2

Linear and multilinear algebra

The study of the geometry of Lagrangian mechanics requires that one be familiar with basic concepts in abstract linear and multilinear algebra. The reader is expected to have encountered at least some of these concepts before, so this chapter serves primarily as a refresher. We also use our discussion of linear algebra as a means of introducing the summation convention in a systematic manner. Since this gets used in computations, the reader may wish to take the opportunity to become familiar with it.

We suppose the reader to have had some exposure to basic concepts in linear algebra, and, for this reason, the pace in this chapter will be rather hurried. For the standard material, we shall simply give the definitions, state some results that we will use in the text, and occasionally give a few elementary examples. For readers looking to improve their facility with basic linear algebra, there is an enormous number of good sources. One that is amenable in depth to our needs, as well as being a fine work of exposition, is the text of Halmos [1996]. Linear algebra and its applications are the subject of Strang [1980]. The basics of tensors on vector spaces and on manifolds are well covered by Abraham, Marsden, and Ratiu [1988, Chapters 2 and 5]. A more in-depth analysis of tensors, especially on manifolds and in infinite dimensions, is the account of Nelson [1967]. We remark that in infinite dimensions, even in Banach spaces, there are some technical issues of which one must be careful.

2.1 Basic concepts and notation

Before we begin, since this is our first chapter with technical content, we take the opportunity to give elementary notation. We shall suppose the reader to be familiar with much of this, so our presentation will be quick and informal. Some of this introductory material can be found in [Halmos 1974b]; see also [Abraham, Marsden, and Ratiu 1988].

2.1.1 Sets and set notation

If x is a point in a set S, then we write $x \in S$. When it is syntactically convenient, we may also write $S \ni x$ rather than $x \in S$. For a set S, $\mathbf{2}^S$ denotes the set of subsets of S. The empty set is denoted by $\emptyset$, and we note that $\emptyset \in \mathbf{2}^S$ for any set S. If A is a subset of S, then we write $A \subset S$. We adopt the convention that, if $A \subset S$, then it is possible that $A = S$. If we wish for A to be a proper subset of S, then we write $A \subsetneq S$. For $A \subset S$, $S \setminus A$ is the ***complement*** of A in S, meaning the set of points in S that are not in A. We shall often define a subset of a set by specifying conditions that must be satisfied by elements of the subset. In this case, the notation

$$\{\, x \in S \mid \ \text{condition(s) on } x\}$$

is used.

If S and T are sets, then $S \times T$ denotes the ***Cartesian product*** of the two sets, which consists of ordered pairs (x, y) with $x \in S$ and $y \in T$. If S is a set, then we will sometimes denote the n-fold Cartesian product of S with itself as S^n:

$$S^n = \underbrace{S \times \cdots \times S}_{n\text{-times}}.$$

If A and B are sets, then $A \cup B$ denotes the ***union*** of A and B, i.e., the new set formed by points that lie in either A or B, and $A \cap B$ denotes the ***intersection*** of A and B, i.e., the set of points lying in both A and B. If $S = A \cup B$ and if $A \cap B = \emptyset$, then we say that S is the ***disjoint union*** of A and B. One can also use disjoint union in a different context. Suppose that J is an arbitrary set, and $\{\, S_j \mid \ j \in J\}$ is a collection of sets, indexed by the set J. Thus we have a set assigned to each element of J. The ***disjoint union*** of all these sets is the set

$$\mathop{\overset{\circ}{\bigcup}}_{j \in J} S_j \triangleq \bigcup_{j \in J} (\{j\} \times S_j).$$

The idea is that one combines all the sets S_j, $j \in J$, but one also retains the notion of membership in a particular one of these sets. Put otherwise, a point in the disjoint union specifies two things: (1) the index j indicating which of the sets the point lies in, and (2) an element in S_j.

2.1.2 Number systems and their properties

The set of integers is denoted by $\mathbb{Z}$, and the set of natural numbers, i.e., $\{1, 2, \dots\}$, is denoted by $\mathbb{N}$. The set $\mathbb{N} \cup \{0\}$ will be denoted by $\mathbb{Z}_+$. The rational numbers are denoted by $\mathbb{Q}$, and consist of fractions of integers. The set of real numbers is denoted by $\mathbb{R}$. The strictly positive real numbers are denoted by $\mathbb{R}_+$, and the nonnegative real numbers are denoted by $\bar{\mathbb{R}}_+$. A subset $I \subset \mathbb{R}$ is an ***interval*** if it has one of the following forms:

1. $]-\infty, a[= \{x \in \mathbb{R} \mid x < a\}$;
2. $]-\infty, a] = \{x \in \mathbb{R} \mid x \leq a\}$;
3. $]a, b[= \{x \in \mathbb{R} \mid a < x < b\}$;
4. $]a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$;
5. $]a, +\infty[= \{x \in \mathbb{R} \mid x > a\}$;
6. $[a, b[= \{x \in \mathbb{R} \mid a \leq x < b\}$;
7. $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$;
8. $[a, +\infty[= \{x \in \mathbb{R} \mid x \geq a\}$;
9. $\mathbb{R}$.

The set of intervals is denoted by $\mathscr{I}$, and we shall reserve the letter I for a typical interval.

If $S \subset \mathbb{R}$, then an ***upper bound*** for S is an element $b \in \mathbb{R} \cup \{+\infty\}$ for which $x < b$ for each $x \in S$. Then $\sup S \in \mathbb{R} \cup \{+\infty\}$ denotes the least upper bound for S. In like manner, a ***lower bound*** for S is an element $a \in \{-\infty\} \cup \mathbb{R}$ for which $a < x$ for each $x \in S$, and $\inf S \in \{-\infty\} \cup \mathbb{R}$ denotes the greatest lower bound for S. We call $\sup S$ the ***supremum*** of S and $\inf S$ the ***infimum*** of S. It may be the case that $\sup S = +\infty$ and/or $\inf S = -\infty$, but it is always the case that $\sup S$ and $\inf S$ exist; see [Halmos 1974b]. In cases where it is certain that $\sup S \in S$ (resp. $\inf S \in S$)—for example if S is finite—we may write $\max S$ for $\sup S$ (resp. $\min S$ for $\inf S$).

The set of complex numbers is denoted by $\mathbb{C}$. We abbreviate $\sqrt{-1}$ as i. For $z \in \mathbb{C}$, $\mathrm{Re}(z)$ denotes the real part of z, and $\mathrm{Im}(z)$ denotes the imaginary part of z. It is also useful to split up the complex plane, and we do this as follows:

$$
\begin{aligned}
\mathbb{C}_+ &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0\}, \qquad & \mathbb{C}_- &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) < 0\}, \\
\overline{\mathbb{C}}_+ &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) \geq 0\}, \qquad & \overline{\mathbb{C}}_- &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) \leq 0\}, \\
\mathrm{i}\mathbb{R} &= \{z \in \mathbb{C} \mid \mathrm{Re}(z) = 0\}.
\end{aligned}
$$

2.1.3 Maps

For sets S and T, a ***map*** f from S to T assigns to each element $x \in S$ one and only one element $f(x) \in T$. We write a map as $f \colon S \to T$, calling S the ***domain*** and T the ***codomain*** of f. If we wish to define a map, along with its domain and codomain, all at once, then we will sometimes employ the notation

$$
\begin{aligned}
f \colon S &\to T \\
x &\mapsto \text{what } f \text{ maps } x \text{ to.}
\end{aligned}
$$

Thus S is the domain of f, T is the codomain of f, and the expression on the right on the last line defines the map, whatever that definition might be.

We are careful to distinguish notationally between a map f and its value at a point, denoted by $f(x)$. We define the ***image*** of f by $\mathrm{image}(f) = \{f(x) \mid x \in S\}$. The map $\mathrm{id}_S \colon S \to S$ defined by $\mathrm{id}_S(x) = x$, for all $x \in S$, is the ***identity map***. If $f \colon S \to T$ and $g \colon T \to U$ are maps, then $g \circ f \colon S \to U$ is the ***composition*** of f and g, defined by $g \circ f(x) = g(f(x))$. If $f \colon S \to T$ is a map, and if $A \subset S$, then we denote by $f|A \colon A \to T$ the ***restriction*** of f, which is defined by $(f|A)(x) = f(x)$, $x \in A$.

A map $f\colon S \to T$ is ***injective*** (or is an ***injection***) if the equality $f(x_1) = f(x_2)$ for $x_1, x_2 \in S$ implies that $x_1 = x_2$. A map $f\colon S \to T$ is ***surjective*** (or is a ***surjection***) if, for each $y \in T$, there exists at least one $x \in S$ such that $f(x) = y$. It is common to see an injective map called ***one-to-one*** or ***1–1***, and a surjective map called ***onto***. A map that is both injective and surjective is ***bijective*** (or is a ***bijection***). Sets S and T for which there exists a bijection $f\colon S \to T$ are sometimes said to be in ***1–1 correspondence***. One verifies the following equivalent characterizations of injective, surjective, and bijective maps:

1. a map $f\colon S \to T$ is injective if and only if there exists a map $f_L\colon T \to S$, called a ***left inverse*** of f, such that $f_L \circ f = \mathrm{id}_S$;
2. a map $f\colon S \to T$ is surjective if and only if there exists a map $f_R\colon T \to S$, called a ***right inverse*** of f, such that $f \circ f_R = \mathrm{id}_T$;
3. a map $f\colon S \to T$ is bijective if and only if there exists a unique map $f^{-1}\colon T \to S$, called the ***inverse*** of f, having the property that $f \circ f^{-1} = \mathrm{id}_T$ and $f^{-1} \circ f = \mathrm{id}_S$.

If $A \subset S$, then $i_A\colon A \to S$ denotes the ***inclusion map***, which assigns to a point $x \in A$ the same point, but thought of as being in S. For a map $f\colon S \to T$ and for $B \subset T$, we write

$$f^{-1}(B) = \{x \in S \mid \ f(x) \in B\},$$

and call this the ***preimage*** of B under f. In the event that $B = \{y\}$ is a singleton, we write $f^{-1}(\{y\}) = f^{-1}(y)$. If $T = \mathbb{R}$ and $y \in \mathbb{R}$, then it is common to call $f^{-1}(y)$ a ***level set*** of f. If $S_1, \dots, S_k$ are sets, then, for $i \in \{1, \dots, k\}$, the map $\mathrm{pr}_i\colon S_1 \times \cdots \times S_k \to S_i$ is the ***projection onto the ith factor***, which assigns $x_i \in S_i$ to $(x_1, \dots, x_i, \dots, x_k) \in S_1 \times \cdots \times S_i \times \cdots \times S_k$.

If one has a collection of maps that are related, in some way, then a useful way of displaying these relationships is with a ***commutative diagram***. This is best illustrated with examples; consider the two diagrams

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "h"] \\ C \arrow[r, "i"'] & D \end{tikzcd} \qquad \begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "g"'] & & B \arrow[dl, "h"] \\ & C & \end{tikzcd}$$

The diagram on the left commutes if $h \circ f = i \circ g$, and the diagram on the right commutes if $h \circ f = g$. More generally, and somewhat imprecisely, a diagram ***commutes*** if all possible maps between two sets, obtained by composing maps on the arrows in the diagram, are equal. Note that, in a commutative diagram, unless it is otherwise stated, the maps are not invertible, so one can follow the arrows only in the forward direction.

2.1.4 Relations

If S is a set, then a ***relation*** in S is a subset $R \subset S \times S$, and two points $x_1, x_2 \in S$ are ***R-related*** if $(x_1, x_2) \in R$. A relation R is an ***equivalence relation*** if

1. for each $x \in S$, $(x, x) \in R$ (***reflexivity***),
2. if $(x_1, x_2) \in R$, then $(x_2, x_1) \in R$ (***symmetry***), and
3. if $(x_1, x_2) \in R$ and $(x_2, x_3) \in R$, then $(x_1, x_3) \in R$ (***transitivity***).

For an equivalence relation R, two members $x_1, x_2 \in S$ are ***equivalent*** if $(x_1, x_2) \in R$. We often write $x_1 \sim x_2$ in this case. Indeed, we will on occasion, with a slight abuse of notation, define an equivalence relation in S by indicating when it holds that $x_1 \sim x_2$ for $x_1, x_2 \in S$. For $x_0 \in S$, $[x_0] = \{\, x \in S \mid\ x \sim x_0 \}$ typically denotes the ***equivalence class*** of x_0, i.e., all those points in S equivalent to x. The set of all equivalence classes is written $S/\sim$. Thus $(S/\sim) \subset \mathbf{2}^S$. The map assigning to $x \in S$ its equivalence class $[x] \in S/\sim$ is called the ***canonical projection*** for the equivalence relation.

Let us give an example of an equivalence relation to illustrate the notation and concepts.

Example 2.1. We take $S = \mathbb{R}^2$ and define a relation in S by

$$R = \{\, ((x, y), (x + t, y + t)) \mid\ t \in \mathbb{R} \}\,.$$

Thus $(x_1, y_1) \sim (x_2, y_2)$ if and only if $x_1 - x_2 = y_1 - y_2$. The equivalence class of $(x, y) \in \mathbb{R}^2$ is the line through (x, y) having slope 1:

$$[(x, y)] = \{\, (x + t, y + t) \mid\ t \in \mathbb{R} \}\,.$$

The set of equivalence classes, $\mathbb{R}^2/\sim$, is therefore the set of lines in $\mathbb{R}^2$ with slope 1. To make this more concrete, we note that the equivalence class $[(x, y)]$ is uniquely determined by the point $(x - y, 0) \in [(x, y)]$, i.e., by the point in $[(x, y)]$ on the x-axis. Thus there is a natural bijection between $\mathbb{R}^2/\sim$ and $\mathbb{R}$. •

2.1.5 Sequences and permutations

A set S is ***finite*** if there exist $n \in \mathbb{N}$ and a bijection $f\colon S \to \{1, \dots, n\}$. If S is not finite, it is ***infinite***. A set S is ***countable*** if there exists a bijection $f\colon S \to \mathbb{N}$, and is ***uncountable*** if it is infinite and not countable. A ***sequence*** in a set S is a map $f\colon \mathbb{N} \to S$, and we typically denote a sequence by writing all of its elements as $\{f(n)\}_{n\in\mathbb{N}}$, or more typically as $\{x_n\}_{n\in\mathbb{N}}$, where $x_n = f(n)$. We shall on occasion want collections of sets more general than sequences, and in such cases we may write $\{x_a\}_{a\in A}$, where A is an arbitrary ***index set*** (e.g., see the discussion on disjoint union). If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence, then a ***subsequence*** is a subset $\{x_{n_k}\}_{k\in\mathbb{N}}$ where $\{n_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathbb{N}$ satisfying

$$n_1 < n_2 < \cdots < n_k < n_{k+1} < \cdots .$$

Let us recall some facts about permutations. For $m \in \mathbb{N}$, let S_m be the set of bijections of $\{1, \dots, m\}$ with itself, i.e., the set of permutations of $\{1, \dots, m\}$. We call S_m the ***permutation group*** of order m. If $\sigma \in S_m$, then we represent it as

$$\begin{pmatrix} 1 & 2 & \cdots & m \\ \sigma(1) & \sigma(2) & \cdots & \sigma(m) \end{pmatrix}.$$

Thus, for $j \in \{1, \dots, m\}$, below the jth element in the $2 \times m$ matrix is $\sigma(j)$. A ***transposition*** is a permutation of the form

$$\begin{pmatrix} 1 & \cdots & i & \cdots & j & \cdots & m \\ 1 & \cdots & j & \cdots & i & \cdots & m \end{pmatrix},$$

i.e., a swapping of two elements of $\{1, \dots, m\}$. A permutation $\sigma \in S_m$ is ***even*** if it is the composition of an even number of transpositions, and is ***odd*** if it is the composition of an odd number of transpositions. This definition may be shown to not depend on the choice of transpositions into which a permutation is decomposed. We define $\operatorname{sgn}\colon S_m \to \{-1, 1\}$ by

$$\operatorname{sgn}(\sigma) = \begin{cases} 1, & \sigma \text{ is even}, \\ -1, & \sigma \text{ is odd}. \end{cases}$$

The number $\operatorname{sgn}(\sigma)$ is the ***sign*** of σ.

2.1.6 Zorn's Lemma

In this section we present Zorn's Lemma, which can probably be safely omitted on a first reading. We make only one nontrivial application of Zorn's Lemma, that being in the proof of Lemma 3.94. However, we make many (seemingly) trivial uses of Zorn's Lemma, some of which we point out. We encourage the interested reader to find as many of these occurrences as they can. Some of them are subtle, as may be ascertained from the fact that Zorn's Lemma may be shown to be equivalent to the seemingly obvious Axiom of Choice.[1] There is a fascinating mathematical tale behind these matters, and we refer to [Moore 1982] for an account of this. A proof of Zorn's Lemma may be found in [Halmos 1974b].

To state Zorn's Lemma, we need to state some definitions from set theory. In the following definition, it is convenient to use the notation for a relation that defines what "$x_1 \sim x_2$" means, rather than defining a subset R of $S \times S$.

[1] The Axiom of Choice says that, given any set of mutually exclusive nonempty sets, there exists at least one set that contains exactly one element in common with each of the nonempty sets.

Definition 2.2. Let S be a set.

(i) A ***partial order*** in a set S is a relation $\preceq$ with the properties that
   (a) $x \preceq x$,
   (b) $x \preceq y$ and $y \preceq z$ implies $x \preceq z$, and
   (c) $x \preceq y \preceq x$ implies $x = x$.

   A ***partially ordered set*** is a pair $(S, \preceq)$, where $\preceq$ is a partial order in S.

(ii) A ***total order*** on S is a partial order for which either $x \preceq y$ or $y \preceq x$ for all distinct $x, y \in S$. A set S equipped with a total order is called a ***chain***.

(iii) An ***upper bound*** for a chain S is an element $x \in S$ for which $y \preceq x$ for all $x \in S$. •

With these definitions, we state Zorn's Lemma.

Theorem 2.3 (Zorn's Lemma). *A partially ordered set $(S, \preceq)$ in which every chain has an upper bound contains at least one maximal element.*

2.2 Vector spaces

This section introduces the notions of vector space, linear map, and dual space, with the objective of presenting the following key concepts. First, it is convenient to consider vector spaces in more abstract terms than to simply think of $\mathbb{R}^n$. In other words, a vector space is a set with two operations satisfying certain properties, not a set of n-tuples. Second, vectors, linear maps, and other objects defined on vector spaces can be written in components, once a basis is available, and the summation convention is a convenient procedure to do so. Third, naturally associated to each vector space is the vector space of linear functions on the vector space; this vector space is called the dual space and is analyzed in Section 2.2.5. We refer the reader to [Halmos 1996] and [Strang 1980] for more detailed expositions.

Vector spaces over the field of real numbers arise in many ways in the book. On occasion, we will also find it useful to have on hand some properties of complex vector spaces. We define $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$; thus $\mathbb{F}$ will be used whenever we intend to say that either $\mathbb{R}$ or $\mathbb{C}$ is allowed. If $a \in \mathbb{C}$, then $\bar{a}$ is the complex conjugate, and if $a \in \mathbb{R}$, then $\bar{a} = a$. In like manner, for $a \in \mathbb{F}$, $|a|$ is the absolute value when $\mathbb{F} = \mathbb{R}$, and the complex modulus when $\mathbb{F} = \mathbb{C}$.

2.2.1 Basic definitions and concepts

We begin with a definition.

Definition 2.4 (Vector space). An $\mathbb{F}$-***vector space*** (or simply a ***vector space*** if $\mathbb{F}$ is understood, or if it is immaterial whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) is a set V equipped with two operations: (1) vector addition, denoted by $v_1 + v_2 \in \mathsf{V}$ for $v_1, v_2 \in \mathsf{V}$, and (2) scalar multiplication, denoted by $a\, v \in \mathsf{V}$ for $a \in \mathbb{F}$ and $v \in \mathsf{V}$. Vector addition must satisfy the rules

(i) $v_1 + v_2 = v_2 + v_1$, $v_1, v_2 \in \mathsf{V}$ (***commutativity***),
(ii) $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$, $v_1, v_2, v_3 \in \mathsf{V}$ (***associativity***),
(iii) there exists a unique vector $0 \in \mathsf{V}$ with the property that $v + 0 = v$ for every $v \in \mathsf{V}$ (***zero vector***), and
(iv) for every $v \in \mathsf{V}$, there exists a unique vector $-v \in \mathsf{V}$ such that $v+(-v) = 0$ (***negative vector***),

and scalar multiplication must satisfy the rules

(v) $a_1(a_2\, v) = (a_1 a_2)v$, $a_1, a_2 \in \mathbb{F}$, $v \in \mathsf{V}$ (***associativity***),
(vi) $1\, v = v$, $v \in \mathsf{V}$,
(vii) $a(v_1 + v_2) = a\, v_1 + a\, v_2$, $a \in \mathbb{F}$, $v_1, v_2 \in \mathsf{V}$ (***distributivity***), and
(viii) $(a_1 + a_2)v = a_1\, v + a_2\, v$, $a_1, a_2 \in \mathbb{F}$, $v \in \mathsf{V}$ (***distributivity*** again). •

Example 2.5. The prototypical $\mathbb{F}$-vector space is $\mathbb{F}^n$, $n \in \mathbb{Z}_+$, the collection of ordered n-tuples, $(x^1, \dots, x^n)$, of elements of $\mathbb{F}$. On this space vector, addition and scalar multiplication are defined by

$$(x^1, \dots, x^n) + (y^1, \dots, y^n) = (x^1 + y^1, \dots, x^n + y^n),$$
$$a(x^1, \dots, x^n) = (ax^1, \dots, ax^n).$$

It is an easy exercise to verify that these operations satisfy the conditions of Definition 2.4. We shall normally denote a typical vector in $\mathbb{F}^n$ by a boldface letter, e.g., $\boldsymbol{x} = (x^1, \dots, x^n)$. If $n = 0$, then we adopt the convention that $\mathbb{F}^n$ is the trivial vector space consisting of only the zero vector. •

The set $\mathbb{R}^n$ will be referred to frequently as the ***n-dimensional Euclidean space***. While it has a vector space structure, it also has other structure that will be of interest to us. In particular, it is possible to talk about differentiable functions on Euclidean space, and this makes possible the developments of Chapter 3 that are integral to our approach.

Let us list some concepts that can be immediately defined once one embraces the notion of a vector space.

Definition 2.6. Let V be an $\mathbb{F}$-vector space.

(i) For subsets $S_1, S_2 \subset \mathsf{V}$, the ***sum*** of S_1 and S_2 is the set

$$S_1 + S_2 = \{ v_1 + v_2 \mid \ v_1 \in S_1,\ v_2 \in S_2 \}.$$

(ii) A subset $\mathsf{U} \subset \mathsf{V}$ of a vector space is a ***subspace*** if $u_1 + u_2 \in \mathsf{U}$ for all $u_1, u_2 \in \mathsf{U}$ and if $a\, u \in \mathsf{U}$ for every $a \in \mathbb{F}$ and $u \in \mathsf{U}$.
(iii) If V_1 and V_2 are vector spaces, the ***direct sum*** of V_1 and V_2 is the vector space $\mathsf{V}_1 \oplus \mathsf{V}_2$ whose set is $\mathsf{V}_1 \times \mathsf{V}_2$ (the Cartesian product), and with vector addition defined by $(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2)$ and scalar multiplication defined by $a(v_1, v_2) = (av_1, av_2)$.
(iv) If U_1 and U_2 are subspaces of V, then we shall also write $\mathsf{V} = \mathsf{U}_1 \oplus \mathsf{U}_2$ if $\mathsf{U}_1 \cap \mathsf{U}_2 = \{0\}$ and if every vector $v \in \mathsf{V}$ can be written as $v = u_1 + u_2$

for some $u_1 \in \mathsf{U}_1$ and $u_2 \in \mathsf{U}_2$.[2] It is often convenient to denote a typical vector in a direct sum $\mathsf{V}_1 \oplus \mathsf{V}_2$ by $v_1 \oplus v_2$.

(v) If $\mathsf{V}_1, \dots, \mathsf{V}_k$ are $\mathbb{F}$-vector spaces, then we may write

$$\bigoplus_{j=1}^{k} \mathsf{V}_j \triangleq \mathsf{V}_1 \oplus \cdots \oplus \mathsf{V}_k. \qquad \bullet$$

The next definition collects some language surrounding the notions of linear independence and bases.

Definition 2.7 (Linear independence and bases). Let V be an $\mathbb{F}$-vector space.

(i) A set $S \subset \mathsf{V}$ of vectors is ***linearly independent*** if, for every finite subset $\{v_1, \dots, v_k\} \subset S$, the equality $c^1 v_1 + \cdots + c^k v_k = 0$ implies that $c^1 = \cdots = c^k = 0$.
(ii) A set of vectors $S \subset \mathsf{V}$ ***generates*** a vector space V if every vector $v \in \mathsf{V}$ can be written as $v = c^1 v_1 + \cdots + c^k v_k$ for some choice of constants $c^1, \dots, c^k \in \mathbb{F}$, and for some $v_1, \dots, v_k \in S$. In this case we write $\mathsf{V} = \operatorname{span}_{\mathbb{F}} \{S\}$.
(iii) A ***basis*** for a vector space V is a collection of vectors that is linearly independent and that generates V.
(iv) A vector space is ***finite-dimensional*** if it possesses a basis with a finite number of elements, and the number of basis elements is the ***dimension*** of V, denoted by $\dim(V)$ (one can prove that this is independent of basis).
(v) If U is a subspace of a finite-dimensional vector space V, then the ***codimension*** of U is $\dim(\mathsf{V}) - \dim(\mathsf{U})$. •

The following result indicates why the notion of a basis is so useful.

Proposition 2.8 (Components of a vector). *If $\{e_1, \dots, e_n\}$ is a basis for an $\mathbb{F}$-vector space V, then, for any $v \in \mathsf{V}$, there exist unique constants $v^1, \dots, v^n \in \mathbb{F}$ such that $v = v^1 e_1 + \cdots + v^n e_n$. These constants are called the* ***components*** *of v relative to the basis.*

Here we begin to adopt the convention that components of vectors are indexed with *super*scripts, while lists of vectors are indexed with *sub*scripts. Let us use this chance to introduce the summation convention, first employed by Einstein [1916], that we shall use.

The summation convention. Whenever an expression contains a repeated index, one as a subscript and the other as a superscript, summation is implied over this index. Thus, for example, we have

[2] Some authors use “external direct sum” for the direct sum in (iii) and “internal direct sum” for the direct sum in (iv).

$$v^i e_i = \sum_{i=1}^{n} v^i e_i,$$

since summation over i is implied. •

Example 2.9. The ***standard basis*** for $\mathbb{F}^n$ is given by

$$\boldsymbol{e}_1 = (1, 0, \dots, 0),\ \boldsymbol{e}_2 = (0, 1, \dots, 0),\ \dots,\ \boldsymbol{e}_n = (0, 0, \dots, 1).$$

The choice of a basis $\{e_1, \dots, e_n\}$ for a general n-dimensional vector space V makes V "look like" $\mathbb{F}^n$, in that every vector in V is uniquely represented by its components $(v^1, \dots, v^n) \in \mathbb{F}^n$ relative to this basis. •

We will have occasion to use the notion of a quotient space.

Definition 2.10 (Quotient space). Let $\mathsf{U} \subset \mathsf{V}$ be a subspace and consider the equivalence relation $\sim_{\mathsf{U}}$ in V given by $v_1 \sim_{\mathsf{U}} v_2$ if $v_2 - v_1 \in \mathsf{U}$. Denote by V/U the set of equivalence classes under this equivalence relation, and call V/U the ***quotient space*** of U in V. The equivalence class containing $v \in \mathsf{V}$ will be denoted by $v + \mathsf{U}$. •

The following result gives some useful properties of the quotient space.

Proposition 2.11. *For a finite-dimensional $\mathbb{F}$-vector space V with $\mathsf{U} \subset \mathsf{V}$ a subspace, the quotient space V/U has the following properties:*

(i) the operations

$$\begin{aligned}(v_1 + \mathsf{U}) + (v_2 + \mathsf{U}) &= (v_1 + v_2) + \mathsf{U}, && v_1, v_1 \in \mathsf{V},\\ a(v + \mathsf{U}) &= a\,v + \mathsf{U}, && a \in \mathbb{F},\ v \in \mathsf{V},\end{aligned}$$

make V/U an $\mathbb{F}$-vector space;

(ii) $\dim(\mathsf{V}/\mathsf{U}) = \dim(\mathsf{V}) - \dim(\mathsf{U})$*;*

(iii) if U' is any complement to U in V (i.e., U' has the property that $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}'$), then U' is naturally isomorphic (a notion we define formally below) to V/U.

Proof. This is left to the reader (Exercise E2.1). ■

2.2.2 Linear maps

We once again begin with the definition.

Definition 2.12 (Linear map). A map $A \colon \mathsf{U} \to \mathsf{V}$ between $\mathbb{F}$-vector spaces U and V is ***linear*** if $A(au) = aA(u)$ and if $A(u_1 + u_2) = A(u_1) + A(u_2)$ for each $a \in \mathbb{F}$ and $u, u_1, u_2 \in \mathsf{U}$. If $\mathsf{U} = \mathsf{V}$, then A is sometimes called a ***linear transformation***. •

Remarks 2.13. 1. Sometimes, when there are multiple algebraic structures present in a problem, it is convenient to call a linear map between $\mathbb{F}$-vector spaces $\mathbb{F}$***-linear***. However, we shall only occasionally make use of this terminology.

2. On occasion we shall make use of the notion of an ***affine map*** between vector spaces U and V, by which we mean a map of the form $u \mapsto A(u) + b$, where $A\colon \mathsf{U} \to \mathsf{V}$ is linear and $b \in \mathsf{V}$. •

The set of linear maps from a vector space U to a vector space V is itself a vector space, which we denote by $L(\mathsf{U};\mathsf{V})$. Vector addition in $L(\mathsf{U};\mathsf{V})$ is given by

$$(A + B)(u) = A(u) + B(u),$$

and scalar multiplication is defined by

$$(aA)(u) = a(A(u)).$$

Note that what is being defined in these two equations is $A + B \in L(\mathsf{U};\mathsf{V})$ in the first case, and $aA \in L(\mathsf{U};\mathsf{V})$ in the second case. One verifies that $\dim(L(\mathsf{U};\mathsf{V})) = \dim(\mathsf{U})\dim(\mathsf{V})$, provided that U and V are finite-dimensional.

Example 2.14. Let us consider linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$. We first note that if $A^a_i \in \mathbb{F}$, $a \in \{1, \dots, m\}$, $i \in \{1, \dots, n\}$, then the map $\boldsymbol{A}\colon \mathbb{F}^n \to \mathbb{F}^m$ defined by

$$\boldsymbol{A}(\boldsymbol{x}) = \Big(\sum_{i=1}^{n} A^1_i x^i, \dots, \sum_{i=1}^{n} A^m_i x^i\Big),$$

is readily verified to be linear. What is more, one can also verify (see Exercise E2.2) that *every* linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$ is of this form. •

The image of a linear map is simply its image as defined in Section 2.1. Let us define some useful related concepts.

Definition 2.15. Let U and V be $\mathbb{F}$-vector spaces and let $A \in L(\mathsf{U};\mathsf{V})$.

(i) The ***kernel*** of A, denoted by $\ker(A)$, is the subset of U defined by $\{\, u \in \mathsf{U} \mid\ A(u) = 0\}$.

(ii) The ***rank*** of A is defined to be $\operatorname{rank}(A) = \dim(\operatorname{image}(A))$. (We shall see below that $\operatorname{image}(A)$ is a subspace, so its dimension is well-defined.)

(iii) If U and V are finite-dimensional, then A has ***maximal rank*** if $\operatorname{rank}(A) = \min\{\dim(\mathsf{U}), \dim(\mathsf{V})\}$.

(iv) If A is invertible, it is called an ***isomorphism***, and U and V are said to be ***isomorphic***. •

In some texts on linear algebra, what we call the kernel of a linear map is called the ***null space***, and what we call the image of a linear map is called the ***range***.

Often, the context might imply the existence of a *natural* isomorphism between vector spaces U and V (e.g., the existence of a basis $\{e_1, \dots, e_n\}$ for an $\mathbb{F}$-vector space V implies a natural isomorphism from V to $\mathbb{R}^n$), and in such cases we may write $\mathsf{U} \simeq \mathsf{V}$.

We record some useful properties of the kernel and image of a linear map [Halmos 1996].

Proposition 2.16. *Let* U *and* V *be* $\mathbb{F}$*-vector spaces and let* $A \in L(\mathsf{U};\mathsf{V})$*. The following statements hold:*

(i) $\ker(A)$ *is a subspace of* U*, and* $\operatorname{image}(A)$ *is a subspace of* V*;*
(ii) if U *is finite-dimensional, then* $\dim(\ker(A)) + \operatorname{rank}(A) = \dim(\mathsf{U})$ *(this is the **rank–nullity formula**).*

If $\mathsf{U} = \mathsf{U}_1 \oplus \cdots \oplus \mathsf{U}_k$ and $\mathsf{V} = \mathsf{V}_1 \oplus \cdots \oplus \mathsf{V}_l$, then $A \in L(\mathsf{U};\mathsf{V})$ may be represented in block form by

$$\begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots \\ A_{l1} & \cdots & A_{lk} \end{bmatrix},$$

where $A_{rs} \in L(\mathsf{U}_s;\mathsf{V}_r)$, $r \in \{1,\dots,l\}$, $s \in \{1,\dots,k\}$. At times it will be convenient to simply write a linear map in $L(\mathsf{U};\mathsf{V})$ in this form without comment.

2.2.3 Linear maps and matrices

In Example 2.14 we examined the structure of linear maps between the vector spaces $\mathbb{F}^n$ and $\mathbb{F}^m$. It is natural to think of such linear maps as matrices. Let us formally discuss matrices and some of their properties. An ***m* × *n* matrix with entries in** $\mathbb{F}$ is a map from $\{1,\dots,m\} \times \{1,\dots,n\}$ to $\mathbb{F}$. The image of $(a,i) \in \{1,\dots,m\} \times \{1,\dots,n\}$ is called the ***(a, i)th component*** of the matrix. A typical matrix is denoted by $\boldsymbol{A}$, i.e., we use a bold font to denote a matrix. It is common to represent a matrix as an array of numbers in the form

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}. \tag{2.1}$$

The reader should not be overly concerned with the location of the indices in the preceding equation. There will be times when it will be natural to think of the indices of the components of a matrix as being both up, both down, or one up and one down, depending on what sort of object is being represented by the matrix. This use of matrices is multiple contexts is a potential point of confusion for newcomers, so we recommend paying some attention to understanding this when it comes up. The set of $m \times n$ matrices with entries in $\mathbb{F}$ we denote by $\mathbb{F}^{m\times n}$, and we note that these matrices can be naturally regarded as linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ (cf. Example 2.14). Certain matrices will be of particular interest. An $n \times n$ matrix $\boldsymbol{A}$ is ***diagonal*** if $A_{ij} = 0$ for $i \neq j$. The $n \times n$ identity matrix (i.e., the diagonal matrix with all 1's on the diagonal) is denoted by $\boldsymbol{I}_n$. The $m \times n$ matrix whose entries are all zero is denoted by $\mathbf{0}_{m\times n}$. There will be times when the size of a matrix of zeros will be evident from context, and we may simply write $\mathbf{0}$ in such cases, to avoid clutter.

Remark 2.17 (Elements of $\mathbb{F}^n$ as column vectors). On occasion, it will be convenient to adopt the standard convention of regarding vectors in $\mathbb{F}^n$ as being $n \times 1$ matrices, i.e., as column vectors. Like writing a matrix as (2.1), this is an abuse, although a convenient one. •

For $\boldsymbol{A} \in \mathbb{F}^{n\times n}$, we denote by

$$\operatorname{tr} \boldsymbol{A} = \sum_{j=1}^{n} A_{jj}$$

the ***trace*** of $\boldsymbol{A}$. The ***determinant*** of $\boldsymbol{A} \in \mathbb{F}^{n\times n}$ is defined by

$$\det \boldsymbol{A} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A_{\sigma(1)1} \cdots A_{\sigma(n)n}.$$

It may be shown that $\det \boldsymbol{A} \neq 0$ if and only if the linear transformation of $\mathbb{F}^n$ defined by $\boldsymbol{A}$ is invertible, in which case we say the matrix $\boldsymbol{A}$ is ***invertible***.

Let us now discuss a useful matrix operation.

Definition 2.18 (Transpose and symmetry). If $\boldsymbol{A} \in \mathbb{F}^{m\times n}$, we denote the ***transpose*** of $\boldsymbol{A}$ by $\boldsymbol{A}^T \in \mathbb{F}^{n\times m}$. Thus,

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix} \qquad \Longrightarrow \qquad \boldsymbol{A}^T = \begin{bmatrix} A_{11} & \cdots & A_{m1} \\ \vdots & \ddots & \vdots \\ A_{1n} & \cdots & A_{mn} \end{bmatrix}.$$

If $\mathbb{F} = \mathbb{R}$ and $\boldsymbol{A} = \boldsymbol{A}^T$ (resp. $\boldsymbol{A} = -\boldsymbol{A}^T$), then $\boldsymbol{A}$ is ***symmetric*** (resp. ***skew-symmetric***). •

Next we observe that, by choosing bases, *every* linear map $A \in L(\mathsf{U}; \mathsf{V})$ between finite-dimensional vector spaces can be represented as in Example 2.14. If $\{f_1, \dots, f_n\}$ is a basis for U and $\{e_1, \dots, e_m\}$ is a basis for V, for each $i \in \{1, \dots, n\}$ we may write

$$A(f_i) = A_i^1 e_1 + \cdots + A_i^m e_m,$$

for some unique choice of constants $A_i^1, \dots, A_i^m \in \mathbb{F}$. By letting i run from 1 to n, we thus define nm constants $A_i^a \in \mathbb{F}$, $i \in \{1, \dots, n\}$, $a \in \{1, \dots, m\}$. If we think of $A_i^a \in \mathbb{F}$ as being the (a, i)th component of a matrix, then this defines the ***matrix representation*** or ***matrix representative*** of A relative to the two bases. In identifying the rows and columns of the matrix representative, we note that the superscript is the row index, and the subscript is the column index. We write this matrix as $[A]$, it being understood in any context what bases are being used. If $u \in \mathsf{U}$ is written as $u = u^1 f_1 + \cdots + u^n f_n$, one readily ascertains that

$$A(u) = \sum_{a=1}^{m} \sum_{i=1}^{n} A_i^a u^i e_a.$$

Thus the components of $A(u)$ are written using the summation convention as $A_i^1 u^i, \dots, A_i^m u^i$. We see that the components of $A(u)$ are obtained from those of u by the usual matrix-vector multiplication, using the matrix representation of A, and thinking of the components of u as forming a column vector.

Let us say a few more things about the summation convention we use.

More on the summation convention. 1. In the usual notion of matrix-vector multiplication, the "up" index for A is the row index, and the "down" index is the column index. Note that we can also compactly write

$$\sum_{a=1}^{m} \sum_{i=1}^{n} A_i^a u^i e_a = A_i^a u^i e_a.$$

2. For $i \in \{1, \dots, n\}$, $a \in \{1, \dots, m\}$, let A_i^a and B_a^i be the components of $A \in L(\mathsf{U}; \mathsf{V})$ and $B \in L(\mathsf{V}; \mathsf{U})$, respectively. If A is the inverse of B, then

$$A_i^a B_a^j = \delta_i^j, \quad \text{and} \quad A_i^a B_b^i = \delta_b^a,$$

where δ_i^j, $i, j \in \mathbb{Z}_+$, denotes the ***Kronecker delta***, defined by

$$\delta_i^j = \begin{cases} 1, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

We will at times also find use for the symbol δ_{ij} that has the same meaning as δ_j^i, i.e., it is 1 when i and j are equal, and 0 otherwise. •

Since a matrix representation depends on a choice of basis, it is sometimes interesting to understand how the matrix representation changes when one changes basis. Let us consider the simplest case when $A \in L(\mathsf{V}; \mathsf{V})$, i.e., A is a linear map from a vector space to itself. Let $\mathscr{B} = \{e_1, \dots, e_n\}$ and $\tilde{\mathscr{B}} = \{\tilde{e}_1, \dots, \tilde{e}_n\}$ be bases for V, and let $[A]_{\mathscr{B}}$ and $[A]_{\tilde{\mathscr{B}}}$ be the corresponding matrix representatives for A. The bases $\mathscr{B}$ and $\tilde{\mathscr{B}}$ can be related by

$$e_i = P_i^j \tilde{e}_j, \qquad i \in \{1, \dots, n\},$$

so defining an invertible $n \times n$ matrix $\boldsymbol{P}$ whose (i, j)th component is P_j^i. One then readily checks that the matrix representatives are related by

$$[A]_{\tilde{\mathscr{B}}} = \boldsymbol{P}[A]_{\mathscr{B}} \boldsymbol{P}^{-1}, \tag{2.2}$$

which is called the ***change of basis formula*** for the matrix representative for A. The transformation of $\mathbb{R}^{n \times n}$ given by $\boldsymbol{A} \mapsto \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$ is called a ***similarity transformation***.

Finally, we comment that the notions of trace and determinant for matrices can be extended to linear maps $A \in L(\mathsf{V};\mathsf{V})$. Indeed, one can define the determinant (resp. trace) of A as the determinant (resp. trace) of its matrix representation in a basis. In basic texts on linear algebra [e.g., Halmos 1996] it is shown that these definitions are independent of basis.

2.2.4 Invariant subspaces, eigenvalues, and eigenvectors

We will on occasion encounter linear maps that have an interesting property relative to some subspace.

Definition 2.19 (A-invariant subspace). For V an $\mathbb{F}$-vector space and for $A \in L(\mathsf{V};\mathsf{V})$, a subspace $\mathsf{U} \subset \mathsf{V}$ is ***A-invariant*** if $A(u) \in \mathsf{U}$ for all $u \in \mathsf{U}$. •

The next result follows immediately from the definition of the matrix representation for a linear map.

Proposition 2.20. *Let V be an $\mathbb{F}$-vector space, let $A \in L(\mathsf{V};\mathsf{V})$, and let $\mathsf{U} \subset \mathsf{V}$ be an A-invariant subspace. If $\{e_1, \dots, e_n\}$ is a basis for V with the property that $\{e_1, \dots, e_k\}$ is a basis for U, then*

$$[A] = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0}_{n-k,k} & \boldsymbol{A}_{22} \end{bmatrix},$$

for $\boldsymbol{A}_{11} \in \mathbb{F}^{k\times k}$, $\boldsymbol{A}_{12} \in \mathbb{F}^{k\times(n-k)}$, and $\boldsymbol{A}_{22} \in \mathbb{F}^{(n-k)\times(n-k)}$.

We shall also require the following construction involving the notion of an invariant subspace. For an $\mathbb{F}$-vector space V, an arbitrary subset $\mathscr{L} \subset L(\mathsf{V};\mathsf{V})$, and a subspace $\mathsf{U} \subset \mathsf{V}$, we denote by $\langle \mathscr{L}, \mathsf{U}\rangle$ the smallest subspace of V containing U that is as well an invariant subspace for each of the linear maps from $\mathscr{L}$. One readily verifies that $\langle \mathscr{L}, \mathsf{U}\rangle$ is generated by vectors of the form

$$L_1 \circ \cdots \circ L_k(u), \qquad L_1, \dots, L_k \in \mathscr{L} \cup \{\mathrm{id}_{\mathsf{V}}\},\ u \in \mathsf{U},\ k \in \mathbb{N}.$$

At times it will be useful to extend a $\mathbb{R}$-vector space to a $\mathbb{C}$-vector space.

Definition 2.21 (Complexification). If V is a $\mathbb{R}$-vector space, its ***complexification*** is the $\mathbb{C}$-vector space $\mathsf{V}_{\mathbb{C}}$ defined by $\mathsf{V}_{\mathbb{C}} = \mathsf{V} \times \mathsf{V}$, with vector addition and scalar multiplication defined by

$$\begin{aligned}(u_1, v_1) + (u_2, v_2) &= (u_1 + u_2, v_1 + v_2),\\ (\alpha + \mathrm{i}\beta)(u, v) &= (\alpha u - \beta v, \beta u + \alpha v).\end{aligned}$$ •

One may verify that these operations satisfy the axioms of a $\mathbb{C}$-vector space. For convenience, if V is a $\mathbb{C}$-vector space, then we write $\mathsf{V}_{\mathbb{C}} = \mathsf{V}$. A $\mathbb{R}$-linear map $A \in L(\mathsf{U};\mathsf{V})$ extends to a $\mathbb{C}$-linear map $A_{\mathbb{C}} \in L(\mathsf{U}_{\mathbb{C}};\mathsf{V}_{\mathbb{C}})$ by $A_{\mathbb{C}}(u, v) = (A(u), A(v))$. If U and V are $\mathbb{C}$-vector spaces, for convenience, we take $A_{\mathbb{C}} = A$.

Of special interest are linear maps from an $\mathbb{F}$-vector space V to itself.

Definition 2.22 (Eigenvalues and eigenvectors). Let V be an $\mathbb{F}$-vector space and let $A \in L(\mathsf{V};\mathsf{V})$. An ***eigenvalue*** for A is an element $\lambda \in \mathbb{C}$ with the property that $A_{\mathbb{C}}(v) = \lambda v$ for some nonzero vector $v \in \mathsf{V}_{\mathbb{C}}$, called an ***eigenvector*** for λ. The set of eigenvalues of A is denoted by $\mathrm{spec}(A)$. •

Remark 2.23. We shall sometimes speak of the eigenvalues and eigenvectors of an $n \times n$ matrix. In so doing, we are thinking of such a matrix as being an element of $L(\mathbb{R}^n;\mathbb{R}^n)$. In elementary courses on linear algebra, one talks about eigenvalues and eigenvectors *only* for matrices. It is our desire to talk about these concepts for general vector spaces that necessitates the idea of complexification. •

In the subsequent discussion, we fix a linear map $A \in L(\mathsf{V};\mathsf{V})$, and assume that $\dim(\mathsf{V}) < +\infty$. The eigenvalues of A are the roots of the ***characteristic polynomial*** $P_A(\lambda) \triangleq \det(\lambda \mathrm{id}_{\mathsf{V}_{\mathbb{C}}} - A_{\mathbb{C}})$, which is a monic polynomial[3] having degree equal to the dimension of V. If $\det(\lambda \mathrm{id}_{\mathsf{V}_{\mathbb{C}}} - A_{\mathbb{C}}) = (\lambda - \lambda_0)^k P(\lambda)$ for a polynomial $P(\lambda)$ having the property that $P(\lambda_0) \neq 0$, then the eigenvalue λ_0 has ***algebraic multiplicity*** k. The eigenvectors for an eigenvalue λ_0 are nonzero vectors in the subspace $\mathsf{U}_{\lambda_0} = \ker(\lambda_0 \mathrm{id}_{\mathsf{V}_{\mathbb{C}}} - A_{\mathbb{C}})$. The ***geometric multiplicity*** of an eigenvalue λ_0 is $\dim(\mathsf{U}_{\lambda_0})$. We let $m_a(\lambda_0)$ denote the algebraic multiplicity and $m_g(\lambda_0)$ denote the geometric multiplicity of λ_0. It is always the case that $m_a(\lambda_0) \geq m_g(\lambda_0)$, and both equality and strict inequality can occur. A useful result concerning a linear map and its characteristic polynomial is the following.

Theorem 2.24 (Cayley–Hamilton Theorem). *Let* V *be a finite-dimensional* $\mathbb{F}$*-vector space and let* $A \in L(\mathsf{V};\mathsf{V})$*. If* P_A *is the characteristic polynomial of* A*, then* $P_A(A) = 0$*. That is to say,* A *satisfies its own characteristic polynomial.*

Finally, we comment that there is an important normal form associated with the eigenvalues and eigenvectors of a linear map. This normal form is called the ***Jordan normal form*** and we shall on occasion refer to it in exercises. Readers with a good course in linear algebra behind them will have encountered this. Others may refer to [Halmos 1996] or [Horn and Johnson 1990].

2.2.5 Dual spaces

The notion of a dual space to a vector space V is extremely important in mechanics. It can also be a potential point of confusion, since there is often a reflex to identify the dual of a vector space with the vector space itself. It will be important for us to understand the distinction between a vector space and its dual. Indeed, the reader should be forewarned that certain physical concepts (e.g., velocity) are naturally regarded as living in a certain vector

[3] A polynomial is ***monic*** when the coefficient of the highest-degree term is 1.

space, while others (e.g., force) are naturally regarded as living in the dual of that vector space. The easiest thing to do is to accept objects living in duals of vector spaces as "brute facts" arising from the mathematical development we utilize.

Definition 2.25 (Dual space). If V is a finite-dimensional $\mathbb{F}$-vector space, the ***dual space*** to V is the set $\mathsf{V}^* = L(\mathsf{V}; \mathbb{F})$ of linear maps from V to $\mathbb{F}$. •

If $\alpha \in \mathsf{V}^*$, we shall alternately write $\alpha(v)$, $\alpha \cdot v$, or $\langle \alpha; v \rangle$ to denote the image in $\mathbb{F}$ of $v \in \mathsf{V}$ under α. If $S \subset \mathsf{V}$, we denote by $\mathrm{ann}(S)$ the ***annihilator*** of S, which is defined by $\mathrm{ann}(S) = \{\alpha \in \mathsf{V}^* \mid \alpha(v) = 0,\ v \in S\}$. Note that, for any nonempty set S, $\mathrm{ann}(S)$ is a subspace. By symmetry (at least for finite-dimensional vector spaces), for $T \subset \mathsf{V}^*$ we define the ***coannihilator*** of T to be the subspace of V defined by $\mathrm{coann}(T) = \{v \in \mathsf{V} \mid \alpha(v) = 0,\ \alpha \in T\}$.

Note that, since $\dim(\mathbb{F}) = 1$, V^* is a vector space having dimension equal to that of V. We shall typically call elements of V^* ***covectors***, although the term ***one-form*** is also common.

Let us see how to represent elements in V^* using a basis for V. Given a basis $\{e_1, \dots, e_n\}$ for V, we define n elements of V^*, denoted by $e^1, \dots, e^n$, by $e^i(e_j) = \delta^i_j$, $i, j \in \{1, \dots, n\}$.

The following result is useful, albeit simple.

Proposition 2.26 (Dual basis). *If $\{e_1, \dots, e_n\}$ is a basis for V, then $\{e^1, \dots, e^n\}$ is a basis for V^*, called the **dual basis**.*

Proof. First let us show that the dual vectors $\{e^1, \dots, e^n\}$ are linearly independent. Let $c_1, \dots, c_n \in \mathbb{F}$ have the property that $c_i e^i = 0$. For each $j \in \{1, \dots, n\}$, we must therefore have $c_i e^i(e_j) = c_i \delta^i_j = c_j = 0$. This gives linear independence. Now let us show that each dual vector $\alpha \in \mathsf{V}^*$ can be expressed as a linear combination of $\{e^1, \dots, e^n\}$. For $\alpha \in \mathsf{V}^*$, define $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ by $\alpha_i = \alpha(e_i)$, $i \in \{1, \dots, n\}$. We claim that $\alpha = \alpha_i e^i$. To prove this, it suffices to check that the two covectors α and $\alpha_i e^i$ agree when applied to any of the basis vectors $\{e_1, \dots, e_n\}$. However, this is obvious, since, for $j \in \{1, \dots, n\}$, we have $\alpha(e_j) = \alpha_j$ and $\alpha_i e^i(e_j) = \alpha_i \delta^i_j = \alpha_j$. ■

If $\{e_1, \dots, e_n\}$ is a basis for V with dual basis $\{e^1, \dots, e^n\}$, then we may write $\alpha \in \mathsf{V}^*$ as $\alpha = \alpha_i e^i$ for some uniquely determined $\alpha_1, \dots, \alpha_n \in \mathbb{F}$, called the ***components*** of α in the given basis. If $v \in \mathsf{V}$ is expressed as $v = v^i e_i$, then we have

$$\alpha(v) = \alpha_i e^i(v^j e_j) = \alpha_i v^j e^i(e_j) = \alpha_i v^j \delta^i_j = \alpha_i v^i.$$

Note that this makes the operation of feeding a vector to a covector look very much like the "dot product," but it is in the best interests of the reader to refrain from thinking this way. One cannot take the dot product of objects in different spaces, and this is the case with $\alpha(v)$, since $\alpha \in \mathsf{V}^*$ and $v \in \mathsf{V}$. The proper generalization of the dot product, called an inner product, is given in Section 2.3.

Remark 2.27 (Elements of $(\mathbb{F}^n)^*$ as row vectors). Let us follow up on Remark 2.17. Elements of $(\mathbb{F}^n)^*$ can be thought of as being $1 \times n$ matrices, i.e., as row vectors. Thus, given $\boldsymbol{\alpha} \in (\mathbb{F}^n)^*$, we can write $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}$. Moreover, given $\boldsymbol{v} = (v^1, \dots, v^n) \in \mathbb{F}^n$, and thinking of this vector as a column vector, we have

$$\boldsymbol{\alpha}(\boldsymbol{v}) = \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix} \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}.$$

This representation serves two purposes. First, it makes the notion of a dual space concrete in the case of $\mathbb{F}^n$. Second, it emphasizes the fact that $(\mathbb{F}^n)^*$ is a different object than $\mathbb{F}^n$. The latter is naturally thought of as being the collection of column vectors, while the former is naturally thought of as being the collection of row vectors. •

Given a linear map between finite-dimensional vector spaces, there is a naturally induced map between the dual spaces.

Definition 2.28 (Dual of a linear map). If $A\colon \mathsf{U} \to \mathsf{V}$ is a linear map between finite-dimensional $\mathbb{F}$-vector spaces U and V, then the linear map $A^*\colon \mathsf{V}^* \to \mathsf{U}^*$ defined by

$$\langle A^*(\alpha); u\rangle = \langle \alpha; A(u)\rangle, \qquad \alpha \in V^*,\ u \in \mathsf{U},$$

is the ***dual*** of A. •

Of course, one should verify that this definition makes sense (i.e., is independent of u) and that A^* is linear. This, however, is easily done. The following result records the matrix representative of the dual of a linear map.

Proposition 2.29 (Matrix representation of the dual). *Let U and V be finite-dimensional $\mathbb{F}$-vector spaces, let $A \in L(\mathsf{U};\mathsf{V})$, and let $\{f_1, \dots, f_n\}$ and $\{e_1, \dots, e_m\}$ be bases for U and V, respectively, with $\{f^1, \dots, f^n\}$ and $\{e^1, \dots, e^m\}$ the dual bases. Then the matrix representation for A^* is given by $[A^*] = [A]^T$. In particular, if we write $\alpha = \alpha_a e^a$, we have $A^*(\alpha) = A^a_i \alpha_a f^i$.*

Remark 2.30. In assigning the rows and columns of the matrix representative of A^*, the subscripts are row indices and the superscripts are column indices. This is the opposite of the convention used for the matrix representative of A. This reflects the general fact that, when dealing with duals of vector spaces, the location of indices is swapped from when one is dealing with the vector spaces themselves. •

This result leads to the following extension of our summation convention.

More on the summation convention. As previously observed, when we write a collection of elements of a vector space, we use subscripts to enumerate them, e.g., $v_1, \dots, v_k$. For collections of elements of the dual space, we use *super*scripts to enumerate them, e.g., $\alpha^1, \dots, \alpha^k$. By contrast, we saw that the components of a vector with respect to a basis are written with indices as superscripts. In keeping with these conventions, the components of a covector with respect to a basis for the dual space are written with indices as subscripts. •

2.3 Inner products and bilinear maps

Readers may have encountered the notion of an inner product, perhaps as the dot product in $\mathbb{R}^n$, or perhaps in the context of Hilbert spaces arising in Fourier analysis. The inner product will arise in its most important role for us as a model for the inertial properties of a mechanical system. We shall also require inner products and norms in certain technical developments, so we present here the basic underlying ideas. We also present a generalization of an inner product that will arise in numerous instances: a symmetric bilinear map. The discussion of symmetric bilinear maps provides a nice segue into the discussion of tensors in Section 2.4.

One key idea arising from this section is that inner products, orthogonality, and bilinear maps on vector spaces are defined in general terms, and should not only be understood as the dot product between elements of $\mathbb{R}^n$. A second important point is the basic distinction between linear maps between vector spaces and bilinear maps on a vector space. Despite their similar appearances (their representations in components are similar), these are different objects. A third key concept in this section is that any nondegenerate bilinear map induces two associated linear maps that naturally transform vectors into covectors, and vice versa. These maps can be understood easily when written in components.

2.3.1 Inner products and norms

The notion of an inner product is fundamental to the approach we take in this book. While we shall see that an inner product naturally arises for us in the differential geometric setting, in this section we consider the linear situation.

Definition 2.31 (Inner product). An ***inner product*** on an $\mathbb{F}$-vector space V assigns to each pair of vectors $v_1, v_2 \in \mathsf{V}$ a number $\langle\langle v_1, v_2\rangle\rangle \in \mathbb{F}$, and this assignment satisfies

(i) $\langle\langle v_1, v_2\rangle\rangle = \overline{\langle\langle v_2, v_1\rangle\rangle}$, $v_1, v_2 \in \mathsf{V}$ (***symmetry***),
(ii) $\langle\langle c_1 v_1 + c_2 v_2, v_3\rangle\rangle = c_1 \langle\langle v_1, v_3\rangle\rangle + c_2 \langle\langle v_2, v_3\rangle\rangle$, $v_1, v_2, v_3 \in \mathsf{V}$, $c_1, c_2 \in \mathbb{F}$ (***bilinearity***), and
(iii) $\langle\langle v, v\rangle\rangle \geq 0$, $v \in \mathsf{V}$, and $\langle\langle v, v\rangle\rangle = 0$ if and only if $v = 0$ (***positive-definiteness***). •

A structure more general than an inner product is that of a norm.

Definition 2.32 (Norm). A ***norm*** on a vector space V assigns to each vector $v \in \mathsf{V}$ a number $\|v\| \in \mathbb{R}$, and this assignment satisfies

(i) $\|v\| \geq 0$, $v \in \mathsf{V}$, and $\|v\| = 0$ if and only if $v = 0$ (***positivity***),
(ii) $\|\lambda v\| = |\lambda| \, \|v\|$, $\lambda \in \mathbb{F}$, $v \in \mathsf{V}$ (***homogeneity***), and
(iii) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$, $v_1, v_2 \in \mathsf{V}$ (***triangle inequality***). •

An ***inner product space*** is a pair $(\mathsf{V}, \langle\!\langle \cdot, \cdot \rangle\!\rangle)$, where $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ is an inner product on V. A ***normed vector space*** is a pair $(\mathsf{V}, \|\cdot\|)$, where $\|\cdot\|$ is a norm on V. An inner product space is an example of a normed vector space with the norm given by $\|v\| = \sqrt{\langle\!\langle v, v \rangle\!\rangle}$. To show that this proposed norm satisfies the triangle inequality is not perfectly trivial, but requires the Cauchy–Schwarz Inequality (given as Theorem 2.36 below).

Examples 2.33. 1. Let $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be the standard basis for $\mathbb{F}^n$. The ***standard inner product*** on $\mathbb{F}^n$ is defined by

$$\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle_{\mathbb{F}^n} = \sum_{i=1}^{n} x^i \bar{y}^i,$$

where $\boldsymbol{x} = (x^1, \dots, y^n)$ and $\boldsymbol{y} = (y^1, \dots, y^n)$. Note that this inner product is defined uniquely (by linearity) by the property that $\langle\!\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle\!\rangle = \delta_{ij}$. The corresponding norm on $\mathbb{F}^n$ is called the ***standard norm***, and is denoted by $\|\cdot\|_{\mathbb{F}^n}$.

2. Let us consider a generalization of the standard inner product on $\mathbb{R}^n$. Let $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ be a symmetric matrix. One may verify (see Proposition 2.39 below) that $\boldsymbol{W}$ has real eigenvalues. Let us further suppose that $\boldsymbol{W}$ has *positive* eigenvalues. Then we define an inner product on $\mathbb{R}^n$ by $\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle_{\boldsymbol{W}} = \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{y}$. This is indeed an inner product. What is more, *every* inner product on $\mathbb{R}^n$ has this form for some suitable $\boldsymbol{W}$, as the reader may verify in Exercise E2.15. The norm associated with the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\boldsymbol{W}}$ is denoted by $\|\cdot\|_{\boldsymbol{W}}$. •

An important concept associated with an inner product space is that of orthogonality.

Definition 2.34 (Orthogonality). Let V be an $\mathbb{F}$-vector space with inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$.

(i) Two vectors v and u are ***orthogonal*** if $\langle\!\langle v, u \rangle\!\rangle = 0$.
(ii) A collection of nonzero vectors $\{v_1, \dots, v_m\}$ is ***orthogonal*** if $\langle\!\langle v_i, v_j \rangle\!\rangle = 0$ for $i \neq j$. If additionally $\|v_i\| = 1$, $i \in \{1, \dots, m\}$, then the collection of vectors is orthonormal.
(iii) An ***orthogonal basis*** (resp. ***orthonormal basis***) for V is a basis $\{e_1, \dots, e_n\}$ that is orthogonal (resp. orthonormal).
(iv) An ***orthogonal family*** in V is a countable collection $\{e_j\}_{j \in \mathbb{N}}$ of nonzero vectors for which $\langle\!\langle e_i, e_j \rangle\!\rangle = 0$ whenever $i \neq j$. If additionally $\langle\!\langle e_i, e_i \rangle\!\rangle = 1$ for $i \in \mathbb{N}$, then the family is ***orthonormal***.

(v) Given a subset S of V, we define its ***orthogonal complement*** $S^{\perp}$ as the subspace $\{ v \in \mathsf{V} \mid \langle\langle v, w\rangle\rangle = 0 \text{ for all } w \in S\}$. •

The situation in which one has a subspace of an inner product space is particularly interesting.

Proposition 2.35. *Let V be an $\mathbb{F}$-vector space with inner product $\langle\langle \cdot, \cdot\rangle\rangle$, and let $\mathsf{U} \subset \mathsf{V}$ be a subspace with $\mathsf{U}^{\perp}$ its orthogonal complement. Then $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}^{\perp}$.*

Let V, $\langle\langle \cdot, \cdot\rangle\rangle$, and U be as in the proposition. Then, for any $v \in \mathsf{V}$, we can write $v = v_1 \oplus v_2$ with $v_1 \in \mathsf{U}$ and $v_2 \in \mathsf{U}^{\perp}$. The element of $L(\mathsf{V}; \mathsf{V})$ given by $v = v_1 \oplus v_2 \mapsto v_1 \oplus 0$ is the ***orthogonal projection*** onto U.

In some of our analysis it will be useful to have the following result, for whose proof we refer to [Abraham, Marsden, and Ratiu 1988].

Theorem 2.36 (Cauchy–Schwarz Inequality). *In an inner product space $(\mathsf{V}, \langle\langle \cdot, \cdot\rangle\rangle)$ we have $|\langle\langle v_1, v_2\rangle\rangle| \le \|v_1\| \, \|v_2\|$. Furthermore, equality holds if and only if v_1 and v_2 are collinear (i.e., if and only if $\dim(\operatorname{span}_{\mathbb{F}} \{v_1, v_2\}) \le 1$).*

2.3.2 Linear maps on inner product spaces

On an inner product space $(\mathsf{V}, \langle\langle \cdot, \cdot\rangle\rangle)$ one can define special types of linear maps.

Definition 2.37 (Symmetric and skew-symmetric linear maps). Let V be a $\mathbb{R}$-vector space with inner product $\langle\langle \cdot, \cdot\rangle\rangle$, and let $A \in L(\mathsf{V}; \mathsf{V})$.

(i) A is ***symmetric*** if $\langle\langle A(v_1), v_2\rangle\rangle = \langle\langle v_1, A(v_2)\rangle\rangle$ for all $v_1, v_2 \in \mathsf{V}$.

(ii) A is ***skew-symmetric*** if $\langle\langle A(v_1), v_2\rangle\rangle = -\langle\langle v_1, A(v_2)\rangle\rangle$ for all $v_1, v_2 \in \mathsf{V}$. •

Remarks 2.38. 1. Without the presence of an inner product, note that these notions *do not* make sense.

2. The notion of symmetric and skew-symmetric linear maps are generalizations of the notions of symmetric and skew-symmetric matrices. Indeed, if ones makes the identification between elements of $\mathbb{R}^{n \times n}$ and elements of $L(\mathbb{R}^n; \mathbb{R}^n)$, as indicated in Example 2.14, then one can easily show that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric (resp. skew-symmetric) if and only if $\boldsymbol{A}$ is a symmetric (resp. skew-symmetric) linear map on the $\mathbb{R}$-inner product space $(\mathbb{R}^n, \langle\langle \cdot, \cdot\rangle\rangle_{\mathbb{R}^n})$.

3. It is possible to extend the notion of a symmetric linear map to a $\mathbb{C}$-vector space. In this case, a $\mathbb{C}$-linear map $A \in L(\mathsf{V}; \mathsf{V})$ is ***Hermitian*** if $\langle\langle A(v_1), v_2\rangle\rangle = \langle\langle v_1, A(v_2)\rangle\rangle$. We shall not have much occasion to use Hermitian linear maps, although they do come up in the proof of Theorem 6.42. •

The eigenvalues of symmetric linear maps have useful properties, as recorded by the following results.

Proposition 2.39 (Eigenvalues of symmetric linear maps). *Let $(\mathsf{V}, \langle\!\langle\cdot,\cdot\rangle\!\rangle)$ be a $\mathbb{R}$-inner product space and let $A \in L(\mathsf{V};\mathsf{V})$ be symmetric. The following statements hold:*

(i) $\operatorname{spec}(A) \subset \mathbb{R}$*;*
(ii) if $\lambda_1, \lambda_2 \in \operatorname{spec}(A)$ are distinct, and if v_i is an eigenvector for λ_i, $i \in \{1,2\}$, then $\langle\!\langle v_1, v_2\rangle\!\rangle = 0$;
(iii) more generally, if V is finite-dimensional, then there exists an orthonormal basis for V consisting of eigenvectors for A.

If $A \in L(\mathsf{V};\mathsf{V})$ is symmetric, we denote by $\lambda_{\min}(A)$ (resp. $\lambda_{\max}(A)$) the smallest (resp. largest) eigenvalue of A.

2.3.3 Bilinear maps

An inner product is an example of a more general algebraic object that we now introduce. In this section we shall restrict our discussion to $\mathbb{F} = \mathbb{R}$, unless otherwise stated.

Definition 2.40 (Symmetric and skew-symmetric bilinear maps). If V is a $\mathbb{R}$-vector space, a ***bilinear map*** on V is a map $B\colon \mathsf{V}\times\mathsf{V} \to \mathbb{R}$ with the property that

$$\begin{aligned}B(c_1v_1 + c_2v_2, c_3v_3 + c_4v_4)& \\ = c_1c_3B(v_1,v_3) &+ c_2c_3B(v_2,v_3) + c_1c_4B(v_1,v_4) + c_2c_4B(v_2,v_4),\end{aligned}$$

for all $v_1, v_2, v_3, v_4 \in \mathsf{V}$, and $c_1, c_2, c_3, c_4 \in \mathbb{R}$. If

(i) $B(v_1, v_2) = B(v_2, v_1)$ for all $v_1, v_2 \in \mathsf{V}$, then B is ***symmetric***, and if
(ii) $B(v_1, v_2) = -B(v_2, v_1)$ for all $v_1, v_2 \in \mathsf{V}$, then B is ***skew-symmetric***.

•

Remark 2.41. As with inner products, the notion of a symmetric bilinear map makes sense for $\mathbb{C}$-vector spaces. In the case that V is a $\mathbb{C}$-vector space, a map $B\colon \mathsf{V}\times\mathsf{V} \to \mathbb{C}$ is ***Hermitian*** if it has the properties of symmetry and bilinearity from Definition 2.31 for $\mathbb{C}$-inner products. This idea will come up only in the proof of Theorem 6.42. •

Note that a symmetric bilinear map has exactly the properties of an inner product, except that we no longer require positive-definiteness. The set of symmetric bilinear maps on V is denoted by $\Sigma_2(\mathsf{V})$. Note that $\Sigma_2(\mathsf{V})$ is a $\mathbb{R}$-vector space with the operations of vector addition and scalar multiplication given by

$$(B_1 + B_2)(v_1, v_2) = B_1(v_1, v_2) + B_2(v_1, v_2), \quad (aB)(v_1, v_2) = a(B(v_1, v_2)).$$

Just as a basis $\{e_1, \dots, e_n\}$ for V induces a basis for V^*, it also induces a basis for $\Sigma_2(\mathsf{V})$. The following result is straightforward to prove (Exercise E2.9).

Proposition 2.42 (Basis for $\Sigma_2(\mathsf{V})$). *Let* V *be a* $\mathbb{R}$*-vector space with* $\{e_1, \dots, e_n\}$ *a basis for* V*. For* $i, j \in \{1, \dots, n\}$*,* $i \leq j$*, define* $s^{ij} \in \Sigma_2(\mathsf{V})$ *by*

$$s^{ij}(u, v) = \begin{cases} u^i v^j, & i = j, \\ u^i v^j + v^i u^j, & i \neq j. \end{cases}$$

Then $\{ s^{ij} \mid i, j \in \{1, \dots, n\},\ i \leq j \}$ *is a basis for* $\Sigma_2(\mathsf{V})$*.*

Note that, for $B\colon \mathsf{V} \times \mathsf{V} \to \mathbb{R}$ a bilinear map, for $u, v \in \mathsf{V}$, and for $\{e_1, \dots, e_n\}$ a basis for V, we have

$$B(u, v) = B(u^i e_i, v^j e_j) = B(e_i, e_j) u^i v^j.$$

Motivated by this, let us call $B_{ij} = B(e_i, e_j)$, $i, j \in \{1, \dots, n\}$, the ***components*** of B in the given basis. Note that there are n^2 components, and these are naturally arranged in a matrix, which we denote by $[B]$, called the ***matrix representation*** of B, whose (i, j)th component is B_{ij}. Clearly, B is symmetric (resp. skew-symmetric) if and only if $[B]$ is symmetric (resp. skew-symmetric). It is important to distinguish between linear maps on V and bilinear maps on V. Although each type of map is represented in a basis by an $n \times n$ matrix, these matrices represent different objects. That these objects are not the same is most directly seen by the manner in which the matrix components are represented; for linear maps there is an up index and a down index, whereas for bilinear maps, both indices are down.

To write a symmetric bilinear map $B \in \Sigma_2(\mathsf{V})$ in terms of the basis of Proposition 2.42, one writes

$$B = \sum_{1 \leq i \leq j \leq n} B_{ij} s^{ij} \quad \text{or simply} \quad B = B_{ij} s^{ij},\ i \leq j. \tag{2.3}$$

Note the slight abuse in the summation convention here resulting from the way in which we defined our basis for $\Sigma_2(\mathsf{V})$. Let us summarize some features of the summation convention as they relate to symmetric bilinear maps.

More on the summation convention. The indices for the matrix of a symmetric bilinear map are both subscripts. This should help distinguish bilinear maps from linear maps, since in the latter there is one index up and one index down. If B is a bilinear map with components B_{ij}, $i, j \in \{1, \dots, n\}$, and if u and v are vectors with components u^i, v^i, $i \in \{1, \dots, n\}$, then

$$B(u, v) = B_{ij} u^i v^j.$$

Note that, in contrast with (2.3), the summation over i and j is unrestricted in the previous expression. •

Example 2.43. Let $\langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{R}^n}$ be the standard inner product on $\mathbb{R}^n$. Its matrix representation with respect to the standard basis $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is the $n \times n$ identity matrix $\boldsymbol{I}_n$. •

For symmetric bilinear maps, there are concepts of definiteness that will be useful for us. To make these definitions, it is convenient to associate with $B \in \Sigma_2(\mathsf{V})$ the function $Q_B \colon \mathsf{V} \to \mathbb{R}$ defined by $Q_B(v) = B(v, v)$, sometimes called the ***quadratic form*** associated with B. Furthermore, a symmetric bilinear map is uniquely determined by its quadratic form (see Exercise E2.12), so we will interchangeably use the expressions "symmetric bilinear map" and "quadratic form" for the same thing.

Definition 2.44 (Definiteness of symmetric bilinear maps). With the preceding notation, $B \in \Sigma_2(\mathsf{V})$ is

(i) ***positive-semidefinite*** if $\operatorname{image}(Q_B) \subset \bar{\mathbb{R}}_+$, is
(ii) ***positive-definite*** if it is positive-semidefinite and $Q_B^{-1}(0) = \{0\}$, is
(iii) ***negative-semidefinite*** if $-B$ is positive-semidefinite, is
(iv) ***negative-definite*** if $-B$ is positive-definite, is
(v) ***semidefinite*** if it is either positive- or negative-semidefinite, is
(vi) ***definite*** if it is either positive- or negative-definite, and is
(vii) ***indefinite*** if it is neither positive- nor negative-semidefinite. •

Remarks 2.45. 1. Note that if V is the zero vector space, i.e., $\mathsf{V} = \{0\}$, then there is only one element of $\Sigma_2(\mathsf{V})$, and that is the zero map. By definition this is positive-definite. This degenerate case will actually come up in the treatment of controllability using vector-valued quadratic forms in Chapter 8, as well as in our treatment of stabilization in Chapter 10.

2. The above definitions can also be applied to a linear map $A \in L(\mathsf{V}; \mathsf{V})$ that is symmetric with respect to an inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ on V. To do this, one considers on V the symmetric bilinear map given by $B_A(v_1, v_2) = \langle\!\langle A(v_1), v_2 \rangle\!\rangle$, and then Definition 2.44 applies to B_A. In particular, if we take $\mathsf{V} = \mathbb{R}^n$ and $\langle\!\langle \cdot, \cdot \rangle\!\rangle = \langle\!\langle \cdot, \cdot \rangle\!\rangle_{\mathbb{R}^n}$ then we see that it is possible to talk about symmetric matrices having the various definiteness properties ascribed to symmetric bilinear maps. This is somewhat related to Exercise E2.22. •

It is in fact easy in principle to determine whether a given $B \in \Sigma_2(\mathsf{V})$ satisfies any of the above conditions. The following theorem indicates how to do this.

Theorem 2.46 (Normal form for symmetric bilinear map). *If B is a symmetric bilinear map on a finite-dimensional $\mathbb{R}$-vector space V, then there exists a basis $\{e_1, \dots, e_n\}$ for V such that the matrix representative for B in this basis is given by*

$$[B] = \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{I}_l & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}.$$

*The number of nonzero elements on the diagonal of this matrix (i.e., $k + l$) is the **rank** of B, denoted by $\operatorname{rank}(B)$. The number of -1's along the diagonal (i.e., l) is the **index** of B, denoted by $\operatorname{ind}(B)$. The number of $+1$'s along* *the diagonal minus the number of* -1*'s along the diagonal (i.e.,* $k - l$*) is the* ***signature*** *of* B*, denoted by* $\operatorname{sig}(B)$*. The map* B *is* ***nondegenerate*** *if* $\operatorname{rank}(B) = \dim(\mathsf{V})$*.*

This theorem is proved by the Gram–Schmidt Procedure that may be familiar to students having had a good course in linear algebra [Halmos 1996]. In Exercise E2.20 the reader may work out the fairly obvious correspondences between rank, index, and signature, and the various forms of definiteness. In Exercise E2.22 we also provide an alternative method for determining the definiteness of a symmetric bilinear map.

2.3.4 Linear maps associated with bilinear maps

In this section we introduce notation for some operations that are natural in coordinates, but which can provide confusion when thought of abstractly. Part of the difficulty may be that the notation fits together in a slick way, and the slickness may cause confusion.

For $B\colon \mathsf{V} \times \mathsf{V} \to \mathbb{R}$ a bilinear map, we define the ***flat map*** $B^\flat\colon \mathsf{V} \to \mathsf{V}^*$ by asking that, for $v \in \mathsf{V}$, $B^\flat(v)$ satisfy $\langle B^\flat(v); u\rangle = B(u, v)$ for all $u \in \mathsf{V}$. If $\{e_1, \dots, e_n\}$ is a basis for V with B_{ij}, $i, j \in \{1, \dots, n\}$, the components of B in this basis, then the map $B^\flat\colon \mathsf{V} \to \mathsf{V}^*$ is given simply by $B^\flat(v) = B_{ij} v^j e^i$. In other words, relative to the basis $\{e_1, \dots, e_n\}$ for V and the dual basis $\{e^1, \dots, e^n\}$ for V^*, the components of the matrix representative for the linear map $B^\flat\colon \mathsf{V} \to \mathsf{V}^*$ are exactly the components of B.

If $B^\flat$ is invertible, then we denote the inverse, called the ***sharp map***, by $B^\sharp\colon \mathsf{V}^* \to \mathsf{V}$. Of course, given a basis $\{e_1, \dots, e_n\}$ for V inducing the dual basis $\{e^1, \dots, e^n\}$ for V^*, the matrix $[B^\sharp]$ of the linear map $B^\sharp$ is exactly the inverse of the matrix $[B^\flat]$ for $B^\flat$. It is convenient to denote the components of the matrix $[B^\sharp]$ by B^{ij}, $i, j \in \{1, \dots, n\}$, if B_{ij}, $i, j = \{1, \dots, n\}$, are the components for B. Thus $B^{ij} B_{jk} = \delta^i_k$, $i, k \in \{1, \dots, n\}$. In this case we have $B^\sharp(\alpha) = B^{ij} \alpha_j e_i$.

Remarks 2.47. 1. Some authors use a different convention for defining $B^\flat$, using $\langle B^\flat(v); u\rangle = B(v, u)$ for all $u \in \mathsf{V}$ to define $B^\flat(v)$. In the text, this will not come up since we will only use $B^\flat$ in the case that B is symmetric. However, there are instances (e.g., Hamiltonian mechanics) where it will make a difference.

2. In the case that the vector space under consideration is itself a dual space, say $\mathsf{V} = \mathsf{U}^*$, and if $B\colon \mathsf{V} \times \mathsf{V} \to \mathbb{R}$ is a bilinear map, then $B^\flat$ is a linear map between $\mathsf{V} = \mathsf{U}^*$ and $\mathsf{V}^* = \mathsf{U}^{**}$. Provided that U is finite-dimensional, it is easy to show that $\mathsf{U}^{**} \simeq \mathsf{U}$ (see Exercise E2.4). For this reason, we will actually write the map induced by B from U^* to U as $B^\sharp$. This will come up in Chapters 10 and 12. •

More on the summation convention. The maps $B^\flat\colon \mathsf{V} \to \mathsf{V}^*$ and $B^\sharp\colon \mathsf{V}^* \to \mathsf{V}$ have the following devices for easily remembering their form in coordinates. The map $B^\flat$ in words is "B-flat," and in coordinates it takes a vector (with indices up) and returns a covector (with indices down). Conversely, the map $B^\sharp$ in words is "B-sharp," and in coordinates it takes a covector (with indices down) and returns a vector (with indices up). Thus the notation for the two maps is intended to reflect what they do to indices. In order for the expression $B^\flat(v) = B_{ij}v^j e^i$ to represent usual matrix-vector multiplication between the matrix representative of B and the components of v, one should interpret the first index in B_{ij} as being the row index, and the second as being the column index. The same is true when interpreting $B^\sharp(\alpha) = B^{ij}\alpha_j e_i$ as matrix-vector multiplication. •

We conclude this section with some remarks on inner products. Since an inner product is an element of the vector space $\Sigma_2(\mathsf{V})$, it too has components. For this reason, it is often convenient to write an inner product not as $\langle\langle\cdot,\cdot\rangle\rangle$, but with a character; we choose $\mathbb{G}$ to represent a generic inner product. With this notation, we let $\mathbb{G}_{\mathbb{R}^n}$ denote the standard inner product on $\mathbb{R}^n$. We can now talk about the components of an inner product $\mathbb{G}$, and talk about its matrix representation $[\mathbb{G}]$. The discussion of the preceding paragraph can be specialized to the case of an inner product, yielding the maps $\mathbb{G}^\flat \colon \mathsf{V} \to \mathsf{V}^*$ and $\mathbb{G}^\sharp \colon \mathsf{V}^* \to \mathsf{V}$, called the ***associated isomorphisms*** for the inner product $\mathbb{G}$. Note that $\mathbb{G}^\flat$ is indeed an isomorphism by virtue of the positive-definiteness of an inner product. The isomorphism $\mathbb{G}^\flat$ allows us to define an inner product on V^* that we denote by $\mathbb{G}^{-1}$. We define this inner product by

$$\mathbb{G}^{-1}(\alpha^1,\alpha^2) = \mathbb{G}(\mathbb{G}^\sharp(\alpha^1),\mathbb{G}^\sharp(\alpha^2)). \tag{2.4}$$

If $\{e_1,\dots,e_n\}$ is a basis for V, then the components of the inner product $\mathbb{G}^{-1}$ relative to the dual basis $\{e^1,\dots,e^n\}$ are $\mathbb{G}^{ij}$, $i,j \in \{1,\dots,n\}$, i.e., the components of the inverse of the matrix of components of $\mathbb{G}$.

2.4 Tensors

Without actually saying so, this has, thus far, been a chapter on tensors. That is, many of the objects in the preceding sections on linear algebra are part of a more general class of objects called tensors. Now we will talk about tensors in a general way. Readers having the benefit of a good algebra course may have seen tensors presented in a different way than is done here. We do not comment on this except to say that the two approaches are equivalent in finite dimensions, and that Nelson [1967] and Abraham, Marsden, and Ratiu [1988, Section 5.1] have more to say about this. Throughout this section, we restrict consideration to $\mathbb{R}$-vector spaces.

Roughly speaking, tensors are multilinear maps whose domain is multiple copies of a vector space and of its dual space, and whose codomain is $\mathbb{R}$. While the essential idea of multilinearity is straightforward, some careful bookkeeping is required to perform computations on tensors, such as the tensor product, and the operations of pull-back and push-forward. The latter two operations arise when, given a linear map between two vector spaces, one wants to "map" the tensor from one vector space to the other. These operations will be natural and important when dealing with more complex objects in later chapters.

2.4.1 Basic definitions

Let us begin by defining multilinear maps. Let $\mathsf{U}_1, \dots, \mathsf{U}_k$ and V be $\mathbb{R}$-vector spaces. A map $A\colon \mathsf{U}_1 \times \cdots \times \mathsf{U}_k \to \mathsf{V}$ is called ***multilinear*** if it is linear in each of the k arguments separately. Precisely, A is multilinear if, for each $i \in \{1, \dots, k\}$, and for each $u_j \in \mathsf{U}_j$, $j \in \{1, \dots, i-1, i+1, \dots, k\}$, the map

$$\mathsf{U}_i \ni u_i \mapsto A(u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_k) \in \mathsf{V}$$

is linear. We let $L(\mathsf{U}_1, \dots, \mathsf{U}_k; \mathsf{V})$ denote the set of multilinear maps from $\mathsf{U}_1 \times \cdots \times \mathsf{U}_k$ to V. If $\mathsf{U}_1 = \cdots = \mathsf{U}_k = \mathsf{U}$, we abbreviate the set of multilinear maps by $L^k(\mathsf{U}; \mathsf{V})$. As with linear maps, sometimes it is useful to refer to a multilinear map between $\mathbb{R}$-vector spaces as being $\mathbb{R}$-***multilinear***, to distinguish between possible multiple algebraic structures in the same setup.

Definition 2.48 (Tensors on vector spaces). Let V be a $\mathbb{R}$-vector space and let $r, s \in \mathbb{Z}_+$. We define the ***tensors of type (r, s)*** to be the set

$$T^r_s(\mathsf{V}) = L(\underbrace{\mathsf{V}^* \times \cdots \times \mathsf{V}^*}_{r \text{ copies}} \times \underbrace{\mathsf{V} \times \cdots \times \mathsf{V}}_{s \text{ copies}}; \mathbb{R}). \qquad \bullet$$

Therefore, a tensor of type (r, s), or simply an (r, s)-tensor, is a $\mathbb{R}$-valued multilinear map taking as inputs r covectors and s vectors. For example, symmetric bilinear maps, and in particular inner products, are $(0, 2)$-tensors, since they bilinearly map elements in $\mathsf{V} \times \mathsf{V}$ into $\mathbb{R}$. In the case in which $r = s = 0$, we adopt the convention that $T^0_0(\mathsf{V}) = \mathbb{R}$.

A basic operation one can perform with tensors is to take the product of $t_1 \in T^{r_1}_{s_1}(\mathsf{V})$ and $t_2 \in T^{r_2}_{s_2}(\mathsf{V})$ to obtain a tensor $t_1 \otimes t_2 \in T^{r_1+r_2}_{s_1+s_2}(\mathsf{V})$ as follows:

$$\begin{aligned}(t_1 \otimes t_2)(\alpha^1, \dots, \alpha^{r_1}, \beta^1, \dots, \beta^{r_2}, v_1, \dots, v_{s_1}, u_1, \dots, u_{s_2})\\ = t_1(\alpha^1, \dots, \alpha^{r_1}, v_1, \dots, v_{s_1}) t_2(\beta^1, \dots, \beta^{r_2}, u_1, \dots, u_{s_2}).\end{aligned}$$

This is the ***tensor product*** of t_1 and t_2. Note that generally $t_1 \otimes t_2 \neq t_2 \otimes t_1$. That is to say, the tensor product is *not* commutative. The tensor product is, however, associative, meaning that $(t_1 \otimes t_2) \otimes t_3 = t_1 \otimes (t_2 \otimes t_3)$. It is also linear with respect to the operations of vector addition and scalar multiplication we defined below (see (2.5)). Associativity and linearity are often useful in computations involving the tensor product.

What we called in Section 2.3 a symmetric bilinear map on V is nothing more than a $(0, 2)$ tensor B that satisfies $B(v_1, v_2) = B(v_2, v_1)$ for $v_1, v_2 \in$

V. We call such $(0,2)$-tensors ***symmetric***. In like manner, a $(0,2)$-tensor B is ***skew-symmetric*** if $B(v_1, v_2) = -B(v_2, v_1)$ for $v_1, v_2 \in \mathsf{V}$. We denote by $\Sigma_2(\mathsf{V})$ the symmetric $(0,2)$-tensors, just as when we discussed symmetric bilinear maps above. Of course, similar definitions can be made for tensors of type $(2,0)$, and the resulting symmetric $(2,0)$-tensors are denoted by $\Sigma_2(\mathsf{V}^*)$.

These notions of symmetry and skew-symmetry may be carried over to more general tensors, namely, to $(r,0)$-tensors and to $(0,s)$-tensors for any $r, s \in \mathbb{N}$. However, we shall not need this degree of generality.

2.4.2 Representations of tensors in bases

Let us now see how one may represent tensors in bases, just as we have done for linear maps, covectors, and symmetric bilinear maps. First one should note that $T^r_s(\mathsf{V})$ is indeed a $\mathbb{R}$-vector space with vector addition and scalar multiplication defined by

$$\begin{aligned} &(t_1 + t_2)(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s) \\ &\qquad = t_1(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s) + t_2(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s), \\ &(at)(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s) = a(t(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s)). \end{aligned} \tag{2.5}$$

Now let $\{e_1, \dots, e_n\}$ be a basis for V and let $\{e^1, \dots, e^n\}$ be the dual basis for V^*. The reader is invited to prove the following result, which gives an induced basis for $T^r_s(\mathsf{V})$ (Exercise E2.25).

Proposition 2.49 (Basis for $T^r_s(\mathsf{V})$). *If V is a finite-dimensional $\mathbb{R}$-vector space with basis $\{e_1, \dots, e_n\}$ and dual basis $\{e^1, \dots, e^n\}$, then the set*

$$\left\{ e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s} \;\middle|\; i_1, \dots, i_r, j_1, \dots, j_s \in \{1, \dots, n\} \right\}$$

is a basis for $T^r_s(\mathsf{V})$.

The ***components*** of $t \in T^r_s(\mathsf{V})$ relative to $\{e_1, \dots, e_n\}$ are

$$t^{j_1 \cdots j_r}_{i_1 \cdots i_s} = t(e^{j_1}, \dots, e^{j_r}, e_{i_1}, \dots, e_{i_s}), \qquad i_1, \dots, i_s, j_1, \dots, j_r \in \{1, \dots, n\}.$$

Thus an (r,s)-tensor has n^{r+s} components. Indeed, one easily verifies that we in fact have

$$t = t^{j_1 \cdots j_r}_{i_1 \cdots i_s} e_{j_1} \otimes \cdots \otimes e_{j_r} \otimes e^{i_1} \otimes \cdots \otimes e^{i_s}.$$

Note here that the summation convention is in full force, taking the place of $r + s$ summation signs. Also note that we can write

$$t(\alpha^1, \dots, \alpha^r, v_1, \dots, v_s) = t^{j_1 \cdots j_r}_{i_1 \cdots i_s} \alpha^1_{j_1} \cdots \alpha^r_{j_r} v^{i_1}_1 \cdots v^{i_s}_s,$$

where again the summation convention is implied with respect to the indices $j_1, \dots, j_r$ and $i_1, \dots, i_s$.

2.4.3 Behavior of tensors under linear maps

Given two $\mathbb{R}$-vector spaces U and V and a linear map $A \in L(\mathsf{U};\mathsf{V})$, there is an induced map between certain of the tensors on U and V. This will be particularly useful to us in the geometric setting of Chapter 3, but it is convenient to initiate the presentation in the context of linear algebra.

Definition 2.50 (Push-forward and pull-back on vector spaces). Let U and V be finite-dimensional $\mathbb{R}$-vector spaces and let $A \in L(\mathsf{U};\mathsf{V})$.

(i) If $t \in T^r_0(\mathsf{U})$, then the ***push-forward*** of t by A is $A_*t \in T^r_0(\mathsf{V})$, defined by

$$A_*t(\beta^1,\dots,\beta^r) = t(A^*(\beta^1),\dots,A^*(\beta^r)).$$

(ii) If $t \in T^0_s(\mathsf{V})$, then the ***pull-back*** of t by A is $A^*t \in T^0_s(\mathsf{U})$, defined by

$$A^*t(u_1,\dots,u_s) = t(A(u_1),\dots,A(u_s)).$$

(iii) If A is an isomorphism and if $t \in T^r_s(\mathsf{U})$, then the ***push-forward*** of t by A is $A_*t \in T^r_s(\mathsf{V})$, defined by

$$\begin{aligned}A_*t(&\beta^1,\dots,\beta^r,v_1,\dots,v_s)\\ &= t(A^*(\beta^1),\dots,A^*(\beta^r),A^{-1}(v_1),\dots,A^{-1}(v_s)).\end{aligned}$$

(iv) If A is an isomorphism and if $t \in T^r_s(\mathsf{V})$, then the ***pull-back*** of t by A is $A^*t \in T^r_s(\mathsf{U})$, defined by

$$\begin{aligned}A^*t(&\alpha^1,\dots,\alpha^r,u_1,\dots,u_s)\\ &= t((A^{-1})^*(\alpha^1),\dots,(A^{-1})^*(\alpha^r),A(u_1),\dots,A(u_s)).\end{aligned}$$ •

Remark 2.51. Note the convenient confluence of notation for A^*t when $A \in L(\mathsf{U};\mathsf{V})$ and $t \in T^0_1(\mathsf{V})$. One can think of A^*t as the pull-back of t by A, or as the dual of A applied to t. •

It is important to note in the above definitions that, unless the linear map A is an isomorphism, the notions of push-forward and pull-back are genuinely restricted to tensors of type $(r,0)$ and $(0,s)$, respectively. This will arise also in the geometric notions of push-forward and pull-back, but with additional restrictions being present in that context.

For completeness, let us provide coordinate formulae for push-forward and pull-back. Let U, V, and A be as above, and let $\{f_1,\dots,f_n\}$ and $\{e_1,\dots,e_m\}$ be bases for U and V, respectively. Let A^a_i, $a \in \{1,\dots,m\}$, $i \in \{1,\dots,n\}$, be the components of A with respect to these bases. If A is invertible, let B^i_a, $a,i \in \{1,\dots,n\}$, be the components of $A^{-1} \in L(\mathsf{V};\mathsf{U})$ with respect to the given bases.

1. If $t \in T^r_0(\mathsf{U})$ has components $t^{i_1\cdots i_r}$, then the components of A_*t are

$$(A_*t)^{a_1\cdots a_r} = A^{a_1}_{i_1}\cdots A^{a_r}_{i_r}t^{i_1\cdots i_r}.$$

2. If $t \in T^0_s(\mathsf{V})$ has components $t_{a_1\cdots a_s}$, then the components of A^*t are
$$(A^*t)_{i_1\cdots i_s} = A^{a_1}_{i_1} \cdots A^{a_s}_{i_s} t_{a_1\cdots a_s}.$$
3. If A is an isomorphism and if $t \in T^r_s(\mathsf{U})$ has components $t^{i_1\cdots i_r}_{j_1\cdots j_s}$, then the components of A_*t are
$$(A_*t)^{a_1\cdots a_r}_{b_1\cdots b_s} = A^{a_1}_{i_1} \cdots A^{a_r}_{i_r} B^{j_1}_{b_1} \cdots B^{j_s}_{b_s} t^{i_1\cdots i_r}_{j_1\cdots j_s}.$$
4. If A is an isomorphism and if $t \in T^r_s(\mathsf{V})$ has components $t^{a_1\cdots a_r}_{b_1\cdots b_s}$, then the components of A^*t are
$$(A^*t)^{i_1\cdots i_r}_{j_1\cdots j_s} = B^{i_1}_{a_1} \cdots B^{i_r}_{a_r} A^{b_1}_{j_1} \cdots A^{b_s}_{j_s} t^{a_1\cdots a_r}_{b_1\cdots b_s}.$$

Push-forwards and pull-backs have certain useful properties with respect to composition of linear maps [Abraham, Marsden, and Ratiu 1988].

Proposition 2.52 (Push-forward, pull-back, and composition). *If U, V, and W are $\mathbb{R}$-vector spaces, if $t_1 \in T^r_s(\mathsf{U})$ and $t_2 \in T^r_s(\mathsf{W})$, and if $A \in L(\mathsf{U};\mathsf{V})$ and $B \in L(\mathsf{V};\mathsf{W})$, then the formulae*
$$(B \circ A)_*t_1 = B_*(A_*t_1) \quad \textit{and} \quad (B \circ A)^*t_2 = A^*(B^*t_2)$$
hold, whenever they are well-defined.

2.5 Convexity

While convexity, the topic of this section, is not quite in the domain of linear algebra, now is as good a time as any to present these ideas. We barely graze the surface of the subjects considered in this section, since we shall need only the most basic of concepts. We refer interested readers to the classic text of Rockafellar [1970] on convex analysis, and to the recent text of Boyd and Vandenberghe [2004] on convex optimization.

Definition 2.53 (Convex set). Let V be a $\mathbb{R}$-vector space.

(i) A subset $A \subset \mathsf{V}$ is ***convex*** if $v_1, v_2 \in A$ implies that
$$\{\,(1-t)v_1 + tv_2 \mid \ t \in [0,1]\} \subset A.$$
(ii) If $A \subset \mathsf{V}$ is a general subset, a ***convex combination*** of vectors $v_1, \dots, v_k \in A$ is a linear combination of the form
$$\sum_{j=1}^k \lambda_j v_j, \qquad \lambda_1, \dots, \lambda_k \geq 0,\ \sum_{j=1}^k \lambda_j = 1,\ k \in \mathbb{N}.$$
(iii) The ***convex hull*** of a general subset $A \subset \mathsf{V}$, denoted by $\operatorname{conv}(A)$, is the smallest convex set containing A. •

Thus a set is convex when the line segment connecting any two points in the set lies within the set. It may be verified that a set is convex if and only if it contains all convex combinations of its points. Our definition of convex hull makes sense since the intersection of convex sets is convex. This allows one to assert the existence of the convex hull. One may also show that $\mathrm{conv}(A)$ consists of the union of all convex combinations of elements of A.

The notion of an affine space is also interesting for us.

Definition 2.54 (Affine subspace). Let V be a $\mathbb{R}$-vector space.

(i) A subset A of V is an ***affine subspace*** of V if there exists $v \in \mathsf{V}$ and a subspace U of V for such that $\mathsf{A} = \{v + u \mid \; u \in \mathsf{U}\}$.

(ii) If $A \subset \mathsf{V}$ is a general subset, an ***affine combination*** of vectors $v_1, \dots, v_k \in A$ is a linear combination of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad \lambda_1, \dots, \lambda_k \in \mathbb{R}, \; \sum_{j=1}^{k} \lambda_j = 1, \; k \in \mathbb{N}.$$

(iii) The ***affine hull*** of a general subset $A \subset \mathsf{V}$, denoted by $\mathrm{aff}(A)$, is the smallest affine subspace of V containing A. •

Thus an affine subspace is a "shifted subspace," possibly shifted by the zero vector. Analogously to the convex hull, the definition we give for the affine hull makes sense, since the intersection of affine subspaces is again an affine subspace. This allows one to conclude the existence of the affine hull. One may also show that $\mathrm{aff}(A)$ consists of the union of all affine combinations of elements of A. We refer to Exercise E2.29 for an alternative characterization of an affine subspace.

Exercises

E2.1 Let V be an $\mathbb{F}$-vector space with $\mathsf{U} \subset \mathsf{V}$ a subspace and $\mathsf{U}' \subset \mathsf{V}$ a complement to U in V (i.e., $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}'$). Show that there is a natural isomorphism between U' and V/U.

E2.2 Show that every linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$ has the form exhibited in Example 2.14.

E2.3 Let $\mathcal{H}^n_2$ denote the set of functions $\mathbb{R}^n \to \mathbb{R}$ that are polynomial functions, homogeneous of degree 2, in the variables $(x^1, \dots, x^n)$.
(a) Show that $\mathcal{H}^n_2$ is a $\mathbb{R}$-vector space.
(b) Provide a basis for $\mathcal{H}^n_2$.
(c) Compute the dimension of $\mathcal{H}^n_2$ as a function of n.

E2.4 Let V be a normed $\mathbb{R}$-vector space, and define a linear map ι_{V} from V to V^{**} by $\iota_{\mathsf{V}}(v)(\alpha) = \alpha(v)$. Show that, if V is finite-dimensional, then ι_{V} is an isomorphism.

E2.5 Let V be an $\mathbb{F}$-vector space with $\mathsf{U} \subset \mathsf{V}$ a subspace. Show that $(\mathsf{V}/\mathsf{U})^* \simeq \mathrm{ann}(\mathsf{U})$.

E2.6 Show that a norm $\|\cdot\|$ on a vector space V is the natural norm associated to an inner product if and only if $\|\cdot\|$ satisfies the ***parallelogram law***:

$$\|u+v\|^2 + \|u-v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big).$$

Hint: *"Only if" is easy, but "if" is not; see [Yosida 1980].*

E2.7 Consider $\mathbb{R}^3$ with the standard inner product and let $\mathsf{U} \subset \mathbb{R}^3$ be a two-dimensional subspace.
(a) Define explicitly the orthogonal projection P_{U} onto U.
(b) Define explicitly the reflection map R_{U} that reflects vectors about U.

E2.8 Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ and define $\langle\langle \boldsymbol{A}, \boldsymbol{B}\rangle\rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}^T)$.
(a) Show that $\langle\langle\cdot,\cdot\rangle\rangle$ is an inner product on $\mathbb{R}^{n\times n}$.
(b) Show that the subspaces of symmetric and skew-symmetric matrices are orthogonal with respect to this inner product.
(c) Show that the orthogonal projection $\operatorname{sym}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ onto the set of symmetric matrices is given by $\operatorname{sym}(\boldsymbol{A}) = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T)$.
(d) Show that the orthogonal projection $\operatorname{skew}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ onto the set of skew-symmetric matrices is given by $\operatorname{skew}(\boldsymbol{A}) = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)$.

E2.9 Prove Proposition 2.42.

E2.10 If $B \in \Sigma_2(\mathsf{V})$, show that, thinking of B as a linear map from V to V^*, we have $B = B^*$.

E2.11 Let $\mathbb{G}$ be an inner product on a vector space V, let $\{e_1,\dots,e_n\}$ be a basis for V, and let $\{e^1,\dots,e^n\}$ be the corresponding dual basis.
(a) Show that the matrix representations of $\mathbb{G}$ and $\mathbb{G}^\flat$ agree.
(b) Show that the matrix representations of $\mathbb{G}^{-1}$ and $\mathbb{G}^\sharp$ agree.

E2.12 Let V be a $\mathbb{R}$-vector space and let $B \in \Sigma_2(\mathsf{V})$.
(a) Prove the ***polarization identity***:

$$4B(v_1,v_2) = B(v_1+v_2, v_1+v_2) - B(v_1-v_2, v_1-v_2).$$

A ***quadratic function*** on V is a function $Q\colon \mathsf{V} \to \mathbb{R}$ satisfying $Q(\lambda v) = \lambda^2 Q(v)$ for each $\lambda \in \mathbb{R}$ and $v \in \mathsf{V}$.
(b) Show how the polarization identity provides a means of defining a symmetric bilinear map given a quadratic function on a vector space.

E2.13 Describe the Gram–Schmidt Procedure for obtaining an orthonormal set of vectors starting from a set of linearly independent vectors, and use this procedure to prove Theorem 2.46.

E2.14 Let $[a,b]$ be an interval and let $L_2([a,b])$ be the set of measurable[4] functions $f\colon [a,b] \to \mathbb{R}$ whose square is integrable over $[a,b]$. We refer to such functions as ***square-integrable***. Show that the map $(f,g) \mapsto \int_a^b f(t)g(t)\,\mathrm{d}t$ is an inner product. This map is referred to as the $L_2([a,b])$-inner product; it is convenient to denote it by $\langle\langle f,g\rangle\rangle_{L_2([a,b])}$ and its norm by $\|f\|_{L_2([a,b])}$.

E2.15 A matrix $\boldsymbol{W} \in \mathbb{F}^{n\times n}$ is ***Hermitian*** if $\boldsymbol{W} = \bar{\boldsymbol{W}}^T$.

[4] Readers not familiar with Lebesgue integration can omit "measurable" without conceptual loss.

(a) Show that a Hermitian matrix $\boldsymbol{W}$ has real eigenvalues.

Suppose now that all eigenvalues of the Hermitian matrix $\boldsymbol{W}$ are positive.

(b) Show that $\langle\langle \boldsymbol{x}, \boldsymbol{y} \rangle\rangle_{\boldsymbol{W}} = \boldsymbol{x}^T \boldsymbol{W} \bar{\boldsymbol{y}}$ defines an inner product on $\mathbb{F}^n$.

(c) Show that every inner product on $\mathbb{F}^n$ is obtained as in (b) for some Hermitian matrix $\boldsymbol{W}$ with positive eigenvalues.

E2.16 Given a symmetric linear map A on a finite-dimensional $\mathbb{R}$-inner product space $(\mathsf{V}, \langle\langle \cdot, \cdot \rangle\rangle)$, let $\{v_1, \dots, v_n\}$ be an orthonormal basis of eigenvectors of A with $\{\lambda_1, \dots, v_n\}$ the corresponding eigenvalues.

(a) Show that

$$A = \sum_{j=1}^{n} \lambda_j v_j \otimes v_j^* \tag{E2.1}$$

(by $v_j^* \in V^*$, $j \in \{1, \dots, n\}$, we mean the dual of the linear map from $\mathbb{R}$ to V defined by $a \mapsto a v_j$).

(b) Give a matrix version of (E2.1) in the case that $\mathsf{V} = \mathbb{R}^n$, $\langle\langle \cdot, \cdot \rangle\rangle$ is the standard inner product on $\mathbb{R}^n$, and (by implication) A is defined by a symmetric matrix.

E2.17 Let $(\mathsf{V}, \langle\langle \cdot, \cdot \rangle\rangle)$ be a $\mathbb{R}$-inner product space with $A \in L(\mathsf{V}; \mathsf{V})$ a symmetric linear map. Show that $\ker(A) = \operatorname{image}(A)^{\perp}$.

E2.18 Given two positive-definite symmetric bilinear maps $\mathbb{G}_1$ and $\mathbb{G}_2$ on $\mathbb{R}^n$, show that there exist two constants $0 < \lambda_1 \le \lambda_2$ such that, for all $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\lambda_1 \mathbb{G}_1(\boldsymbol{x}, \boldsymbol{x}) \le \mathbb{G}_2(\boldsymbol{x}, \boldsymbol{x}) \le \lambda_2 \mathbb{G}_1(\boldsymbol{x}, \boldsymbol{x}).$$

E2.19 Recall the list of seven definiteness properties for symmetric bilinear maps from Definition 2.44. Show that this list provides a complete classification of $\Sigma_2(\mathsf{V})$ if V is finite-dimensional.

E2.20 Let V be a finite-dimensional $\mathbb{R}$-vector space and let $B \in \Sigma_2(\mathsf{V})$.

(a) Show that B is positive-semidefinite if and only if $\operatorname{sig}(B) + \operatorname{ind}(B) = \operatorname{rank}(B)$.

(b) Show that B is positive-definite if and only if $\operatorname{sig}(B) + \operatorname{ind}(B) = n$.

(c) Show that B is negative-semidefinite if and only if $\operatorname{ind}(B) = \operatorname{rank}(B)$.

(d) Show that B is negative-definite if and only if $\operatorname{ind}(B) = n$.

E2.21 Let V be an n-dimensional $\mathbb{R}$-vector space and let $B \in \Sigma_2(\mathsf{V})$.

(a) Show that $\dim(\operatorname{image}(B^\flat)) = \operatorname{rank}(B)$.

(b) Show that $\dim(\ker(B^\flat)) = n - \operatorname{rank}(B)$.

E2.22 Let V be a finite-dimensional $\mathbb{R}$-vector space, let $\{e_1, \dots, e_n\}$ be a basis for V, and let $B \in \Sigma_2(\mathsf{V})$. Let $\{\lambda_1, \dots, \lambda_n\}$ be the eigenvalues of the matrix $[B]$.

(a) Show that $\operatorname{ind}(B)$ is equal to the number of negative eigenvalues of $[B]$.

(b) Show that $\operatorname{sig}(B) + \operatorname{ind}(B)$ is equal to the number of positive eigenvalues of $[B]$.

(c) Show that $n - \operatorname{rank}(B)$ is equal to the number of zero eigenvalues of $[B]$.

(d) Are the numerical values (apart from their sign) of the eigenvalues of B meaningful?

E2.23 For V a finite-dimensional $\mathbb{R}$-vector space, show that the following pairs of vector spaces are naturally isomorphic by providing explicit isomorphisms:

(a) $T_0^1(\mathsf{V})$ and V;

(b) $T_1^0(\mathsf{V})$ and V^*;
(c) $T_1^1(\mathsf{V})$ and $L(\mathsf{V};\mathsf{V})$;
(d) $T_2^0(\mathsf{V})$ and $L(\mathsf{V};\mathsf{V}^*)$;
(e) $T_0^2(\mathsf{V})$ and $L(\mathsf{V}^*;\mathsf{V})$.

E2.24 If the components of $t \in T_2^0(\mathbb{R}^2)$ in the standard basis are

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

what is the representation of t as a sum of tensor products?

E2.25 Prove Proposition 2.49.

E2.26 Let $A \in L(\mathbb{R}^2;\mathbb{R}^2)$ have matrix representation

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

in the standard basis. Next, let

$$\boldsymbol{T} = \begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}.$$

Answer the following questions.
(a) If $\boldsymbol{T}$ is the matrix representation of a $(1,1)$-tensor t, what are the push-forward and pull-back of t by A?
(b) If $\boldsymbol{T}$ is the matrix representation of a $(2,0)$-tensor t, what are the push-forward and pull-back of t by A?
(c) If $\boldsymbol{T}$ is the matrix representation of a $(0,2)$-tensor t, what are the push-forward and pull-back of t by A?

E2.27 For each of the following subsets of $\mathbb{R}^2$, determine its convex hull, its affine hull, and the subspace generated by the set:
(a) $\{(1,0),(0,1)\}$;
(b) $\operatorname{span}_{\mathbb{R}}\{(1,0)\}$;
(c) $\{(1,1)+a(1,-1) \mid a \in \mathbb{R}\}$;
(d) $\{(0,1),(-\frac{\sqrt{3}}{2},-\frac{1}{2}),(\frac{\sqrt{3}}{2},-\frac{1}{2})\}$.

E2.28 Let S be a subset of a $\mathbb{R}$-vector space V and let $\operatorname{aff}(S)$ be the affine hull of S. Since $\operatorname{aff}(S)$ is an affine subspace, there exists a subspace $\mathsf{U}_S \subset \mathsf{V}$ such that $\operatorname{aff}(S) = \{v+u \mid u \in \mathsf{U}_S\}$ for some $v \in V$. Show that U_S is generated by the vectors of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad \lambda_1,\dots,\lambda_k \in \mathbb{R},\ \sum_{j=1}^{k} \lambda_j = 0,\ v_1,\dots,v_k \in S,\ k \in \mathbb{N}.$$

E2.29 Let V be a $\mathbb{R}$-vector space. Show that $\mathsf{A} \subset \mathsf{V}$ is an affine subspace if and only if there exists a subspace U of V and a vector $[v] \in \mathsf{V}/\mathsf{U}$ for which $\mathsf{A} = \pi_{\mathsf{U}}^{-1}([v])$, where $\pi_{\mathsf{U}}\colon \mathsf{V} \to \mathsf{V}/\mathsf{U}$ denotes the canonical projection.

Springer

http://www.springer.com/978-0-387-22195-3

Geometric Control of Mechanical Systems
Modeling, Analysis, and Design for Simple Mechanical Control Systems
Bullo, F.; Lewis, A.D.
2005, XXIV, 727 p., Hardcover
ISBN: 978-0-387-22195-3