \subset W$, *is defined to be the set*

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

THEOREM 1.7.3. *Let* $T : V \to W$ *be a linear tranformation. Then the sets* $\ker(T)$ *and* $R(T)$ *are subspaces of* V *and* W *respectively.*

PROOF. By Theorem 1.2.2, we need only show that the two subsets are closed under vector addition and scalar multiplication. We will illustrate one closure axiom for each of the two subsets and leave the other as an exercise for the reader. For example, to show that $\ker(T)$ is closed under vector addition, choose $\mathbf{v}_1, \mathbf{v}_2 \in \ker(T)$, i.e. $T(\mathbf{v}_1) = \mathbf{0}$ and $T(\mathbf{v}_2) = \mathbf{0}$. Then $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$, and so $\mathbf{v}_1 + \mathbf{v}_2 \in \ker(T)$.

Likewise, to show that $R(T)$ is closed under scalar multiplication, choose $\mathbf{w} \in R(T)$ and take $s \in \mathbf{R}$. To show that $s\mathbf{w} \in R(T)$, we need to produce $\mathbf{v} \in V$ such that $T(\mathbf{v}) = s\mathbf{w}$. But since $\mathbf{w} \in R(T)$, there is $\mathbf{v}_0 \in V$ such that $T(\mathbf{v}_0) = \mathbf{w}$. Take $\mathbf{v} = s\mathbf{v}_0$, so that $T(\mathbf{v}) = T(s\mathbf{v}_0) = sT(\mathbf{v}_0) = s\mathbf{w}$. Hence $s\mathbf{w} \in R(T)$, and so $R(T)$ is closed under scalar multiplication.

□

We present an example of a standard technique for finding a basis for the kernel of a linear transformation.

EXAMPLE 1.7.4. *Let* $T : \mathbf{R}^3 \to \mathbf{R}$ *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 1.2.6. Suppose that $a \neq 0$*. Introducing free variables, we can write* $x = (-b/a)s + (-c/a)t$*,* $y = s$*, and* $z = t$*, so that*

$$\ker(T) = \left\{ \Big(-\frac{sb}{a} - \frac{tc}{a}, s, t \Big) \mid s, t \in \mathbf{R} \right\}.$$

Note that for $\mathbf{x} \in \ker(T)$*, we can write* $\mathbf{x} = s\big((-b/a), 1, 0\big) + t\big((-c/a), 0, 1\big)$*. In other words,* $\mathbf{x}$ *can be written as a linear combination of the vectors* $\mathbf{b}_1 = (-b/a, 1, 0)$ *and* $\mathbf{b}_2 = (-c/a, 0, 1)$*. The reader can verify that* $S = \{\mathbf{b}_1, \mathbf{b}_2\}$ *is linearly independent, a reflection of the fact that* s *and* t *were independent free variables. Hence* S *is a basis for* $\ker(T)$ *and so* $\dim(\ker(T)) = 2$*.*

For a linear transformation $T : V \to W$, 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 $R(T) = W$. Recall that T is a *one-to-one* function if $T(\mathbf{x}_1) = T(\mathbf{x}_2)$ implies that $\mathbf{x}_1 = \mathbf{x}_2$.

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

EXAMPLE 1.7.5. *Consider euclidean spaces* $\mathbf{R}^n$*,* $\mathbf{R}^m$ *for* $m < n$*. The projection map* $\mathrm{Pr} : \mathbf{R}^n \to \mathbf{R}^m$*, given by*

$$\mathrm{Pr}(x_1, \ldots, x_n) = (x_1, \ldots, x_m),$$

is a linear transformation that is onto but not one-to-one. The inclusion map $\mathrm{In} : \mathbf{R}^m \to \mathbf{R}^n$ *given by*

$$\mathrm{In}(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$$

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

There is a powerful characterization of one-to-one linear transformations which has no parallel for general functions. We prove the theorem below in detail; the reader is encouraged to pay close attention to where the fact that T is a linear transformation comes into play.

THEOREM 1.7.6. *A linear transformation* $T : V \to W$ *is one-to-one if and only if* $\ker(T) = \{\mathbf{0}\}$*.*

PROOF. Fiirst, assume T is one-to-one. By Theorem 1.5.3, $T(\mathbf{0}) = \mathbf{0}$ and so $\mathbf{0} \in \ker(T)$. Moreover, if $\mathbf{v} \in \ker(T)$ then

$$T(\mathbf{v}) = \mathbf{0} = T(\mathbf{0})$$

and so $\mathbf{v} = \mathbf{0}$, since T is one-to-one. Hence $\ker(T) = \{\mathbf{0}\}$.

Conversely, assume now that $\ker(T) = \{\mathbf{0}\}$. Suppose that $T(\mathbf{x}_1) = T(\mathbf{x}_2)$. Then

$$\mathbf{0} = T(\mathbf{x}_1) - T(\mathbf{x}_2) = T(\mathbf{x}_1 - \mathbf{x}_2)$$

which says that $\mathbf{x}_1 - \mathbf{x}_2 \in \ker(T) = \{\mathbf{0}\}$. So $\mathbf{x}_1 = \mathbf{x}_2$ and T is one-to-one.

□

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

THEOREM 1.7.7. *Let $T : V \to W$ be a linear transformation, where V is finite dimensional. Then*

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

PROOF. The proof involves a standard technique in linear algebra of *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 $\dim(R(T)) \le n$.

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.

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

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

forms a basis for V, and second, that the set $\left\{\mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n\right\}$ forms a basis for $\ker(T)$. We illustrate the first step of this process. Choose $\mathbf{b}_{r+1} \notin \operatorname{Span}\{\mathbf{e}'_1, \dots, \mathbf{e}'_r\}$. 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 $\mathbf{e}'_{r+1} \in \ker(T)$. Add $\mathbf{e}'_{r+1}$ to E. Repeated application of this process yields E'. We leave the verification that $\left\{\mathbf{e}'_{r+1}, \dots, \mathbf{e}'_n\right\}$ forms a basis for $\ker(T)$ to the reader.

$\square$

We will frequently refer to the dimension of the range of a linear transformation, or, alternately, to the rank of the matrix considered as a linear transformation:

DEFINITION 1.7.8. *The* rank of a linear transformation $T : V \to W$ *is defined to be the dimension of* $R(T)$.

The following example illustrates both the statement of the theorem and the notion of completing a basis used in the proof of Theorem 1.7.7.

EXAMPLE 1.7.9. *Let V be a vector space with dimension n and let W be a subspace of V with dimension r. 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 $\mathrm{Pr}_{B',B} : V \to V$ as follows: For any vector $\mathbf{v} \in V$, there are unique scalars $v_1, \dots, v_n$ such that $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$. Define

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

We leave it as an exercise to show that $\mathrm{Pr}_{B',B}$ is a linear transformation. Clearly $W = R(\mathrm{Pr}_{B',B})$ and so $\dim(R(\mathrm{Pr}_{B',B})) = r$. Theorem 1.7.7 then implies that $\dim(\ker(\mathrm{Pr}_{B',B})) = n - r$, a fact which is also seen by noting that $\{\mathbf{e}_{r+1}, \dots, \mathbf{e}_n\}$ is a basis for $\ker(\mathrm{Pr}_{B',B})$.

As the notation implies, the map $\mathrm{Pr}_{B',B}$ 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 1.7.5 above.

Theorem 1.7.7 has a number of important corollaries for finite-dimensional vector spaces. We leave the proofs to the reader.

Corollary 1.7.10. *Let $T : V \to W$ be a* one-to-one *linear transformation between finite-dimensional vector spaces. Then*

$$\dim(V) \leq \dim(W).$$

Corollary 1.7.11. *Let $T : V \to W$ be an* onto *linear transformation between finite-dimensional vector spaces. Then*

$$\dim(V) \geq \dim(W).$$

Note that these two corollaries combined yield a proof of Theorem 1.5.6.

Finally, we mention a partial converse to Theorem 1.5.6 which holds for linear transformations between finite-dimensional vector spaces, although not to more general functions or to linear transformations between infinite-dimensional vector spaces.

Corollary 1.7.12. *Let $T : V \to W$ be a linear transformation with* $\dim(V) = \dim(W) < \infty$*. Then T is one-to-one if and only if T is onto.*

The concept of linear isomorphism gives a first 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 a geometric context.

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 $T : V \to W$, the inverse $T^{-1} : W \to V$ is defined on all of W, where $T \circ T^{-1} = \mathrm{Id}_W$ and $T^{-1} \circ T = \mathrm{Id}_V$ (where Id_V and Id_W are the identity transformations of V and W described in Example 1.6.3). We summarize this in the following theorem.

Theorem 1.7.13. *Let $T : V \to W$ be a linear isomorphism. Then there is a unique linear isomorphism $T^{-1} : W \to V$ such that $T \circ T^{-1} = \mathrm{Id}_W$ and $T^{-1} \circ T = \mathrm{Id}_V$. Moreover, if B is a basis for V and B' is a basis for W, then $\left[T^{-1}\right]_{B,B'} = [T]^{-1}_{B',B}$, using the notation of Theorem 1.6.1.*

Proof. The most important fact to be proved in this theorem is that the inverse of a linear transformation, which exists purely on set-theoretic grounds, is in fact a linear transformation. We leave this as an exercise.

□

Theorem 1.5.6 shows that if two vector spaces are isomorphic, then they have the same dimension. The converse is also true, again only for finite-dimensional vector spaces.

Theorem 1.7.14. *Let V and W be vector spaces with same finite dimension n. Then V and W are isomorphic.*

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for V and let $\{\mathbf{e}'_1, \ldots, \mathbf{e}'_n\}$ be a basis for W. Define a linear transformation according to Theorem 1.6.5 by $T(\mathbf{e}_i) = \mathbf{e}'_i$. It is an exercise to show that this T is one-to-one and hence, by Corollary 1.7.12, onto.

□

The above theorem justifies the statement following Example 1.1.3: 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 1.7.15. *Let $T : V \to W$ be a linear transformation between n-dimensional vector spaces and let B be a basis for V and B' be a basis for W. Let $A = [T]_{B',B}$ be the matrix representing T relative to B, B'. Then T is a linear isomorphism if and only if* $\det(A) \neq 0$.

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

THEOREM 1.7.16. *Let V be a finite-dimensional vector space, and let $T : V \to V$ 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

$$[T]_{B_2,B_2} = [I]_{B_2,B_1} [T]_{B_1,B_1} [T]_{B_1,B_2} ,$$

and that $[I]_{B_2,B_1} = [I]^{-1}_{B_1,B_2}$, where $I : V \to V$ is the identity transformation. □

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

1.8. 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.

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

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

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

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 1.1.4. 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 1.8.1. *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 any $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have

$$\begin{aligned}(T_1+T_2)(\mathbf{v}_1+\mathbf{v}_2) &= T_1(\mathbf{v}_1+\mathbf{v}_2)+T_2(\mathbf{v}_1+\mathbf{v}_2)\\ &= (T_1(\mathbf{v}_1)+T_1(\mathbf{v}_2))+(T_2(\mathbf{v}_1)+T_2(\mathbf{v}_2)) \quad T_1,\ T_2 \text{ are linear}\\ &= (T_1(\mathbf{v}_1)+T_2(\mathbf{v}_1))+(T_1(\mathbf{v}_2)+T_2(\mathbf{v}_2))\\ &= (T_1+T_2)(\mathbf{v}_1)+(T_1+T_2)(\mathbf{v}_2).\end{aligned}$$

Also, for any $c \in \mathbf{R}$ and $\mathbf{v} \in V$,

$$\begin{aligned}(T_1+T_2)(c\mathbf{v}) &= T_1(c\mathbf{v})+T_2(c\mathbf{v})\\ &= cT_1(\mathbf{v})+cT_2(\mathbf{v}) \quad T_1,\ T_2 \text{ are linear}\\ &= c(T_1(\mathbf{v})+T_2(\mathbf{v}))\\ &= c(T_1+T_2)(\mathbf{v}).\end{aligned}$$

Hence $T_1+T_2 \in V^*$. The fact that sT_1 is also linear for any $s \in \mathbf{R}$ is proved similarly. Note that the zero covector $O \in V^*$ is defined by $O(\mathbf{v}) = 0$ for all $\mathbf{v} \in V$. □

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

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

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

Define covectors $E_i \in V^*$ by how they act on the basis B according to Theorem 1.6.5: $E_i(\mathbf{e}_i) = 1$ and $E_i(\mathbf{e}_j) = 0$ for $j \neq i$. In other words, for $\mathbf{v} = v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n$,

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

We show that $B^* = \{E_1, \ldots, E_n\}$ is a basis for V^*. To show that B^* is linearly independent, suppose that $c_1E_1 + \cdots + c_nE_n = O$ (an equality of linear transformations). That means that for all $\mathbf{v} \in V$,

$$c_1E_1(\mathbf{v}) + \cdots + c_nE_n(\mathbf{v}) = O(\mathbf{v}) = 0.$$

In particular, for $\mathbf{v} = \mathbf{e}_i$,

$$\begin{aligned}0 &= c_1E_1(\mathbf{e}_i) \quad + \cdots + c_nE_n(\mathbf{e}_i)\\ &= c_iE_i(\mathbf{e}_i)\\ &= c_i\end{aligned}$$

for all $i = 1, \ldots, n$. Hence B^* is a linearly independent set.

To show that B^* spans V^*, choose an arbitrary $T \in V^*$, i.e. $T : V \to \mathbf{R}$ is a linear transformation. We need to find scalars $c_1, \ldots, c_n$ such that $T = c_1E_1 + \cdots + c_nE_n$. Following the idea of the preceding argument for linear independence, define $c_i = T(\mathbf{e}_i)$.

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

$$T(\mathbf{v}) = c_1E_1(\mathbf{v}) + \cdots + c_nE_n(\mathbf{v}).$$

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

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

On the other hand,

$$\begin{aligned} (c_1E_1 + \cdots + c_nE_n)(\mathbf{v}) &= c_1E_1(\mathbf{v}) + \cdots + c_nE_n(\mathbf{v}) \\ &= c_1v_1 + \cdots + c_nv_n. \end{aligned}$$

Hence $T = c_1E_1 + \cdots + c_nE_n$, and B^* spans V^*.

$\square$

Definition 1.8.3. *Let $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for V. The basis $B^* = \{E_1, \ldots, E_n\}$ for V^*, where $E_i : V \to \mathbf{R}$ are linear transformations defined by their action on the basis vectors as*

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

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

Example 1.8.4. *Let B_0 be the standard basis for $\mathbf{R}^n$, i.e.*

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

with 1 in the i^{th} component (see Example 1.4.6). The basis $B_0^* = \{E_1, \ldots, E_n\}$ dual to B_0 is known as the standard basis for $(\mathbf{R}^n)^*$. *Note that if $\mathbf{v} = (v_1, \ldots, v_n)$, then $E_i(\mathbf{v}) = v_i$. In other words, in the language of Example 1.7.5, E_i is the projection onto the i^{th} component.*

We note that Theorem 1.6.1 gives a standard method of writing a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ as an $m \times n$ matrix. Linear one-forms $T \in (\mathbf{R}^n)^*$, $T : \mathbf{R}^n \to \mathbf{R}$ are no exception, when the codomain of scalars is considered as a 1-dimensional vector space. 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:

$$\begin{aligned} [E_1] &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \\ &\vdots \\ [E_n] &= \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}. \end{aligned}$$

We now show how to apply the "dual" construction to linear tranformations between vector spaces V and W. Given a linear transformation $\Psi : V \to W$, we will construct a new linear transformation

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

(Note this construction "reverses the arrow" of the transformation Ψ.)

Take an element of the domain $T \in W^*$, i.e. $T : W \to \mathbf{R}$ is a linear transformation. We wish to assign to T a linear transformation $S = \Psi^*(T) \in V^*$, i.e.

$S : V \to \mathbf{R}$. In other words, given $T \in W^*$, we want to be able to describe $S(\mathbf{v}) = (\Psi^*(T))(\mathbf{v})$ for all $\mathbf{v} \in V$ in such a way that S has the properties of a linear transformation.

To do so, define $\Psi^* : W^* \to V^*$ as

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

$\Psi^*(T)$ is called the *pullback of T by Ψ*.

THEOREM 1.8.5. *Let $\Psi : V \to W$ be a linear transformation and let $\Psi^* : W^* \to V^*$ 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.

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 $T : W \to \mathbf{R}$ is a linear transformation, then $\Psi^*(T) : V \to \mathbf{R}$ 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 for a fixed T and for any vector $\mathbf{v} \in V$ and scalar $s \in \mathbf{R}$, $(\Psi^*(T))(s\mathbf{v}) = s(\Psi^*(T))(\mathbf{v})$ 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 the remaining details, including the verification that for all $T_1, T_2 \in W^*$ that $\Psi^*(T_1 + T_2) = \Psi^*(T_1) + \Psi^*(T_2)$, as an exercise.

□

It is worth mentioning that the definition of the pullback in Theorem 1.8.5 is the sort of "canonical" construction typical of abstract algebra. It can be expressed by the diagram

$$\begin{tikzcd}
V \arrow[r, "\Psi"] \arrow[dr, dashed, "\Psi^* T"'] & W \arrow[d, "T"] \\
& \mathbf{R}
\end{tikzcd}$$

Note also, in other notation, that $\Psi^*(T) = T \circ \Psi$.

EXAMPLE 1.8.6 (The matrix form of a pullback). *Let $\Psi : \mathbf{R}^3 \to \mathbf{R}^2$ be given by $\Psi(x, y, z) = (2x + y - z, x + 3z)$ and $T \in (\mathbf{R}^2)^*$ 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 1.6.1, we have

$$[\Psi] = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 3 \end{bmatrix},$$

$[T] = \begin{bmatrix} 1 & -5 \end{bmatrix}$ *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 1.8.4 and the more general matrix representation from the method following Theorem 1.6.5. Using the standard bases $B_0 = \{E_1, E_2\}$ for $(\mathbf{R}^2)^*$ and $B_0' = \{E_1', E_2', E_3'\}$ for $(\mathbf{R}^3)^*$, where $E_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $E_1' = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $E_2' = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, and $E_3' = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, we note that $\Psi^*(E_1) = 2E_1' + E_2' - E_3'$ and $\Psi^*(E_2) = E_1' + 3E_3'$. Hence

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

In other words, the proper way to see the matrix for a pullback is using the dual basis. Now, when the calculations for the pullback of $T(u, v) = u - 5v$ by Ψ are written using the column vector $[T]_{B_0} = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, *we see that*

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

Note that 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 from Section 0.1 that the *Cartesian product of* $V_1, \dots, V_k$ is the set of ordered n-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 1.8.7. *Let $V_1, \dots, V_k$ and W be vector spaces. A function*

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

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

$$T(\mathbf{x}_1 + \mathbf{y}, \mathbf{x}_2, \dots, \mathbf{x_k}) = T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x_k}) + T(\mathbf{y}, \mathbf{x}_2, \dots, \mathbf{x_k})$$
$$\vdots$$
$$T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}, \mathbf{x_k} + \mathbf{y}) = T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x_k}) + T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}, \mathbf{y})$$

and

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

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

EXAMPLE 1.8.8 (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 1.8.9 (The determinant as an n-form on $\mathbf{R}^n$). *For any $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbf{R}^n$, let A be the matrix whose columns are given by the vectors $\mathbf{a}_i$:*

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n].$$

Define the map $\Omega : \mathbf{R}^n \times \cdots \times \mathbf{R}^n \to \mathbf{R}$ by

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

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 on $\mathbf{R}^n$.

EXAMPLE 1.8.10. *Let $G_0 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ be the function defined by*

$$G_0(\mathbf{x}, \mathbf{y}) = x_1 y_1 + \cdots + x_n y_n.$$

for any $\mathbf{x} = (x_1, \dots, x_n)$ 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 1.8.11. *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 $G_A : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ 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:

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

Likewise,

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

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 1.8.12. *Define* $S : \mathbf{R}^2 \times \mathbf{R}^2 \to \mathbf{R}$ *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 any $c \in \mathbf{R}$, $S(c\mathbf{x}, \mathbf{y}) = cS(\mathbf{x}, \mathbf{y})$. *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.*

Given a vector space V of dimension n and a bilinear form $b : V \times V \to \mathbf{R}$, there is a standard way to represent b by means of an $(n \times n)$ matrix B, assuming a basis is specified.

PROPOSITION 1.8.13. *Let* V *be a vector space with basis* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *and let* $b : V \times V \to \mathbf{R}$ *be a bilinear form. Let* $B = [b_{ij}]$, *where* $b_{ij} = b(\mathbf{e}_i, \mathbf{e}_j)$. *Then, for any* $\mathbf{v}, \mathbf{w} \in V$, *we have*

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

where the product on the right is the matrix product of the transpose of $\mathbf{w}$ *with the matrix product* $B\mathbf{v}$. *Here* $\mathbf{v}$ *and* $\mathbf{w}$ *are written as column vectors relative to the basis.*

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

□

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 1.8.11 is $B = A^T A$.

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 detail in Chapter 3.

DEFINITION 1.8.14. *Suppose $T : V \to W$ is a linear transformation between vector spaces V and W. Let $B : W \times W \to \mathbf{R}$ be a bilinear form on W. Then the* pullback of B by T, *denoted T^*B, is the bilinear form $T^*B : V \times V \to \mathbf{R}$ defined by*

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

PROPOSITION 1.8.15. *Let U, V and W be vector spaces and let $T_1 : U \to V$ and $T_2 : V \to W$ be linear tranformations. Let $B : W \times W \to \mathbf{R}$ be a bilinear form on W. Then*

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

PROOF. For any vectors $\mathbf{u}_1, \mathbf{u}_2 \in U$,

$$\begin{aligned}((T_2 \circ T_1)^* B)(\mathbf{u}_1, \mathbf{u}_2) &= B((T_2 \circ T_1)(\mathbf{u}_1), (T_2 \circ T_1)(\mathbf{u}_2)) \\ &= B(T_2(T_1(\mathbf{u}_1)), T_2(T_1(\mathbf{u}_2))),\end{aligned}$$

and

$$\begin{aligned}(T_1^*(T_2^* B))(\mathbf{u}_1, \mathbf{u}_2) &= (T_2^* B)(T_1(\mathbf{u}_1), T_1(\mathbf{u}_2)) \\ &= B(T_2(T_1(\mathbf{u}_1)), T_2(T_1(\mathbf{u}_2))).\end{aligned}$$

□

1.9. 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: Given two vectors $\mathbf{v}, \mathbf{w}$ in a vector space V, 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* can also be defined, but only in a very crude way. Let W be a subspace of a vector space V. Further, let B' be a basis for W and B be a basis for V such that $B' \subset B$; in this case we can say B *extends* B'. Then we might say that a vector $\mathbf{v} \in V$ is *vertical to W relative to the bases* (B', B) if $\mathrm{Pr}_{B',B}(\mathbf{v}) = \mathbf{0}$, where $\mathrm{Pr}_{B',B}$ is the projection onto W (see Example 1.7.9). Said in another way, $\mathbf{v}$ is vertical to W relative to (B', B) if $\mathbf{v} \in \ker(\mathrm{Pr}_{B',B})$. We could then say that $\mathbf{v}$ is *perpendicular to a vector* $\mathbf{w} \in V$ if $\mathbf{v}$ is vertical to the subspace $W = \mathrm{Span}\{\mathbf{w}\}$, assuming we had made choices of bases (B', B) for W and V.

The reader should note that both these definitions of basic geometric notions are somewhat stilted, since we are defining the terms parallel and perpendicular without reference to the notion of angle. The above definition of verticality is especially awkward, since it is completely dependent on the choice of bases involved. In fact, it should be noted again 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 by means of an additional mathematical structure which generalizes the familiar "dot product" (Example 1.8.10).

DEFINITION 1.9.1. *An* inner product *on a vector space V is a function $G : V \times V \to \mathbf{R}$ 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 1.8.13 must be symmetric and positive definite.

Recall that bilinear means linear in each component. In view of the symmetry condition (I2), in practice this requires checking linearity in just one component. Property (I3), relatively simple to state, has important geometric consequences that we will see shortly.

Example 1.9.2 (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 $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$. We saw in Example 1.8.10 that G_0 is a bilinear form on $\mathbf{R}^n$. The reader may verify property (I2). To see property (I3), note that $G_0(\mathbf{v}, \mathbf{v}) = v_1^2 + \cdots + v_n^2$, a quantity which is always nonnegative and is zero exactly when $v_1 = \cdots = v_n = 0$, i.e. when $\mathbf{v} = \mathbf{0}$.*

Note that Example 1.9.2 can be generalized to any finite-dimensional vector space V. Given a basis $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ 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 1.9.3. *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 (V, G) will be determined not by the existence of an inner product on V, but by the choice of inner product G.

Note that it is easy to construct new inner products from a given inner product structure.

Example 1.9.4. *Let A be any* invertible *$n \times n$ matrix. Define a bilinear form G_A on $\mathbf{R}^n$ as in Example 1.8.11: $G_A(\mathbf{v}, \mathbf{w}) = G_0(A\mathbf{v}, A\mathbf{w})$, where G_0 is the standard inner product from Example 1.9.2. 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 see how this construction looks in 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

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

Given an inner product, it is possible to define geometric notions like length, distance, magnitude and direction.

DEFINITION 1.9.5. *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 1.9.6 (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 1.9.7. *Let* (V, G) *be an inner product space. The* angle $\angle$ *between two nonzero vectors* $\mathbf{v}, \mathbf{w} \in V$ *is defined to be*

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

Note that the definition of angle makes sense as a result of Theorem 1.9.6.

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 $G(\mathbf{v}, \mathbf{w}) = 0$. The notion of orthogonality, in turn, distinguishes "special" bases for V as well as a further method for producing new subspaces of V from a given set of vectors in V.

THEOREM 1.9.8. *Let* (V, G) *be an inner product space with* $\dim(V) = n$*. There exists a basis* $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *satisfying the following two properties:*

(O1) *For* $i = 1, \ldots, n$*,* $G(\mathbf{u}_i, \mathbf{u}_i) = 1$*;*

(O2) *For* $i \neq j$*,* $G(\mathbf{u}_i, \mathbf{u}_j) = 0$*.*

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

The proof relies on an important procedure, similar to the proof of Theorem 1.7.7, known as *Gram-Schmidt orthogonalization*. Beginning with any given basis, the procedure constructs a new basis satisfying (O1) and (O2). We refer the reader to any elementary linear algebra text for the standard proof. In the next section, we will carry out the details in a similar context.

We also state without proof a kind of converse to Theorem 1.9.8. This theorem is actually a restatement of the comment following Example 1.9.2.

THEOREM 1.9.9. *Let* V *be a vector space and let* $B = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *be a basis for* B*. 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 1.6.5. Then G_B *is an inner product.*

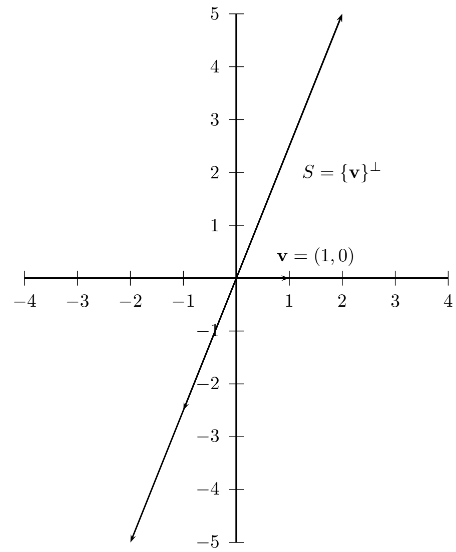

FIGURE 1.3. The orthogonal subspace to $\mathbf{v} = (1, 0)$ in Example 1.9.12.

Given a vector $\mathbf{w}$ in an inner product space (V, G), the set W of all vectors orthogonal to $\mathbf{w}$ can be seen to be a subspace of V. One could appeal directly to Theorem 1.2.2, or one could note that W is the kernel of the linear transformation $i_{\mathbf{w}} : V \to \mathbf{R}$ given by $i_{\mathbf{w}}(\mathbf{v}) = G(\mathbf{w}, \mathbf{v})$.

More generally, we have the following theorem. For any set S, define

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

THEOREM 1.9.10. *Let S be any set of vectors in an inner product space (V, G). The set $S^\perp$ is a subspace of V.*

PROOF. Exercise.

□

$S^\perp$ is called the *orthogonal subspace to S*.

EXAMPLE 1.9.11. *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 R^3 (see Example 1.9.2). Then*

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

See Example 1.2.4.

EXAMPLE 1.9.12. *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 1.9.4. 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 Figure 1.3.

THEOREM 1.9.13. *Let (V, G) be an inner product space. Let*

$$S = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \subset V$$

be a finite subset of V and let $W = \operatorname{Span}(S)$. Then $W^\perp = S^\perp$.

PROOF. Take a vector $\mathbf{v} \in W^{\perp}$, so that $G(\mathbf{w}, \mathbf{v}) = 0$ for all $\mathbf{w} \in W$. In particular, for $\mathbf{w} = \mathbf{w}_i \in S$, $G(\mathbf{w}_i, \mathbf{v}) = 0$. 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 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k$. Relying on the linearity of G in the first component,

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

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

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

□

COROLLARY 1.9.14. *Let B be a basis for a subspace $W \subset V$. Then $W^{\perp} = B^{\perp}$.*

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

THEOREM 1.9.15. *Let W be a subspace of an inner product space (V, G). Then $W \cap W^{\perp} = \{\mathbf{0}\}$.*

The proof is an immediate consequence of the positive definite condition Definition 1.9.1 (I3) of inner products.

THEOREM 1.9.16. *Let W be a subspace of an inner product space (V, G) with $\dim(V) = n$. Then*

$$\dim(W) + \dim(W^{\perp}) = n.$$

PROOF. Let B' be an orthonormal basis for W, extended to an orthonormal basis B for V, which may be chosen according to Theorem 1.9.8. Consider the function $\mathrm{Pr}_{B',B} : V \to V$ given by the projection of a vector $\mathbf{v}$ onto the subspace W relative to the bases (B', B) (see Example 1.7.9). It is an exercise to show that $W^{\perp} = \ker(\mathrm{Pr}_{B',B})$ and that $W = R(\mathrm{Pr}_{B',B})$. The theorem is then a consequence of Theorem 1.7.7.

□

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

DEFINITION 1.9.17. *Let (V, G) be an inner product space. A linear transformation $T : V \to V$ 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 Section 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 following property of linear isometries is easily verified.

PROPOSITION 1.9.18. *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 a linear isometry.*

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

THEOREM 1.9.19. *Let (V, G) be an inner product space, and let $T : V \to V$ 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.*

COROLLARY 1.9.20. *Let $T : V \to V$ be a linear isometry. Then $\det(T) = \pm 1$. In particular, T is invertible.*

In $\mathbf{R}^3$, this implies in geometric language that linear isometries consist of rotations (with determinant 1) composed with reflections (with determinant -1).

PROPOSITION 1.9.21. *Let $T : V \to V$ be a linear isometry. Then its inverse T^{-1} is also a linear isometry.*

PROOF. Assuming T is a linear isometry, apply Proposition 1.8.15 to $G = (\mathrm{Id})^* G = (T \circ T^{-1})^* G$ and use the assumption that $T^* G = G$.

□

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 1.8.2, and so by Theorem 1.7.14, the two vector spaces are isomorphic. A choice of an inner product on V, however, induces a distinguished or canonical isomorphism between them.

THEOREM 1.9.22. *Let G be an inner product on a finite-dimensional vector space V. Then the function*

$$\Phi : V \to V^*$$

defined by $\Phi(\mathbf{v}) = T_{\mathbf{v}}$, where $T_{\mathbf{v}}(\mathbf{w}) = G(\mathbf{v}, \mathbf{w})$, 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 $s \in \mathbf{R}$, $\Phi(s\mathbf{v}) = T_{s\mathbf{v}}$ and so for all $\mathbf{w} \in V$, $T_{s\mathbf{v}}(\mathbf{w}) = G(s\mathbf{v}, \mathbf{w}) = sG(\mathbf{v}, \mathbf{w}) = sT_{\mathbf{v}}(\mathbf{w})$. Hence $T_{s\mathbf{v}} = sT_{\mathbf{v}}$ 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 $\Phi(\mathbf{v}) = O$, 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 1.7.6, 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 1.7.12). However, we will show directly that Φ is onto in order to exhibit the inverse transformation $\Phi^{-1} : V^* \to V$.

Let $T \in V^*$. We need to find $\mathbf{v}_T \in V$ such that $\Phi(\mathbf{v}_T) = T$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for (V, G), as guaranteed by Theorem 1.9.8. 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.

□

The reader should notice the similarity between the construction of Φ^{-1} and the procedure outlined in the proof of Theorem 1.8.2.

The fact that the map Φ in Theorem 1.9.22 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 in the symplectic setting shortly.

1.10. Geometric structures II: Linear symplectic forms

In this section, we outline the essentials of linear symplectic geometry, which will in turn form the basis for one of the main differential geometric structures that we will pursue later in the text. The presentation here will parallel the development of inner product structures in Section 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 Chapter 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 Section 8, we will be more detailed in the presentation and proof of the statements in this section.

DEFINITION 1.10.1. *Let V be a vector space. Let $\omega : V \times V \to \mathbf{R}$ be a function satisfying the following properties:*

(S1) *ω is a bilinear form on V;*

(S2) *ω is skew-symmetric: For all $\mathbf{v}, \mathbf{w} \in V$, $\omega(\mathbf{v}, \mathbf{w}) = -\omega(\mathbf{w}, \mathbf{v})$;*

(S3) *ω is nondegenerate: If $\omega(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in V$, then $\mathbf{v} = \mathbf{0}$.*

Then ω is a linear symplectic form *on V. The pair (V, ω) is called a* symplectic vector space.

Note the main difference between (S1)–(S3) and (I1)–(I3) in Definition 1.9.1 is that a linear symplectic form is skew-symmetric, or anti-commutative, compared 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, for all $\mathbf{v} \in V$, $\omega(\mathbf{v}, \mathbf{v}) = 0$ as a consequence of (S2).

EXAMPLE 1.10.2. *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, $0 = \omega_0(\mathbf{v}, (0, 1)) = v_1$. 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 Figure 1.4.*

EXAMPLE 1.10.3. *Generalizing Example 1.10.2, consider the euclidean vector space $\mathbf{R}^{2n}$. Define the function*

$$\omega_0(\mathbf{v}, \mathbf{w}) = (a_1t_1 - b_1s_1) + \cdots + (a_nt_n - b_ns_n),$$

where $\mathbf{v} = (a_1, b_1, \ldots, a_n, b_n)$ and $\mathbf{w} = (s_1, t_1, \ldots, s_n, t_n)$. The verification that ω_0 is a symplectic form proceeds exactly as in Example 1.10.2; it will be called the

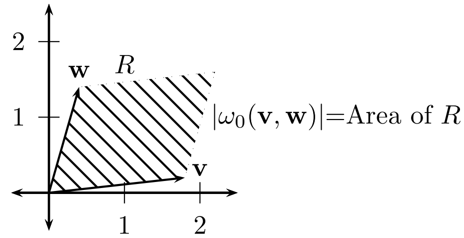

FIGURE 1.4. The standard symplectic form on $\mathbf{R}^2$.

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 must be even. This is the first significant difference between symmetric and skew-symmetric bilinear forms. Theorem 1.10.4 below should be read alongside Theorems 1.9.3 and 1.9.8, which show that every finite-dimensional vector space carries an inner product, to which corresponds distinguished orthonormal bases.

THEOREM 1.10.4. *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 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 orthogonalization process that is at the heart of the proof of Theorem 1.9.8. The reader should also compare the technique here with the technique in the proof of Theorem 1.7.7.

Choose any nonzero $\mathbf{v} \in V$ and define $\mathbf{e}_1 = \mathbf{v}$.

Since $\mathbf{e}_1 \neq \mathbf{0}$, property (S3) 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 1.4.3 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 construction of $\mathbf{f}_1$.

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 $\mathrm{Span}(B_k) = V$, then B_k is the desired basis and we are done.

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

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

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

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

and similarly,

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

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 $\mathrm{Span}(B_k)$ since otherwise the previous calculations would imply that $\omega(\mathbf{e}_{k+1}, \mathbf{w}_{k+1}) = 0$. Define further

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

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

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

Note that the proof relies in an essential way on the nondegeneracy condition (S3) of ω. We will see another proof of this result below in Theorem 1.10.24.

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

$$B_0 = \{\mathbf{e}_1, \mathbf{f}_1, \ldots, \mathbf{e}_n, \mathbf{f}_n\} \tag{1.2}$$

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* $(\mathbf{R}^{2n}, \omega_0)$.

Of course, there is no unique symplectic basis for a given vector space V.

The following is another indication of the importance of the nondegeneracy condition (S3) of Definition 1.10.1. It gives further evidence of why (S3) is the correct symplectic analog to the inner product positive definite condition (I3) in Definition 1.9.1.

THEOREM 1.10.5. *A linear symplectic form ω on a finite-dimensional vector space V induces a linear isomorphism $\Psi : V \to V^*$ defined by $\Psi(\mathbf{v}) = T_{\mathbf{v}} \in V^*$, where $T_{\mathbf{v}}(\mathbf{w}) = \omega(\mathbf{v}, \mathbf{w})$.*

The proof exactly parallels the proof of Theorem 1.9.22.

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 1.10.6. *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, 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 1.9.10.

Following the idea of Section 9, for any set $S \subset (V, \omega)$ we can define the set

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

THEOREM 1.10.7. *Let S be a set of vectors in a symplectic vector space (V, ω). The set S^ω is a subspace of V.*

PROOF. Exercise.

□

S^ω is called the *ω-orthogonal subspace to S*.

EXAMPLE 1.10.8. *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 $\omega_0\big((a,b),(x,y)\big) = ay - bx = 0$. 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 $(x, y) = \big(x, (b/a)x\big) = (x/a)\big(a, b\big)$. In fact, if $W = \mathrm{Span}(S)$, then $W = W^{\omega_0}$. See Figure 1.5.*

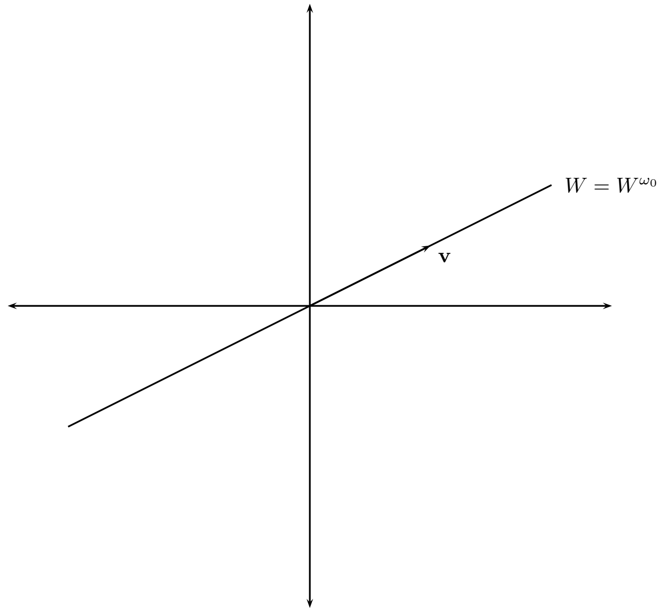

FIGURE 1.5. Symplectic orthogonality in $(\mathbf{R}^2, \omega_0)$, see Example 1.10.8.

EXAMPLE 1.10.9. *Let* $S = \{\mathbf{v}\} \subset (\mathbf{R}^4, \omega_0)$ *where* $\mathbf{v} = (1, 2, 3, 4)$. *Then*

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

We can write a basis for S^{ω_0} *by assigning free variables to* x_1, x_2, y_2 *and writing* $y_1 = 2s + 4t - 3u$, *so that*

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

and so elements of S^{ω_0} *can be written as linear combinations of the linear independent set* $\{(1, 2, 0, 0), (0, 4, 1, 0), (0, -3, 0, 1)\}$. *This shows that* $\dim(S^{\omega_0}) = 3$.

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

THEOREM 1.10.10. *Let* S *be a set of vectors in a symplectic vector space* (V, ω) *and let* $W = \operatorname{Span}(S)$. *Then* $W^\omega = S^\omega$.

PROOF. Let $\mathbf{v} \in W^\omega$. Then for all $\mathbf{w} \in W$, $\omega(\mathbf{w}, \mathbf{v}) = 0$. Since $S \subset W = \operatorname{Span}(S)$, for all $\mathbf{s} \in S$, we have $\mathbf{s} \in W$ and so $\omega(\mathbf{s}, \mathbf{v}) = 0$ and $\mathbf{w} \in S^\omega$. So $W^\omega \subset S^\omega$.

Now suppose $\mathbf{v} \in S^\omega$. Then for all $\mathbf{s} \in S$, $\omega(\mathbf{s}, \mathbf{v}) = 0$. But since $W = \operatorname{Span}(S)$, for all $\mathbf{w} \in W$, there are vectors $\mathbf{s}_1, \ldots, \mathbf{s}_k \in S$ and constants $c_1, \ldots, c_n$ such that

$$\mathbf{w} = c_1\mathbf{s}_1 + \cdots + c_k\mathbf{s}_k.$$

But then by the bilinearity of ω we have

$$\begin{aligned}\omega(\mathbf{w},\mathbf{v}) &= \omega(c_1\mathbf{s}_1+\cdots+c_k\mathbf{s}_k,\mathbf{v})\\ &= c_1\omega(\mathbf{s}_1,\mathbf{v})+\cdots+c_k\omega(\mathbf{s}_k,\mathbf{v})\\ &= 0.\end{aligned}$$

So $\mathbf{v}\in W^\omega$ and $S^\omega\subset W^\omega$.

Hence $S^\omega = W^\omega$.

□

COROLLARY 1.10.11. *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 1.9.16 concerning the dimension of the orthogonal subspace has a direct parallel in the symplectic setting.

THEOREM 1.10.12. *Let W be a subspace of a finite dimensional symplectic vector space (V,ω). Then*

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

PROOF. We rely on the isomorphism $\Psi : V\to V^*$ given by $\Psi(\mathbf{v}) = i(\mathbf{v})\omega$ of Theorem 1.10.5. Consider the map $T : V\to W^*$ given by $T(\mathbf{v}) = \Psi(\mathbf{v})|_W$. On the one hand, T is onto. To see this, let $\mathcal{B} = \{\mathbf{w}_1,\ldots,\mathbf{w}_{2n}\}$ be a basis for V such that $\{\mathbf{w}_1,\ldots,\mathbf{w}_k\}$ is a basis for W. Given $\alpha\in W^*$, define $\tilde\alpha\in V^*$ to be $\tilde\alpha(\mathbf{v}) = c_1\alpha(\mathbf{w_1})+\cdots+c_k\alpha(\mathbf{w}_k)$, where we are writing $\mathbf{v} = c_1\mathbf{w}_1+\cdots+c_{2n}\mathbf{w}_{2n}$ according to the basis $\mathcal{B}$. The reader can check that $\tilde\alpha\in V^*$. Let $\mathbf{v}_\alpha\in V$ be such that $\Psi(\mathbf{v}_\alpha)=\tilde\alpha$. Then $T(\mathbf{v}_\alpha)=\alpha$, and so T is onto.

In addition, practically by definition,

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

We thus rely on Theorem 1.7.7:

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

the last equality due to Theorem 1.8.2.

□

The analog of Theorem 1.9.15 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 1.10.13. *Let W be a subspace of a symplectic vector space (V,ω). 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 1.10.14. *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 1.10.12.

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 1.10.4 to construct a symplectic basis for W which necessarily has an even number of elements, as claimed.

$\square$

EXAMPLE 1.10.15 (Examples of ω-orthogonal subspaces). *The subspace $W =$* Span(S)*, where S is the set in Example 1.10.8, is lagrangian; note that this means it is both isotropic and coisotropic.*

The subspace $W =$ Span(S)*, where S is the set in Example 1.10.9, 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 any subspace $W =$ Span$(\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_k, \mathbf{f}_k\})$ *where $k \leq n$ 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 1.10.16. *Let (V, ω) be a symplectic vector space. A linear transformation $T : V \to V$ 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 Section 8, T is a linear symplectomorphism if $T^*\omega = \omega$.*

We list here some basic properties of linear symplectomorphisms.

PROPOSITION 1.10.17. *Let (V, ω) be a finite-dimensional symplectic vector space. Then:*

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

PROOF. The reader should refer to Proposition 1.8.15 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 1.7.6, this means that T is one-to-one, and so by Corollary 1.7.12, T is also onto.

To do this, suppose $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{v} \in \ker(T)$. 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,

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

a contradiction. Hence $\ker(T) = \{\mathbf{0}\}$, which implies that T is a linear isomorphism. In particular, T is invertible.

The third follows from the fact that

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

assuming that T is a linear symplectomorphism.

□

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

THEOREM 1.10.18. *Let (V, ω) be a symplectic vector space with dimension $\dim V = 2n$. If $T : V \to V$ is a linear symplectomorphism and*

$$\{\mathbf{e}_1, \mathbf{f}_1, \dots, \mathbf{e}_n, \mathbf{f}_n\}$$

is a symplectic basis for V, 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 and T is the linear isomorphism defined (according to Theorem 1.6.5) by $T(\mathbf{e}_i) = \mathbf{u}_i$ and $T(\mathbf{f}_i) = \mathbf{v}_i$. Then T is a linear symplectormorphism.

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 caclulation holds for any symplectic basis. In particular, for

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

we have

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

□

PROPOSITION 1.10.19. *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.

□

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 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 $(\mathbf{R}^{2n}, \omega_0)$ is given by Equation (1.2) following Theorem 1.10.4. Writing vectors $\mathbf{v}, \mathbf{w}$ in standard components as column vectors, the reader can verify using Proposition 1.8.13 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 1.10.20. *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,

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

and

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

Hence $T^*\omega_0 = \omega_0$ is equivalent to the matrix equation $A^T J A = J$.

□

A $(2n) \times (2n)$ matrix satisfying the condition that $A^T J A = J$ will be called a *symplectic matrix*. We write $\mathrm{Sp}(2n)$ 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 1.10.21. *Let $A \in \mathrm{Sp}(2n)$. Then:*

(a) *A is invertible;*
(b) *$A^T \in \mathrm{Sp}(2n)$, and*
(c) *$A^{-1} \in \mathrm{Sp}(2n)$.*

PROOF. Suppose $A \in \mathrm{Sp}(2n)$, 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

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

which shows that

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

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

We leave the proof of (c) as an exercise.

□

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 1.10.22. *If $A \in \mathrm{Sp}(2n)$, then*

$$\det(A) = 1.$$

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

The following statement concerns the eigenvalues of a symplectic matrix.

THEOREM 1.10.23. *Suppose λ is an eigenvalue of the symplectic matrix $A \in \mathrm{Sp}(2n)$ with multiplicity k. Then $1/\lambda$, $\overline{\lambda}$, and $1/\overline{\lambda}$ are also eigenvalues of multiplicity k, where $\overline{\lambda}$ is the complex conjugate of λ.*

PROOF. Consider the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$; note that 0 cannot be a root since then A could not be invertible. It is always the case that $\overline{\lambda}$ is a root of p if λ is, since p is a real polynomial, and that the multiplicities of λ and $\overline{\lambda}$ are the same. We have

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I) \\ &= \det(J(A - \lambda I)J^{-1}) \\ &= \det(JAJ^{-1} - \lambda I) \\ &= \det((A^{-1})^T - \lambda I) \quad \text{since } A^T J A = J \\ &= \det((A^{-1} - \lambda I)^T) \\ &= \det(A^{-1} - \lambda I) \\ &= \det(A^{-1}(I - \lambda A)) \\ &= \det(A^{-1}) \det(I - \lambda A) \\ &= \det(I - \lambda A) \quad \text{by Theorem 1.10.22} \\ &= \lambda^{2n} \det\left(\frac{1}{\lambda} I - A\right) \\ &= \lambda^{2n} p\left(\frac{1}{\lambda}\right). \end{aligned}$$

This shows that if λ is a root of p, then so is $1/\lambda$ (and hence $1/\overline{\lambda}$ 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 $1/\lambda$ is a root of p with multiplicity k.

□

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 1.9.8 and 1.10.4 are not compatible, in the sense that it cannot produce a basis that is simultaneously symplectic and orthogonal. Nevertheless, it is possible to construct such a basis by resorting to techniques of 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 [**3**]. 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 complex conjugation. The most important property of hermitian matrices for our purposes is that they have n linearly independent orthonormal eigenvectors (with respect to the standard hermitian product $\langle \mathbf{x}, \mathbf{y}\rangle = \overline{\mathbf{x}}^T\mathbf{y}$) whose corresponding eigenvalues are real and nonzero.

Theorem 1.10.24. *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 basis*

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

which is symplectic and G-orthogonal, i.e.

$$G(\widetilde{\mathbf{u}}_i, \widetilde{\mathbf{v}}_j) = 0 \quad \text{for all } i, j; \quad G(\widetilde{\mathbf{u}}_i, \widetilde{\mathbf{u}}_j) = G(\widetilde{\mathbf{v}}_i, \widetilde{\mathbf{v}}_j) = 0 \quad \text{for } i \neq j.$$

Moreover, the basis can be chosen so that $G(\widetilde{\mathbf{u}}_i, \widetilde{\mathbf{u}}_i) = G(\widetilde{\mathbf{v}}_i, \widetilde{\mathbf{v}}_i)$ for all i.

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 1.9.8. 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 $A = [a_{ij}]$, where $a_{ij} = \omega(\mathbf{e}_i, \mathbf{e}_j)$. 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,

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

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_{ij} is hermitian:

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

Now by the fundamental property of hermitian matrices mentioned above, there are $2n$ linearly independent eigenvectors of iA which 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

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

(being vectors with complex components, the bar representing complex conjugation as always). The corresponding eigenvalues will be denoted

$$\pm\mu_1, \ldots, \pm\mu_n.$$

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 $j = 1, \ldots, n$, we choose pairs λ_j and $\mathbf{x}_j$ as follows: From each pair of eigenvectors $\mathbf{y}_j, \overline{\mathbf{y}}_j$ 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 $\mathbf{x}_j = \overline{\mathbf{y}}_j$, while if $\lambda_j = -\mu_j$, choose $\mathbf{x}_j = \mathbf{y}_j$. In this way, we have $A\mathbf{x}_j = i\lambda_j \mathbf{x}_j$.

Write

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

with vectors $\mathbf{u}_j$ and $\mathbf{v}_j$ having have real components. We claim that the set $\mathcal{B}'' = \{\mathbf{u}_1, \mathbf{v}_1, \ldots, \mathbf{u}_n, \mathbf{v}_n\}$ is a G-orthogonal basis for V. The fact that $\mathcal{B}''$ is a basis for V is a result of the fundamental property of the eigenvectors of a hermitian matrix. To show that $\mathcal{B}''$ is G-orthogonal, we note that the condition $\overline{\mathbf{x}}_j^T \mathbf{x}_k = \delta_k^j$ 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

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

which implies 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 that, in a similar way, 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 to the reader to find scalars c_j and d_j such that for $\widetilde{\mathbf{u}}_j = c_j\mathbf{u}_j$ and $\widetilde{\mathbf{v}}_j = d_j\mathbf{v}_j$,

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

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

$$\mathcal{B} = \{\widetilde{\mathbf{u}}_1,\widetilde{\mathbf{v}}_1,\ldots,\widetilde{\mathbf{u}}_n,\widetilde{\mathbf{v}}_n\}$$

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:

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

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

THEOREM 1.10.25. *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 $\Phi : (\mathbf{R}^{2n}, \omega_0) \to (\mathbf{R}^{2n}, \omega_0)$ such that $\Phi(E(r_1, \dots, r_n)) = E$, where*

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

The constants are uniquely determined when ordered $r_1 \leq \cdots \leq r_n$.

PROOF. Let G be the inner product defined by the matrix A, i.e. $G(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A \mathbf{y}$. The ellipsoid E is then characterized as

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

According to Theorem 1.10.24, there is a basis

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

which 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 $\Phi : \mathbf{R}^{2n} \to \mathbf{R}^{2n}$ 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

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

we have

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

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

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

we then have

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

and so

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

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,

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

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 $\Phi_1, \Phi_2 : \mathbf{R}^{2n} \to \mathbf{R}^{2n}$ and n-tuples $(r_1, \dots, r_n)$, $(r_1', \dots, r_n')$ with $r_1 \leq \cdots \leq r_n$ and $r_1' \leq \cdots \leq r_n'$ 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 $\mathbf{x}$ is the column vector representation of $(x_1, y_1, \dots, x_n, y_n)$, D is the diagonal matrix $D = \operatorname{diag}\left[1/(r_1)^2, 1/(r_1)^2, \dots, 1/(r_n)^2, 1/(r_n)^2\right]$ and D' is the diagonal matrix $D' = \operatorname{diag}\left[1/(r_1')^2, 1/(r_1')^2, \dots, 1/(r_n')^2, 1/(r_n')^2\right]$. This implies that

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

The fact that as a symplectic matrix, Φ satisfies $\Phi^T J \Phi = J$, along with the fact that $J^{-1} = -J$, together imply $\Phi^T = -J\Phi^{-1}J$ and so

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

This shows that JD is similar to JD', and so the two matrices have the same eigenvalues. The reader may verify that the eigenvalues of JD are $\pm i r_j$ and those of JD' 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 1.10.25 prompts the following definition.

DEFINITION 1.10.26. *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 $\sigma(E) = (r_1, \dots, r_n)$, $r_1 \leq \cdots \leq r_n$ such that there is a linear symplectomorphism Φ with $\Phi(E(r_1, \dots, r_n)) = E$.*

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

1.11. For further reading

With the exception of Sections 8 and 10, much of the material in this chapter can be found in any textbook on linear algebra. The notation here generally follows that in [**3**].

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 [**5**, Chapter 1] or to [**32**].

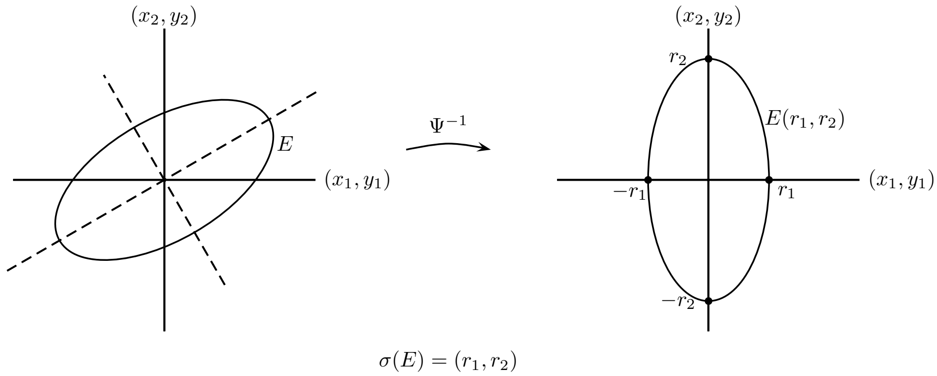

FIGURE 1.6. Linear symplectomorphisms and the symplectic spectrum.

1.12. Exercises

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

(1) Prove Theorem 1.4.3.
(2) Prove Theorem 1.4.10.
(3) Prove Theorem 1.4.13
(4) Let $T : V \to W$ be a linear isomorphism between vector spaces V and W, and let $T^{-1} : W \to V$ be the inverse of T, i.e. $T(T^{-1}(\mathbf{w})) = \mathbf{w}$ and $T^{-1}(T(\mathbf{v})) = \mathbf{v}$. Show that T^{-1} is a *linear* transformation.
(5) Complete the proof of Theorem 1.7.14.
(6) Consider the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ of $\mathbf{R}^3$, where

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

(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 $\mathbf{R}^3$.

(7) For each of the linear transformations Ψ and linear one-forms T below, compute Ψ^*T.
(a) $\Psi : \mathbf{R}^3 \to \mathbf{R}^3$, $\Psi(u, v, w) = (2u, 3u - v - w, u + 2w)$,

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

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

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

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

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

(8) Let $\alpha \in (\mathbf{R}^3)^*$ be given by $\alpha(x, y, z) = 4y + z$.
(a) Describe and find a basis for $\ker \alpha$.

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

(9) Consider the linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ 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 1.8.10.

(b) Compute T^*S, where S is the bilinear form in Example 1.8.12.

(10) For any linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$, show that

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

where A is an $n \times n$ matrix, $A[T]$ is the matrix product of $[T]$ with A, where $[T]$ is written as a matrix relative to the standard basis of $\mathbf{R}^n$, and G_A and $G_{A[T]}$ are defined according to Example 1.8.11.

(11) Prove the following converse to Proposition 1.8.13: Let B be an $n \times n$ matrix and let $\mathcal{B}$ be a basis for the n-dimensional vector space V. Then the function $b : V \times V \to \mathbf{R}$ 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 $\mathcal{B}$, is a bilinear form.

(12) Use Exercise 11 to give five examples of bilinear forms on $\mathbf{R}^3$ and five examples of bilinear forms on $\mathbf{R}^4$.

(13) Let b, B and $\mathcal{B}$ be as given in Exercise 11, and let $T : V \to V$ be a linear transformation. Show that

$$T^*b = \tilde{b},$$

where $\tilde{b}$ is the bilinear form corresponding to the matrix $A^T BA$, where $A = [T]_{\mathcal{B}}$ is the matrix representation of T relative to the basis $\mathcal{B}$.

(14) 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 1.9.4. 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) $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$.

(15) Show that the function $G(\mathbf{v}, \mathbf{w}) = v_1w_1 + 2v_1w_2 + 2v_2w_1 + 5v_2w_2$ 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.

(16) 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 1.9.9). Find $G(\mathbf{v}, \mathbf{w})$ where $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$. Find $\angle((1, 0), (0, 1))$.

(17) Prove Theorem 1.9.10.

(18) For the following subspaces W of $\mathbf{R}^n$, find a basis for $W^\perp$, the orthogonal subspace of W relative to the standard inner product on $\mathbf{R}^n$

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

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

(19) Provide the details for the proof of Theorem 1.9.16

(20) 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 $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of V such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis for W and that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis for $W^{\perp}$. (See Theorem 1.9.16.)
(c) Show that if V is finite-dimensional, then $(W^{\perp})^{\perp} \subset W$, and so by (a) that $W = (W^{\perp})^{\perp}$.

(21) Prove Proposition 1.9.18.

(22) Prove Proposition 1.9.21.

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

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

is also an orthonormal basis for V.

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

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

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

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

(26) 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 1.10.4 to find a symplectic basis $\{\mathbf{e}_1, \mathbf{f}_1, \mathbf{e}_2, \mathbf{f}_2\}$ for $\mathbf{R}^4$ relative to S.

(27) Use the procedure in Theorem 1.10.4 to construct three different symplectic bases for $\mathbf{R}^4$ by making appropriate choices at different stages of the process.

(28) Let $\mathbf{R}^4$ be equipped with the standard linear symplectic form ω_0. Decide if the following subspaces are isotropic, coisotropic, Lagrangian, or symplectic.
(a) $W_1 = \operatorname{Span}\{(1,0,-1,3)\}$;
(b) $W_2 = \operatorname{Span}\{(3,1,0,-1),(2,1,2,1)\}$;
(c) $W_3 = \operatorname{Span}\{(1,0,2,-1),(0,1,1,-1)\}$;
(d) $W_4 = \operatorname{Span}\{(1,1,1,0),(2,-1,0,1),(0,2,0,-1)\}$;

(e) $W_5 = \ker T$, where $T : \mathbf{R}^4 \to \mathbf{R}^2$ is given by

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

(29) Prove Theorem 1.10.7.
(30) Prove Theorem 1.10.12.
(31) Let W_1 and W_2 be subspaces of a symplectic vector space (V, ω). Show that if $W_1 \subset W_2$, then $W_2^\omega \subset W_1^\omega$.
(32) Show that if W is a subspace of a symplectic vector space (V, ω), then $(W^\omega)^\omega = W$.
(33) 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?
(34) Prove Proposition 1.10.19.
(35) Prove Theorem 1.10.5.
(36) For each of the examples in Exercise 24, write the isomorphism Ψ from Theorem 1.10.5 explicitly in terms of the standard bases of $\mathbf{R}^4$ and $(\mathbf{R}^4)^*$.
(37) Let W be a subspace of a finite-dimensional symplectic vector space (V, ω). Let $\Psi : V \to V^*$ be the isomorphism described in Theorem 1.10.5.

(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 $\Psi(W) = (W^\omega)^0$.

(38) Provide the details for the proof of Theorem 1.10.24. In particular:
- Show that the set $\mathcal{B}''$ is a basis for V.
- Show that $\mathbf{u}_j^T \mathbf{v}_k = \mathbf{v}_j^T \mathbf{u}_k = 0$.
- Find scalars c_j and d_j such that for $\tilde{\mathbf{u}}_j = c_j \mathbf{u}_j$ and $\tilde{\mathbf{v}}_j = d_j \mathbf{v}_j$,

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

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

(39) 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 $A^T J A = J$.

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

$$A = \left[\, \mathbf{a}_1 \,\vdots\, \mathbf{b}_1 \,\vdots\, \cdots \,\vdots\, \mathbf{a}_n \,\vdots\, \mathbf{b}_n \,\right]$$

is a symplectic matrix.

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