CHAPTER 1

Vector Space Theory

In order to understand the structure and geometry of multivariate distributions and associated statistical problems, it is essential that we be able to distinguish those aspects of multivariate distributions that can be described without reference to a coordinate system and those that cannot. Finite dimensional vector space theory provides us with a framework in which it becomes relatively easy to distinguish between coordinate free and coordinate concepts. It is fair to say that the material presented in this chapter furnishes the language we use in the rest of this book to describe many of the geometric (coordinate free) and coordinate properties of multivariate probability models. The treatment of vector spaces here is far from complete, but those aspects of the theory that arise in later chapters are covered. Halmos (1958) has been followed quite closely in the first two sections of this chapter, and because of space limitations, proofs sometimes read "see Halmos (1958)."

The material in this chapter runs from the elementary notions of basis, dimension, linear transformation, and matrix to inner product space, orthogonal projection, and the spectral theorem for self-adjoint linear transformations. In particular, the linear space of linear transformations is studied in detail, and the chapter ends with a discussion of what is commonly known as the singular value decomposition theorem. Most of the vector spaces here are finite dimensional real vector spaces, although excursions into infinite dimensions occur via applications of the Cauchy–Schwarz Inequality. As might be expected, we introduce complex coordinate spaces in the discussion of determinants and eigenvalues.

Multilinear algebra and tensors are not covered systematically, although the outer product of vectors and the Kronecker product of linear transformations are covered. It was felt that the simplifications and generality obtained by introducing tensors were not worth the price in terms of added notation, vocabulary, and abstractness.

1.1. VECTOR SPACES

Let R denote the set of real numbers. Elements of R, called scalars, are denoted by $\alpha, \beta, \ldots$.

Definition 1.1. A set V, whose elements are called vectors, is called a real vector space if:

(I) to each pair of vectors $x, y \in V$, there is a vector $x + y \in V$, called the sum of x and y, and for all vectors in V,

(i) $x + y = y + x$.
(ii) $(x + y) + z = x + (y + z)$.
(iii) There exists a unique vector $0 \in V$ such that $x + 0 = x$ for all x.
(iv) For each $x \in V$, there is a unique vector $-x$ such that $x + (-x) = 0$.

(II) For each $\alpha \in R$ and $x \in V$, there is a vector denoted by $\alpha x \in V$, called the product of α and x, and for all scalars and vectors,

(i) $\alpha(\beta x) = (\alpha\beta)x$.
(ii) $1x = x$.
(iii) $(\alpha + \beta)x = \alpha x + \beta x$.
(iv) $\alpha(x + y) = \alpha x + \alpha y$.

In II(iii), $(\alpha + \beta)x$ means the sum of the two scalars, α and β, times x, while $\alpha x + \beta x$ means the sum of the two vectors, αx and βx. This multiple use of the plus sign should not cause any confusion. The reason for calling V a real vector space is that multiplication of vectors by real numbers is permitted.

A classical example of a real vector space is the set R^n of all ordered n-tuples of real numbers. An element of R^n, say x, is represented as

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad x_i \in R, \quad i = 1, \ldots, n,$$

and x_i is called the ith coordinate of x. The vector $x + y$ has ith coordinate $x_i + y_i$ and αx, $\alpha \in R$, is the vector with coordinates αx_i, $i = 1, \ldots, n$. With

$0 \in R^n$ representing the vector of all zeroes, it is routine to check that R^n is a real vector space. Vectors in the coordinate space R^n are always represented by a column of n real numbers as indicated above. For typographical convenience, a vector is often written as a row and appears as $x' = (x_1, \dots, x_n)$. The prime denotes the *transpose* of the vector $x \in R^n$.

The following example provides a method of constructing real vector spaces and yields the space R^n as a special case.

◆ **Example 1.1.** Let $\mathcal{X}$ be a set. The set V is the collection of all the real-valued functions defined on $\mathcal{X}$. For any two elements $x_1, x_2 \in V$, define $x_1 + x_2$ as the function on $\mathcal{X}$ whose value at t is $x_1(t) + x_2(t)$. Also, if $\alpha \in R$ and $x \in V$, αx is the function on $\mathcal{X}$ given by $(\alpha x)(t) \equiv \alpha x(t)$. The symbol $0 \in V$ is the zero function. It is easy to verify that V is a real vector space with these definitions of addition and scalar multiplication. When $\mathcal{X} = \{1, 2, \dots, n\}$, then V is just the real vector space R^n and $x \in R^n$ has as its ith coordinate the value of x at $i \in \mathcal{X}$. Every vector space discussed in the sequel is either V (for some set $\mathcal{X}$) or a linear subspace (to be defined in a moment) of some V. ◆

Before defining the dimension of a vector space, we need to discuss linear dependence and independence. The treatment here follows Halmos (1958, Sections 5–9). Let V be a real vector space.

Definition 1.2. A finite set of vectors $\{x_i | i = 1, \dots, k\}$ is *linearly dependent* if there exist real numbers $\alpha_1, \dots, \alpha_k$, not all zero, such that $\Sigma \alpha_i x_i = 0$. Otherwise, $\{x_i | i = 1, \dots, k\}$ is *linearly independent*.

A brief word about summation notation. Ordinarily, we do not indicate indices of summation on a summation sign when the range of summation is clear from the context. For example, in Definition 1.2, the index i was specified to range between 1 and k before the summation on i appeared; hence, no range was indicated on the summation sign.

An arbitrary subset $S \subseteq V$ is linearly independent if every finite subset of S is linearly independent. Otherwise, S is linearly dependent.

Definition 1.3. A *basis* for a vector space V is a linearly independent set S such that every vector in V is a linear combination of elements of S. V is *finite dimensional* if it has a finite set S that is a basis.

◆ **Example 1.2.** Take $V = R^n$ and let $\varepsilon_i' = (0, \dots, 0, 1, 0, \dots, 0)$ where the one occurs as the ith coordinate of ε_i, $i = 1, \dots, n$. For $x \in R^n$, it is clear that $x = \Sigma x_i \varepsilon_i$ where x_i is the ith coordinate of x. Thus every vector in R^n is a linear combination of $\varepsilon_1, \dots, \varepsilon_n$. To show that $\{\varepsilon_i | i = 1, \dots, n\}$ is a linearly independent set, suppose $\Sigma \alpha_i \varepsilon_i = 0$ for some scalars α_i, $i = 1, \dots, n$. Then $x = \Sigma \alpha_i \varepsilon_i = 0$ has α_i as its ith coordinate, so $\alpha_i = 0$, $i = 1, \dots, n$. Thus $\{\varepsilon_i | i = 1, \dots, n\}$ is a basis for R^n and R^n is finite dimensional. The basis $\{\varepsilon_i | i = 1, \dots, n\}$ is called the *standard basis* for R^n. ◆

Let V be a finite dimensional real vector space. The basic properties of linearly independent sets and bases are:

(i) If $\{x_1, \dots, x_m\}$ is a linearly independent set in V, then there exist vectors $x_{m+1}, \dots, x_{m+k}$ such that $\{x_1, \dots, x_{m+k}\}$ is a basis for V.
(ii) All bases for V have the same number of elements. The *dimension* of V is defined to be the number of elements in any basis.
(iii) Every set of $n + 1$ vectors in an n-dimensional vector space is linearly dependent.

Proofs of the above assertions can be found in Halmos (1958, Sections 5–8). The dimension of a finite dimensional vector space is denoted by $\dim(V)$. If $\{x_1, \dots, x_n\}$ is a basis for V, then every $x \in V$ is a unique linear combination of $\{x_1, \dots, x_n\}$—say $x = \Sigma \alpha_i x_i$. That every x can be so expressed follows from the definition of a basis and the uniqueness follows from the linear independence of $\{x_1, \dots, x_n\}$. The numbers $\alpha_1, \dots, \alpha_n$ are called the *coordinates* of x in the basis $\{x_1, \dots, x_n\}$. Clearly, the coordinates of x depend on the order in which we write the basis. Thus by a basis we always mean an ordered basis.

We now introduce the notion of a subspace of a vector space.

Definition 1.4. A nonempty subset $M \subseteq V$ is a *subspace* (or *linear manifold*) of V if, for each $x, y \in M$ and $\alpha, \beta \in R$, $\alpha x + \beta y \in M$.

A subspace M of a real vector space V is easily shown to satisfy the vector space axioms (with addition and scalar multiplication inherited from V), so subspaces are real vector spaces. It is not difficult to verify the following assertions (Halmos, 1958, Sections 10–12):

(i) The intersection of subspaces is a subspace.
(ii) If M is a subspace of a finite dimensional vector space V, then $\dim(M) \leqslant \dim(V)$.

(iii) Given an m-dimensional subspace M of an n-dimensional vector space V, there is a basis $\{x_1, \ldots, x_m, \ldots, x_n\}$ for V such that $\{x_1, \ldots, x_m\}$ is a basis for M.

Given any set $S \subseteq V$, span(S) is defined to be the intersection of all the subspaces that contain S—that is, span(S) is the smallest subspace that contains S. It is routine to show that span(S) is equal to the set of all linear combinations of elements of S. The subspace span(S) is often called the subspace spanned by the set S.

If M and N are subspaces of V, then span($M \cup N$) is the set of all vectors of the form $x + y$ where $x \in M$ and $y \in N$. The suggestive notation $M + N \equiv \{z | z = x + y, x \in M, y \in N\}$ is used for span($M \cup N$) when M and N are subspaces. Using the fact that a linearly independent set can be extended to a basis in a finite dimensional vector space, we have the following. Let V be finite dimensional and suppose M and N are subspaces of V.

(i) Let $m = \dim(M)$, $n = \dim(N)$, and $k = \dim(M \cap N)$. Then there exist vectors $x_1, \ldots, x_k, y_{k+1}, \ldots, y_m$, and $z_{k+1}, \ldots, z_n$ such that $\{x_1, \ldots, x_k\}$ is a basis for $M \cap N$, $\{x_1, \ldots, x_k, y_{k+1}, \ldots, y_m\}$ is a basis for M, $\{x_1, \ldots, x_k, z_{k+1}, \ldots, z_n\}$ is a basis for N, and $\{x_1, \ldots, x_k, y_{k+1}, \ldots, y_m, z_{k+1}, \ldots, z_n\}$ is a basis for $M + N$. If $k = 0$, then $\{x_1, \ldots, x_k\}$ is interpreted as the empty set.

(ii) $\dim(M + N) = \dim(M) + \dim(N) - \dim(M \cap N)$.

(iii) There exists a subspace $M_1 \subseteq V$ such that $M \cap M_1 = \{0\}$ and $M + M_1 = V$.

Definition 1.5. If M and N are subspaces of V that satisfy $M \cap N = \{0\}$ and $M + N = V$, then M and N are *complementary* subspaces.

The technique of decomposing a vector space into two (or more) complementary subspaces arises again and again in the sequel. The basic property of such a decomposition is given in the following proposition.

Proposition 1.1. Suppose M and N are complementary subspaces in V. Then each $x \in V$ has a unique representation $x = y + z$ with $y \in M$ and $z \in N$.

Proof. Since $M + N = V$, each $x \in V$ can be written $x = y_1 + z_1$ with $y_1 \in M$ and $z_1 \in N$. If $x = y_2 + z_2$ with $y_2 \in M$ and $z_2 \in N$, then $0 = x -$

$x = (y_1 - y_2) + (z_1 - z_2)$. Hence $(y_2 - y_1) = (z_1 - z_2)$ so $(y_2 - y_1) \in M \cap N = \{0\}$. Thus $y_1 = y_2$. Similarly, $z_1 = z_2$. □

The above proposition shows that we can decompose the vector space V into two vector spaces M and N and each x in V has a unique piece in M and in N. Thus x can be represented as (y, z) with $y \in M$ and $z \in N$. Also, note that if $x_1, x_2 \in V$ and have the representations (y_1, z_1), (y_2, z_2), then $\alpha x_1 + \beta x_2$ has the representation $(\alpha y_1 + \beta y_2, \alpha z_1 + \beta z_2)$, for $\alpha, \beta \in R$. In other words the function that maps x into its decomposition (y, z) is linear. To make this a bit more precise, we now define the direct sum of two vector spaces.

Definition 1.6. Let V_1 and V_2 be two real vector spaces. The *direct sum* of V_1 and V_2, denoted by $V_1 \oplus V_2$, is the set of all ordered pairs $\{x, y\}$, $x \in V_1$, $y \in V_2$, with the linear operations defined by $\alpha_1\{x_1, y_1\} + \alpha_2\{x_2, y_2\} \equiv \{\alpha_1 x_1 + \alpha_2 x_2, \alpha_1 y_1 + \alpha_2 y_2\}$.

That $V_1 \oplus V_2$ is a real vector space with the above operations can easily be verified. Further, identifying V_1 with $\{\{x_1, 0\} | x \in V_1\} \equiv \tilde{V}_1$ and V_2 with $\{\{0, y\} | y \in V_2\} \equiv \tilde{V}_2$, we can think of V_1 and V_2 as complementary subspaces of $V_1 \oplus V_2$, since $\tilde{V}_1 + \tilde{V}_2 = V_1 \oplus V_2$ and $\tilde{V}_1 \cap \tilde{V}_2 = \{0, 0\}$, which is the zero element in $V_1 \oplus V_2$. The relation of the direct sum to our previous decomposition of a vector space should be clear.

◆ **Example 1.3.** Consider $V = R^n$, $n \geqslant 2$, and let p and q be positive integers such that $p + q = n$. Then R^p and R^q are both real vector spaces. Each element of R^n is a n-tuple of real numbers, and we can construct subspaces of R^n by setting some of these coordinates equal to zero. For example, consider $M = \{x \in R^n | x = \begin{pmatrix} y \\ 0 \end{pmatrix}$ with $y \in R^p, 0 \in R^q\}$ and $N = \{x \in R^n | x = \begin{pmatrix} 0 \\ z \end{pmatrix}$ with $0 \in R^p$ and $z \in R^q\}$. It is clear that $\dim(M) = p$, $\dim(N) = q$, $M \cap N = \{0\}$, and $M + N = R^n$. The identification of R^p with M and R^q with N shows that it is reasonable to write $R^p \oplus R^q = R^{p+q}$. ◆

1.2. LINEAR TRANSFORMATIONS

Linear transformations occupy a central position, both in vector space theory and in multivariate analysis. In this section, we discuss the basic properties of linear transforms, leaving the deeper results for consideration after the introduction of inner products. Let V and W be real vector spaces.

Definition 1.7. Any function A defined on V and taking values in W is called a *linear transformation* if $A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A(x_1) + \alpha_2 A(x_2)$ for all $x_1, x_2 \in V$ and $\alpha_1, \alpha_2 \in R$.

Frequently, $A(x)$ is written Ax when there is no danger of confusion. Let $\mathcal{L}(V, W)$ be the set of all linear transformations on V to W. For two linear transformations A_1 and A_2 in $\mathcal{L}(V, W)$, $A_1 + A_2$ is defined by $(A_1 + A_2)(x) = A_1 x + A_2 x$ and $(\alpha A)(x) = \alpha Ax$ for $\alpha \in R$. The zero linear transformation is denoted by 0. It should be clear that $\mathcal{L}(V, W)$ is a real vector space with these definitions of addition and scalar multiplication.

◆ **Example 1.4.** Suppose $\dim(V) = m$ and let $x_1, \dots, x_m$ be a basis for V. Also, let $y_1, \dots, y_m$ be arbitrary vectors in W. The claim is that there is a unique linear transformation A such that $Ax_i = y_i$, $i = 1, \dots, m$. To see this, consider $x \in V$ and express x as a unique linear combination of the basis vectors, $x = \Sigma \alpha_i x_i$. Define A by

$$Ax = \sum_i^n \alpha_i A x_i = \sum_1^n \alpha_i y_i.$$

The linearity of A is easy to check. To show that A is unique, let B be another linear transformation with $Bx_i = y_i$, $i = 1, \dots, n$. Then $(A - B)(x_i) = 0$ for $i = 1, \dots, n$, and $(A - B)(x) = (A - B)(\Sigma \alpha_i x_i) = \Sigma \alpha_i (A - B)(x_i) = 0$ for all $x \in V$. Thus $A = B$. ◆

The above example illustrates a general principle—namely, a linear transformation is completely determined by its values on a basis. This principle is used often to construct linear transformations with specified properties. A modification of the construction in Example 1.4 yields a basis for $\mathcal{L}(V, W)$ when V and W are both finite dimensional. This basis is given in the proof of the following proposition.

Proposition 1.2. If $\dim(V) = m$ and $\dim(W) = n$, then $\dim(\mathcal{L}(V, W)) = mn$.

Proof. Let $x_1, \dots, x_m$ be a basis for V and let $y_1, \dots, y_n$ be a basis for W. Define a linear transformation A_{ji}, $i = 1, \dots, m$ and $j = 1, \dots, n$, by

$$A_{ji}(x_k) = \begin{cases} 0 & \text{if } k \neq i \\ y_j & \text{if } k = i \end{cases}.$$

For each (j, i), A_{ji} has been defined on a basis in V so the linear transformation A_{ji} is uniquely determined. We now claim that $\{A_{ji}|i = 1,\ldots, m; j = 1,\ldots, n\}$ is a basis for $\mathcal{L}(V, W)$. To show linear independence, suppose $\Sigma\Sigma\alpha_{ji}A_{ji} = 0$. Then for each $k = 1,\ldots, m$,

$$0 = \sum_j \sum_i \alpha_{ji} A_{ji}(x_k) = \sum_j \alpha_{jk} y_j.$$

Since $\{y_1,\ldots, y_m\}$ is a linearly independent set, this implies that $\alpha_{jk} = 0$ for all j and k. Thus linear independence holds. To show every $A \in \mathcal{L}(V, W)$ is a linear combination of the A_{ji}, first note that Ax_k is a vector in W and thus is a unique linear combination of $y_1,\ldots, y_n$, say $Ax_k = \Sigma_j a_{jk} y_j$ where $a_{jk} \in R$. However, the linear transformation $\Sigma\Sigma a_{ji}A_{ji}$ evaluated at x_k is

$$\sum_j \sum_i a_{ji} A_{ji}(x_k) = \sum_j a_{jk} y_j.$$

Since A and $\Sigma\Sigma a_{ji}A_{ji}$ agree on a basis in V, they are equal. This completes the proof since there are mn elements in the basis $\{A_{ji}|i = 1,\ldots, m; j = 1,\ldots, n\}$ for $\mathcal{L}(V, W)$. □

Since $\mathcal{L}(V, W)$ is a vector space, general results about vector spaces, of course, apply to $\mathcal{L}(V, W)$. However, linear transformations have many interesting properties not possessed by vectors in general. For example, consider vector spaces V_i, $i = 1, 2, 3$. If $A \in \mathcal{L}(V_1, V_2)$ and $B \in \mathcal{L}(V_2, V_3)$, then we can compose the functions B and A by defining $(BA)(x) = B(A(x))$. The linearity of A and B implies that BA is a linear transformation on V_1 to V_3—that is, $BA \in \mathcal{L}(V_1, V_3)$. Usually, BA is called the product of B and A.

There are two special cases of $\mathcal{L}(V, W)$ that are of particular interest. First, if $A, B \in \mathcal{L}(V, V)$, then $AB \in \mathcal{L}(V, V)$ and $BA \in \mathcal{L}(V, V)$, so we have a multiplication defined in $\mathcal{L}(V, V)$. However, this multiplication is not commutative—that is, AB is not, in general, equal to BA. Clearly, $A(B + C) = AB + AC$ for $A, B, C \in \mathcal{L}(V, V)$. The identity linear transformation in $\mathcal{L}(V, V)$, usually denoted by I, satisfies $AI = IA = A$ for all $A \in \mathcal{L}(V, V)$, since $Ix = x$ for all $x \in V$. Thus $\mathcal{L}(V, V)$ is not only a vector space, but there is a multiplication defined in $\mathcal{L}(V, V)$.

The second special case of $\mathcal{L}(V, W)$ we wish to consider is when $W = R$—that is, W is the one-dimensional real vector space R. The space $\mathcal{L}(V, R)$ is called the *dual space* of V and, if $\dim(V) = n$, then $\dim(\mathcal{L}(V, R)) = n$. Clearly, $\mathcal{L}(V, R)$ is the vector space of all real-valued linear functions defined on V. We have more to say about $\mathcal{L}(V, R)$ after the introduction of inner products on V.

Understanding the geometry of linear transformations usually begins with a specification of the range and null space of the transformation. These objects are now defined. Let $A \in \mathcal{L}(V, W)$ where V and W are finite dimensional.

Definition 1.8. The *range of A*, denoted by $\mathfrak{R}(A)$, is

$$\mathfrak{R}(A) \equiv \{u | u \in W, Ax = u \text{ for some } x \in V\}.$$

The *null space of A*, denoted by $\mathfrak{N}(A)$, is

$$\mathfrak{N}(A) \equiv \{x | x \in V, Ax = 0\}.$$

It is routine to verify that $\mathfrak{R}(A)$ is a subspace of W and $\mathfrak{N}(A)$ is a subspace of V. The *rank of A*, denoted by $r(A)$, is the dimension of $\mathfrak{R}(A)$.

Proposition 1.3. If $A \in \mathcal{L}(V, W)$ and $n = \dim(V)$, then $r(A) + \dim(\mathfrak{N}(A)) = n$.

Proof. Let M be a subspace of V such that $M \oplus \mathfrak{N}(A) = V$, and consider a basis $\{x_1, \dots, x_k\}$ for M. Since $\dim(M) + \dim(\mathfrak{N}(A)) = n$, we need to show that $k = r(A)$. To do this, it is sufficient to show that $\{Ax_1, \dots, Ax_k\}$ is a basis for $\mathfrak{R}(A)$. If $0 = \Sigma\alpha_i Ax_i = A(\Sigma\alpha_i x_i)$, then $\Sigma\alpha_i x_i \in M \cap \mathfrak{N}(A)$ so $\Sigma\alpha_i x_i = 0$. Hence $\alpha_1 = \cdots = \alpha_k = 0$ as $\{x_1, \dots, x_k\}$ is a basis for M. Thus $\{Ax_1, \dots, Ax_k\}$ is a linearly independent set. To verify that $\{Ax_1, \dots, Ax_k\}$ spans $\mathfrak{R}(A)$, suppose $w \in \mathfrak{R}(A)$. Then $w = Ax$ for some $x \in V$. Write $x = y + z$ where $y \in M$ and $z \in \mathfrak{N}(A)$. Then $w = A(y + z) = Ay$. Since $y \in M$, $y = \Sigma\alpha_i x_i$ for some scalars $\alpha_1, \dots, \alpha_k$. Therefore, $w = A(\Sigma\alpha_i x_i) = \Sigma\alpha_i Ax_i$. □

Definition 1.9. A linear transformation $A \in \mathcal{L}(V, V)$ is called *invertible* if there exists a linear transformation, denoted by A^{-1}, such that $AA^{-1} = A^{-1}A = I$.

The following assertions hold; see Halmos (1958, Section 36):

(i) A is invertible iff $\mathfrak{R}(A) = V$ iff $Ax = 0$ implies $x = 0$.
(ii) If $A, B, C \in \mathcal{L}(V, V)$ and if $AB = CA = I$, then A is invertible and $B = C = A^{-1}$.
(iii) If A and B are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. If A is invertible and $\alpha \neq 0$, then $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$ and $(A^{-1})^{-1} = A$.

In terms of bases, invertible transformations are characterized by the following.

Proposition 1.4. Let $A \in \mathcal{L}(V, V)$ and suppose $\{x_1, \dots, x_n\}$ is a basis for V. The following are equivalent:

(i) A is invertible.
(ii) $\{Ax_1, \dots, Ax_n\}$ is a basis for V.

Proof. Suppose A is invertible. Since $\dim(V) = n$, we must show $\{Ax_1, \dots, Ax_n\}$ is a linearly independent set. Thus if $0 = \Sigma\alpha_i Ax_i = A(\Sigma\alpha_i x_i)$, then $\Sigma\alpha_i x_i = 0$ since A is invertible. Hence $\alpha_i = 0$, $i = 1, \dots, n$, as $\{x_1, \dots, x_n\}$ is a basis for V. Therefore, $\{Ax_1, \dots, Ax_n\}$ is a basis.

Conversely, suppose $\{Ax_1, \dots, Ax_n\}$ is a basis. We show that $Ax = 0$ implies $x = 0$. First, write $x = \Sigma\alpha_i x_i$ so $Ax = 0$ implies $\Sigma\alpha_i Ax_i = 0$. Hence $\alpha_i = 0$, $i = 1, \dots, n$, as $\{Ax_1, \dots, Ax_n\}$ is a basis. Thus $x = 0$, so A is invertible. □

We now introduce real matrices and consider their relation to linear transformations. Consider vector spaces V and W of dimension m and n, respectively, and bases $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_n\}$ for V and W. Each $x \in V$ has a unique representation $x = \Sigma\alpha_i x_i$. Let $[x]$ denote the column vector of coordinates of x in the given basis. Thus $[x] \in R^m$ and the ith coordinate of $[x]$ is α_i, $i = 1, \dots, m$. Similarly, $[y] \in R^n$ is the column vector of $y \in W$ in the basis $\{y_1, \dots, y_n\}$. Consider $A \in \mathcal{L}(V, W)$ and express Ax_j in the given basis of W; $Ax_j = \Sigma_i a_{ij} y_i$ for unique scalars a_{ij}, $i = 1, \dots, n$, $j = 1, \dots, m$. The $n \times m$ rectangular array of real scalars

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \\ a_{n1} & & \cdots & a_{nm} \end{bmatrix} \equiv \{a_{ij}\}$$

is called the *matrix* of A relative to the two given bases. Conversely, given any $n \times m$ rectangular array of real scalars $\{a_{ij}\}$, $i = 1, \dots, n$, $j = 1, \dots, m$, the linear transformation A defined by $Ax_j = \Sigma_i a_{ij} y_i$ has as its matrix $[A] = \{a_{ij}\}$.

Definition 1.10. A rectangular array $\{a_{ij}\}: m \times n$ of real scalars is called an $m \times n$ *matrix*. If $A = \{a_{ij}\}: m \times n$ is a matrix and $B = \{b_{ij}\}: n \times p$ is a matrix, then $C = AB$, called the *matrix product* of A and B (in that order) is defined to be the matrix $\{c_{ij}\}: m \times p$ with $c_{ij} = \Sigma_k a_{ik} b_{kj}$.

In this book, the distinction between linear transformations, matrices, and the matrix of a linear transformation is always made. The notation $[A]$ means the matrix of a linear transformation with respect to two given bases. However, symbols like A, B, or C may represent either linear transformations or real matrices; care is taken to clearly indicate which case is under consideration.

Each matrix $A = \{a_{ij}\} : m \times n$ defines a linear transformation on R^n to R^m as follows. For $x \in R^n$ with coordinates $x_1, \ldots, x_n$, Ax is the vector y in R^m with coordinates $y_i = \sum_j a_{ij} x_j$, $i = 1, \ldots, m$. Of course, this is the usual row by column rule of a matrix operating on a vector. The matrix of this linear transformation in the standard bases for R^n and R^m is just the matrix A. However, if the bases are changed, then the matrix of the linear transformation changes. When $m = n$, the matrix $A = \{a_{ij}\}$ determines a linear transformation on R^n to R^n via the above definition of a matrix times a vector. The matrix A is called nonsingular (or invertible) if there exists a matrix, denoted by A^{-1}, such that $AA^{-1} = A^{-1}A = I_n$ where I_n is the $n \times n$ identity matrix consisting of ones on the diagonal and zeroes off the diagonal. As with linear transformations, A^{-1} is unique and exists iff $Ax = 0$ implies $x = 0$.

The symbol $\mathcal{L}_{n,m}$ denotes the real vector space of $m \times n$ real matrices with the usual operations of addition and scalar multiplication. In other words, if $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ are elements of $\mathcal{L}_{n,m}$, then $A + B = \{a_{ij} + b_{ij}\}$ and $\alpha A = \{\alpha a_{ij}\}$. Notice that $\mathcal{L}_{n,m}$ is the set of $m \times n$ matrices (m and n are in reverse order). The reason for writing $\mathcal{L}_{n,m}$ is that an $m \times n$ matrix determines a linear transformation from R^n to R^m. We have made the choice of writing $\mathcal{L}(V, W)$ for linear transformations from V to W, and it is an unpleasant fact that the dimensions of a matrix occur in reverse order to the dimensions of the spaces V and W. The next result summarizes the relations between linear transformations and matrices.

Proposition 1.5. Consider vector spaces V_1, V_2, and V_3 with bases $\{x_1, \ldots, x_{n_1}\}$, $\{y_1, \ldots, y_{n_2}\}$, and $\{z_1, \ldots, z_{n_3}\}$, respectively. For $x \in V_1$, $y \in V_2$, and $z \in V_3$, let $[x]$, $[y]$, and $[z]$ denote the vector of coordinates of x, y, and z in the given bases, so $[x] \in R^{n_1}$, $[y] \in R^{n_2}$, and $[z] \in R^{n_3}$. For $A \in \mathcal{L}(V_1, V_2)$ and $B \in \mathcal{L}(V_2, V_3)$ let $[A]$ ($[B]$) denote the matrix of $A(B)$ relative to the bases $\{x_1, \ldots, x_{n_1}\}$ and $\{y_1, \ldots, y_{n_2}\}$ ($\{y_1, \ldots, y_{n_2}\}$ and $\{z_1, \ldots, z_{n_3}\}$). Then:

(i) $[Ax] = [A][x]$.
(ii) $[BA] = [B][A]$.
(iii) If $V_1 = V_2$ and A is invertible, $[A^{-1}] = [A]^{-1}$. Here, $[A^{-1}]$ and $[A]$ are matrices in the bases $\{x_1, \ldots, x_{n_1}\}$ and $\{x_1, \ldots, x_{n_1}\}$.

Proof. A few words are in order concerning the notation in (i), (ii), and (iii). In (i), $[Ax]$ is the vector of coordinates of $Ax \in V_2$ with respect to the basis $\{y_1, \ldots, y_{n_2}\}$ and $[A][x]$ means the matrix $[A]$ times the coordinate vector $[x]$ as defined previously. Since both sides of (i) are linear in x, it suffices to verify (i) for $x = x_j, j = 1, \ldots, n_1$. But $[A][x_j]$ is just the column vector with coordinates a_{ij}, $i = 1, \ldots, n_2$, and $Ax_j = \sum_i a_{ij} y_i$, so $[Ax_j]$ is the column vector with coordinates a_{ij}, $i = 1, \ldots, n_2$. Hence (i) holds.

For (ii), $[B][A]$ is just the matrix product of $[B]$ and $[A]$. Also, $[BA]$ is the matrix of the linear transformation $BA \in \mathcal{L}(V_1, V_3)$ with respect to the bases $\{x_1, \ldots, x_{n_1}\}$ and $\{z_1, \ldots, z_{n_3}\}$. To show that $[BA] = [B][A]$, we must verify that, for all $x \in V$, $[BA][x] = [B][A][x]$. But by (i), $[BA][x] = [BAx]$ and, using (i) twice, $[B][A][x] = [B][Ax] = [BAx]$. Thus (ii) is established.

In (iii), $[A]^{-1}$ denotes the inverse of the matrix $[A]$. Since A is invertible, $AA^{-1} = A^{-1}A = I$ where I is the identity linear transformation on V_1 to V_1. Thus by (ii), with I_n denoting the $n \times n$ identity matrix, $I_n = [I] = [AA^{-1}] = [A][A^{-1}] = [A^{-1}A] = [A^{-1}][A]$. By the uniqueness of the matrix inverse, $[A^{-1}] = [A]^{-1}$. □

Projections are the final topic in this section. If V is a finite dimensional vector space and M and N are subspaces of V such that $M \oplus N = V$, we have seen that each $x \in V$ has a unique piece in M and a unique piece in N. In other words, $x = y + z$ where $y \in M$, $z \in N$, and y and z are unique.

Definition 1.11. Given subspaces M and N in V such that $M \oplus N = V$, if $x = y + z$ with $y \in M$ and $z \in N$, then y is called the *projection* of x on M along N and z is called the projection of x on N along M.

Since M and N play symmetric roles in the above definition, we concentrate on the projection on M.

Proposition 1.6. The function P mapping V into V whose value at x is the projection of x on M along N is a linear transformation that satisfies

(i) $\mathfrak{R}(P) = M$, $\mathfrak{N}(P) = N$.
(ii) $P^2 = P$.

Proof. We first show that P is linear. If $x = y + z$ with $y \in M$, $z \in N$, then by definition, $Px = y$. Also, if $x_1 = y_1 + z_1$ and $x_2 = y_2 + z_2$ are the decompositions of x_1 and x_2, respectively, then $\alpha_1 x_1 + \alpha_2 x_2 = (\alpha_1 y_1 + \alpha_2 y_2) + (\alpha_1 z_1 + \alpha_2 z_2)$ is the decomposition of $\alpha_1 x_1 + \alpha_2 x_2$. Thus $P(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 P x_1 + \alpha_2 P x_2$ so P is linear. By definition $Px \in M$, so $\mathfrak{R}(P)$

$\subseteq M$. But if $x \in M$, $Px = x$ and $\mathcal{R}(P) = M$. Also, if $x \in N$, $Px = 0$ so $\mathcal{N}(P) \supseteq N$. However, if $Px = 0$, then $x = 0 + x$, and therefore $x \in N$. Thus $\mathcal{N}(P) = N$. To show $P^2 = P$, note that $Px \in M$ and $Px = x$ for $x \in M$. Hence, $Px = P(Px) = P^2x$, which implies that $P = P^2$. □

A converse to Proposition 1.6 gives a complete description of all linear transformations on V to V that satisfy $A^2 = A$.

Proposition 1.7. If $A \in \mathcal{L}(V, V)$ and satisfies $A^2 = A$, then $\mathcal{R}(A) \oplus \mathcal{N}(A) = V$ and A is the projection on $\mathcal{R}(A)$ along $\mathcal{N}(A)$.

Proof. To show $\mathcal{R}(A) \oplus \mathcal{N}(A) = V$, we must verify that $\mathcal{R}(A) \cap \mathcal{N}(A) = \{0\}$ and that each $x \in V$ is the sum of a vector in $\mathcal{R}(A)$ and a vector in $\mathcal{N}(A)$. If $x \in \mathcal{R}(A) \cap \mathcal{N}(A)$, then $x = Ay$ for some $y \in V$ and $Ax = 0$. Since $A^2 = A$, $0 = Ax = A^2y = Ay = x$ and $\mathcal{R}(A) \cap \mathcal{N}(A) = \{0\}$. For $x \in V$, write $x = Ax + (I - A)x$ and let $y = Ax$ and $z = (I - A)x$. Then $y \in \mathcal{R}(A)$ by definition and $Az = A(I - A)x = (A - A^2)x = 0$, so $z \in \mathcal{N}(A)$. Thus $\mathcal{R}(A) \oplus \mathcal{N}(A) = V$.

The verification that A is the projection on $\mathcal{R}(A)$ along $\mathcal{N}(A)$ goes as follows. A is zero on $\mathcal{N}(A)$ by definition. Also, for $x \in \mathcal{R}(A)$, $x = Ay$ for some $y \in V$. Thus $Ax = A^2y = Ay = x$, so $Ax = x$ and $x \in \mathcal{R}(A)$. However, the projection on $\mathcal{R}(A)$ along $\mathcal{N}(A)$, say P, also satisfies $Px = x$ for $x \in \mathcal{R}(A)$ and $Px = 0$ for $x \in \mathcal{N}(A)$. This implies that $P = A$ since $\mathcal{R}(A) \oplus \mathcal{N}(A) = V$. □

The above proof shows that the projection on M along N is the unique linear transformation that is the identity on M and zero on N. Also, it is clear that P is the projection on M along N iff $I - P$ is the projection on N along M.

1.3. INNER PRODUCT SPACES

The discussion of the previous section was concerned mainly with the linear aspects of vector spaces. Here, we introduce inner products on vector spaces so that the geometric notions of length, angle, and orthogonality become meaningful. Let us begin with an example.

◆ **Example 1.5.** Consider coordinate space R^n with the standard basis $\{\varepsilon_1, \dots, \varepsilon_n\}$. For $x, y \in R^n$, define $x'y \equiv \Sigma x_i y_i$ where x and y have coordinates $x_1, \dots, x_n$ and $y_1, \dots, y_n$. Of course, x' is the transpose of the vector x and $x'y$ can be thought of as the $1 \times n$

matrix x' times the $n \times 1$ matrix y. The real number $x'y$ is sometimes called the scalar product (or inner product) of x and y. Some properties of the scalar product are:

(i) $x'y = y'x$ (symmetry).
(ii) $x'y$ is linear in y for fixed x and linear in x for fixed y.
(iii) $x'x = \Sigma_1^n x_i^2 \geqslant 0$ and is zero iff $x = 0$.

The *norm* of x, defined by $\|x\| = (x'x)^{1/2}$, can be thought of as the distance between x and $0 \in R^n$. Hence, $\|x - y\| = (\Sigma(x_i - y_i)^2)^{1/2}$ is usually called the distance between x and y. When x and y are both not zero, then the cosine of the angle between x and y is $x'y/\|x\|\|y\|$ (see Halmos, 1958, p. 118). Thus we have a geometric interpretation of the scalar product. In particular, the angle between x and y is $\pi/2(\cos \pi/2 = 0)$ iff $x'y = 0$. Thus we say x and y are orthogonal (perpendicular) iff $x'y = 0$. ◆

Let V be a real vector space. An inner product on V is obtained by simply abstracting the properties of the scalar product on R^n.

Definition 1.12. An *inner product* on a real vector space V is a real valued function on $V \times V$, denoted by $(\cdot, \cdot)$, with the following properties:

(i) $(x, y) = (y, x)$ (symmetry).
(ii) $(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1(x_1, y) + \alpha_2(x_2, y)$ (linearity).
(iii) $(x, x) \geqslant 0$ and $(x, x) = 0$ only if $x = 0$ (positivity).

From (i) and (ii) it follows that $(x, \alpha_1 y_1 + \alpha_2 y_2) = \alpha_1(x, y_1) + \alpha_2(x, y_2)$. In other words, inner products are linear in each variable when the other variable is fixed. The *norm* of x, denoted by $\|x\|$, is defined to be $\|x\| = (x, x)^{1/2}$ and the *distance* between x and y is $\|x - y\|$. Hence geometrically meaningful names and properties related to the scalar product on R^n have become definitions on V. To establish the existence of inner products on finite dimensional vector spaces, we have the following proposition.

Proposition 1.8. Suppose $\{x_1, \ldots, x_n\}$ is a basis for the real vector space V. The function $(\cdot, \cdot)$ defined on $V \times V$ by $(x, y) = \Sigma_1^n \alpha_i \beta_i$, where $x = \Sigma \alpha_i x_i$ and $y = \Sigma \beta_i x_i$, is an inner product on V.

Proof. Clearly $(x, y) = (y, x)$. If $x = \Sigma \alpha_i x_i$ and $z = \Sigma \gamma_i x_i$, then $(\alpha x + \gamma z, y) = \Sigma(\alpha\alpha_i + \gamma\gamma_i)\beta_i = \alpha\Sigma\alpha_i\beta_i + \gamma\Sigma\gamma_i\beta_i = \alpha(x, y) + \gamma(z, y)$. This establishes the linearity. Also, $(x, x) = \Sigma\alpha_i^2$, which is zero iff all the α_i are zero and this is equivalent to x being zero. Thus $(\cdot, \cdot)$ is an inner product on V. □

A vector space V with a given inner product $(\cdot, \cdot)$ is called an *inner product space*.

Definition 1.13. Two vectors x and y in an inner product space $(V, (\cdot, \cdot))$ are *orthogonal*, written $x \perp y$, if $(x, y) = 0$. Two subsets S_1 and S_2 of V are *orthogonal*, written $S_1 \perp S_2$, if $x \perp y$ for all $x \in S_1$ and $y \in S_2$.

Definition 1.14. Let $(V, (\cdot, \cdot))$ be a finite dimensional inner product space. A set of vectors $\{x_1, \ldots, x_k\}$ is called an *orthonormal set* if $(x_i, x_j) = \delta_{ij}$ for $i, j = 1, \ldots, k$ where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$. A set $\{x_1, \ldots, x_k\}$ is called an *orthonormal basis* if the set is both a basis and is orthonormal.

First note that an orthonormal set $\{x_1, \ldots, x_k\}$ is linearly independent. To see this, suppose $0 = \Sigma\alpha_i x_i$. Then $0 = (0, x_j) = (\Sigma\alpha_i x_i, x_j) = \Sigma\alpha_i(x_i, x_j) = \Sigma_i\alpha_i\delta_{ij} = \alpha_j$. Hence $\alpha_j = 0$ for $j = 1, \ldots, k$ and the set $\{x_1, \ldots, x_k\}$ is linearly independent.

In Proposition 1.8, the basis used to define the inner product is, in fact, an orthonormal basis for the inner product. Also, the standard basis for R^n is an orthonormal basis for the scalar product on R^n—this scalar product is called the *standard inner product* on R^n. An algorithm for constructing orthonormal sets from linearly independent sets is now given. It is known as the *Gram–Schmidt orthogonalization procedure*.

Proposition 1.9. Let $\{x_1, \ldots, x_k\}$ be a linearly independent set in the inner product space $(V, (\cdot, \cdot))$. Define vectors $y_1, \ldots, y_k$ as follows:

$$y_1 = \frac{x_1}{\|x_1\|}$$

and

$$y_{i+1} = \frac{x_{i+1} - \sum_{j=1}^{i} (x_{i+1}, y_j) y_j}{\|x_{i+1} - \sum_{j=1}^{i} (x_{i+1}, y_j) y_j\|},$$

for $i = 1, \ldots, k - 1$. Then $\{y_1, \ldots, y_k\}$ is an orthonormal set and $\text{span}\{x_1, \ldots, x_i\} = \text{span}\{y_1, \ldots, y_i\}$, $i = 1, \ldots, k$.

Proof. See Halmos (1958, Section 65). □

An immediate consequence of Proposition 1.9 is that if $\{x_1,\ldots, x_n\}$ is a basis for V, then $\{y_1,\ldots, y_n\}$ constructed above is an orthonormal basis for $(V,(\cdot,\cdot))$. If $\{y_1,\ldots, y_n\}$ is an orthonormal basis for $(V,(\cdot,\cdot))$, then each x in V has the representation $x = \Sigma(x, y_i)y_i$ in the given basis. To see this, we know $x = \Sigma\alpha_i y_i$ for unique scalars $\alpha_1,\ldots, \alpha_n$. Thus

$$(x, y_j) = \sum_i \alpha_i (y_i, y_j) = \sum_i \alpha_i \delta_{ij} = \alpha_j.$$

Therefore, the coordinates of x in the orthonormal basis are (x, y_i), $i = 1,\ldots, n$. Also, it follows that $(x, x) = \Sigma(x, y_i)^2$.

Recall that the dual space of V was defined to be the set of all real-valued linear functions on V and was denoted by $\mathcal{L}(V, R)$. Also $\dim(V) = \dim(\mathcal{L}(V, R))$ when V is finite dimensional. The identification of V with $\mathcal{L}(V, R)$ via a given inner product is described in the following proposition.

Proposition 1.10. If $(V,(\cdot,\cdot))$ is a finite dimensional inner product space and if $f \in \mathcal{L}(V, R)$, then there exists a vector $x_0 \in V$ such that $f(x) = (x_0, x)$ for $x \in V$. Conversely, $(x_0, \cdot)$ is a linear function on V for each $x_0 \in V$.

Proof. Let $x_1,\ldots, x_n$ be an orthonormal basis for V and set $\alpha_i = f(x_i)$ for $i = 1,\ldots, n$. For $x_0 = \Sigma\alpha_i x_i$, it is clear that $(x_0, x_j) = \alpha_j = f(x_j)$. Since the two linear functions f and $(x_0, \cdot)$ agree on a basis, they are the same function. Thus $f(x) = (x_0, x)$ for $x \in V$. The converse is clear. □

Definition 1.15. If S is a subset of V, the *orthogonal complement* of S, denoted by $S^\perp$, is $S^\perp = \{x | x \perp y \text{ for all } y \in S\}$.

It is easily verified that $S^\perp$ is a subspace of V for any set S, and $S \perp S^\perp$. The next result provides a basic decomposition for a finite dimensional inner product space.

Proposition 1.11. Suppose M is a k-dimensional subspace of an n-dimensional inner product space $(V,(\cdot,\cdot))$. Then

(i) $M \cap M^\perp = \{0\}$.
(ii) $M \oplus M^\perp = V$.
(iii) $(M^\perp)^\perp = M$.

Proof. Let $\{x_1,\ldots, x_n\}$ be a basis for V such that $\{x_1,\ldots, x_k\}$ is a basis for M. Applying the Gram–Schmidt process to $\{x_1,\ldots, x_n\}$, we get an orthonormal basis $\{y_1, \ldots, y_n\}$ such that $\{y_1, \ldots, y_k\}$ is a basis for M. Let $N = \text{span}\{y_{k+1}, \ldots, y_n\}$. We claim that $N = M^\perp$. It is clear that $N \subseteq M^\perp$ since $y_j \perp M$ for $j = k+1, \ldots, n$. But if $x \in M^\perp$, then $x = \Sigma_1^n (x, y_i) y_i$ and $(x, y_i) = 0$ for $i = 1, \ldots, k$ since $x \in M^\perp$, that is, $x = \Sigma_{k+1}^n (x, y_i) y_i \in N$. Therefore, $M = N^\perp$. Assertions (i) and (ii) now follow easily. For (iii), $M^\perp$ is spanned by $\{y_{k+1}, \ldots, y_n\}$ and, arguing as above, $(M^\perp)^\perp$ must be spanned by $y_1, \ldots, y_k$, which is just M. □

The decomposition, $V = M \oplus M^\perp$, of an inner product space is called an *orthogonal direct sum decomposition*. More generally, if $M_1, \ldots, M_k$ are subspaces of V such that $M_i \perp M_j$ for $i \neq j$ and $V = M_1 \oplus M_2 \oplus \cdots \oplus M_k$, we also speak of the orthogonal direct sum decomposition of V. As we have seen, every direct sum decomposition of a finite dimensional vector space has associated with it two projections. When V is an inner product space and $V = M \oplus M^\perp$, then the projection on M along $M^\perp$ is called the *orthogonal projection* onto M. If P is the orthogonal projection onto M, then $I - P$ is the orthogonal projection onto $M^\perp$. The thing that makes a projection an orthogonal projection is that its null space must be the orthogonal complement of its range. After introducing adjoints of linear transformations, a useful characterization of orthogonal projections is given.

When $(V, (\cdot, \cdot))$ is an inner product space, a number of special types of linear transformations in $\mathcal{L}(V, V)$ arise. First, we discuss the adjoint of a linear transformation. For $A \in \mathcal{L}(V, V)$, consider (x, Ay). For x fixed, (x, Ay) is a linear function of y, and, by Proposition 1.9, there exists a unique vector (which depends on x) $z(x) \in V$ such that $(x, Ay) = (z(x), y)$ for all $y \in V$. Thus z defines a function from V to V that takes x into $z(x)$. However, the verification that $z(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 z(x_1) + \alpha_2 z(x_2)$ is routine. Thus the function z is a linear transformation on V to V, and this leads to the following definition.

Definition 1.16. For $A \in \mathcal{L}(V, V)$, the unique linear transformation in $\mathcal{L}(V, V)$, denoted by A', which satisfies $(x, Ay) = (A'x, y)$, for all $x, y \in V$, is called the *adjoint* (or *transpose*) of A.

The uniqueness of A' in Definition 1.16 follows from the observation that if $(Bx, y) = (Cx, y)$ for all $x, y \in V$, then $((B - C)x, y) = 0$. Taking $y = (B - C)x$ yields $((B - C)x, (B - C)x) = 0$ for all x, so $(B - C)x = 0$ for all x. Hence $B = C$.

Proposition 1.12. If $A, B \in \mathcal{L}(V, V)$, then $(AB)' = B'A'$, and if A is invertible, then $(A^{-1})' = (A')^{-1}$. Also, $(A')' = A$.

Proof. $(AB)'$ is the transformation in $\mathcal{L}(V, V)$ that satisfies $((AB)'x, y) = (x, ABy)$. Using the definition of A' and B', $(x, ABy) = (A'x, By) = (B'A'x, y)$. Thus $(AB)' = B'A'$. The other assertions are proved similarly. □

Definition 1.17. A linear transformation in $\mathcal{L}(V, V)$ is called:

(i) *Self-adjoint* (or symmetric) if $A = A'$.
(ii) *Skew symmetric* if $A' = -A$.
(iii) *Orthogonal* if $(Ax, Ay) = (x, y)$ for $x, y \in V$.

For self-adjoint transformations, A is:

(iv) *Non-negative definite* (or positive semidefinite) if $(x, Ax) \geqslant 0$ for $x \in V$.
(v) *Positive definite* if $(x, Ax) > 0$ for all $x \neq 0$.

The remainder of this section is concerned with a variety of descriptions and characterizations of the classes of transformations defined above.

Proposition 1.13. Let $A \in \mathcal{L}(V, V)$. Then

(i) $\mathcal{R}(A) = (\mathcal{N}(A'))^{\perp}$.
(ii) $\mathcal{R}(A) = \mathcal{R}(AA')$.
(iii) $\mathcal{N}(A) = \mathcal{N}(A'A)$.
(iv) $r(A) = r(A')$

Proof. Assertion (i) is equivalent to $(\mathcal{R}(A))^{\perp} = \mathcal{N}(A')$. But $x \in \mathcal{N}(A')$ means that $0 = (y, A'x)$ for all $y \in V$, and this is equivalent to $x \perp \mathcal{R}(A)$ since $(y, A'x) = (Ay, x)$. This proves (i). For (ii), it is clear that $\mathcal{R}(AA') \subseteq \mathcal{R}(A)$. If $x \in \mathcal{R}(A)$, then $x = Ay$ for some $y \in V$. Write $y = y_1 + y_2$ where $y_1 \in \mathcal{R}(A')$ and $y_2 \in (\mathcal{R}(A'))^{\perp}$. From (i), $(\mathcal{R}(A'))^{\perp} = \mathcal{N}(A)$, so $Ay_2 = 0$. Since $y_1 \in \mathcal{R}(A')$, $y_1 = A'z$ for some $z \in V$. Thus $x = Ay = Ay_1 = AA'z$, so $x \in \mathcal{R}(AA')$.

To prove (iii), if $Ax = 0$, then $A'Ax = 0$, so $\mathcal{N}(A) \subseteq \mathcal{N}(A'A)$. Conversely, if $A'Ax = 0$, then $0 = (x, A'Ax) = (Ax, Ax)$, so $Ax = 0$, and $\mathcal{N}(A'A) \subseteq \mathcal{N}(A)$.

Since $\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = \dim(V)$, $\dim(\mathcal{R}(A')) + \dim(\mathcal{N}(A')) = \dim(V)$, and $\mathcal{R}(A) = (\mathcal{N}(A'))^{\perp}$, it follows that $r(A) = r(A')$. □

If $A \in \mathcal{L}(V, V)$ and $r(A) = 0$, then $A = 0$ since A must map everything into $0 \in V$. We now discuss the rank one linear transformations and show that these can be thought of as the "building blocks" for $\mathcal{L}(V, V)$.

Proposition 1.14. For $A \in \mathcal{L}(V, V)$, the following are equivalent:

(i) $r(A) = 1$.
(ii) There exist $x_0 \neq 0$ and $y_0 \neq 0$ in V such that $Ax = (y_0, x)x_0$ for $x \in V$.

Proof. That (ii) implies (i) is clear since, if $Ax = (y_0, x)x_0$, then $\mathcal{R}(A) = \text{span}\{x_0\}$, which is one-dimensional. Thus suppose $r(A) = 1$. Since $\mathcal{R}(A)$ is one-dimensional, there exists $x_0 \in \mathcal{R}(A)$ with $x_0 \neq 0$ and $\mathcal{R}(A) = \text{span}\{x_0\}$. As $Ax \in \mathcal{R}(A)$ for all x, $Ax = \alpha(x)x_0$ where $\alpha(x)$ is some scalar that depends on x. The linearity of A implies that $\alpha(\beta_1 x_1 + \beta_2 x_2) = \beta_1\alpha(x_1) + \beta_2\alpha(x_2)$. Thus α is a linear function on V and, by Proposition 1.10, $\alpha(x) = (y_0, x)$ for some $y_0 \in V$. Since $\alpha(x) \neq 0$ for some $x \in V$, $y_0 \neq 0$. Therefore, (i) implies (ii). □

This description of the rank one linear transformations leads to the following definition.

Definition 1.18. Given $x, y \in V$, the *outer product* of x and y, denoted by $x \square y$, is the linear transformation on V to V whose value at z is $(x \square y)z = (y, z)x$.

Thus $x \square y \in \mathcal{L}(V, V)$ and $x \square y = 0$ iff x or y is zero. When $x \neq 0$ and $y \neq 0$, $\mathcal{R}(x \square y) = \text{span}\{x\}$ and $\mathcal{N}(x \square y\} = (\text{span}\{y\})^{\perp}$. The result of Proposition 1.14 shows that every rank one transformation is an outer product of two nonzero vectors. The following properties of outer products are easily verified:

(i) $x \square (\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 x \square y_1 + \alpha_2 x \square y_2$.
(ii) $(\alpha_1 x_1 + \alpha_2 x_2) \square y = \alpha_1 x_1 \square y + \alpha_2 x_2 \square y$.
(iii) $(x \square y)' = y \square x$.
(iv) $(x_1 \square y_1)(x_2 \square y_2) = (y_1, x_2) x_1 \square y_2$.

One word of caution: the definition of the outer product depends on the inner product on V. When there is more than one inner product for V, care must be taken to indicate which inner product is being used to define the outer product. The claim that rank one linear transformations are the building blocks for $\mathcal{L}(V, V)$ is partially justified by the following proposition.

Proposition 1.15. Let $\{x_1,\dots, x_n\}$ be an orthonormal basis for $(V,(\cdot,\cdot))$. Then $\{x_i \square x_j;\ i, j = 1,\dots, n\}$ is a basis for $\mathcal{L}(V, V)$.

Proof. If $A \in \mathcal{L}(V, V)$, A is determined by the n^2 numbers $\alpha_{ij} = (x_i, Ax_j)$. But the linear transformation $B = \Sigma\Sigma\alpha_{ij}x_i \square x_j$ satisfies

$$(x_i, Bx_j) = \left(x_i, \left(\sum_k \sum_l \alpha_{kl} x_k \square x_l\right) x_j\right) = \sum_k \sum_l \alpha_{kl}(x_l, x_j)(x_i, x_k) = \alpha_{ij}.$$

Thus $B = A$ so every $A \in \mathcal{L}(V, V)$ is a linear combination of $\{x_i \square x_j | i, j = 1,\dots, n\}$. Since $\dim(\mathcal{L}(V, V)) = n^2$, the result follows. □

Using outer products, it is easy to give examples of self-adjoint linear transformations. First, since linear combinations of self-adjoint linear transformations are again self-adjoint, the set M of self-adjoint transformations is a subspace of $\mathcal{L}(V, V)$. Also, the set N of skew symmetric transformations is a subspace of $\mathcal{L}(V, V)$. It is clear that the only transformation that is both self-adjoint and skew symmetric is 0, so $M \cap N = \{0\}$. But if $A \in \mathcal{L}(V, V)$, then

$$A = \frac{A + A'}{2} + \frac{A - A'}{2}, \qquad \frac{A + A'}{2} \in M, \qquad \text{and} \qquad \frac{A - A'}{2} \in N.$$

This shows that $\mathcal{L}(V, V) = M \oplus N$. To give examples of elements of M, let $x_1,\dots, x_n$ be an orthonormal basis for $(V,(\cdot,\cdot))$. For each i, $x_i \square x_i$ is self-adjoint, so for scalars α_i, $B = \Sigma\alpha_i x_i \square x_i$ is self-adjoint. The geometry associated with the transformation B is interesting and easy to describe. Since $\|x_i\| = 1$, $(x_i \square x_i)^2 = x_i \square x_i$, so $x_i \square x_i$ is a projection on $\text{span}\{x_i\}$ along $(\text{span}\{x_i\})^\perp$—that is, $x_i \square x_i$ is the orthogonal projection on $\text{span}\{x_i\}$ as the null space of $x_i \square x_i$ is the orthogonal complement of its range. Let $M_i = \text{span}\{x_i\}$, $i = 1,\dots, k$. Each M_i is a one-dimensional subspace of $(V,(\cdot,\cdot))$, $M_i \perp M_j$ if $i \neq j$, and $M_1 \oplus M_2 \oplus \cdots \oplus M_n = V$. Hence, V is the direct sum of n mutually orthogonal subspaces and each $x \in V$ has the unique representation $x = \Sigma(x, x_i)x_i$ where $(x, x_i)x_i = (x_i \square x_i)x$ is the projection of x onto M_i, $i = 1,\dots, n$. Since B is linear, the value of Bx is completely determined by the value of B on each M_i, $i = 1,\dots, n$. However, if $y \in M_j$, then $y = \alpha x_j$ for some $\alpha \in R$ and $By = \alpha Bx_j = \alpha\Sigma\alpha_i(x_i \square x_i)x_j = \alpha\alpha_j x_j = \alpha_j y$. Thus when B is restricted to M_j, B is α_j times the identity transformation, and understanding how B transforms vectors has become particularly simple. In summary, take $x \in V$ and write $x = \Sigma(x, x_i)x_i$; then $Bx = \Sigma\alpha_i(x, x_i)x_i$. What is especially fascinating and useful is that every self-adjoint transformation in $\mathcal{L}(V, V)$ has the representation $\Sigma\alpha_i x_i \square x_i$ for some orthonormal basis for V and some scalars $\alpha_1,\dots, \alpha_n$. This fact is known as the spectral theorem and is discussed in more detail later in this chapter. For the time being, we are content with the following observation about the self-adjoint transformation $B = \Sigma\alpha_i x_i \square x_i$: B is positive definite iff $\alpha_i > 0$, $i = 1,\ldots, n$. This follows since $(x, Bx) = \Sigma\alpha_i(x, x_i)^2$ and $x = 0$ iff $(x, x_i)^2 = 0$ for all $i = 1,\ldots, n$. For exactly the same reasons, B is non-negative definite iff $\alpha_i \geqslant 0$ for $i = 1,\ldots, n$. Proposition 1.16 introduces a useful property of self-adjoint transformations.

Proposition 1.16. If A_1 and A_2 are self-adjoint linear transformations in $\mathcal{L}(V, V)$ such that $(x, A_1x) = (x, A_2x)$ for all x, then $A_1 = A_2$.

Proof. It suffices to show that $(x, A_1y) = (x, A_2y)$ for all $x, y \in V$. But

$$\begin{aligned}(x + y, A_1(x + y)) &= (x, A_1x) + (y, A_1y) + 2(x, A_1y)\\ &= (x + y, A_2(x + y))\\ &= (x, A_2x) + (y, A_2y) + 2(x, A_2y).\end{aligned}$$

Since $(z, A_1z) = (z, A_2z)$ for all $z \in V$, we see that $(x, A_1y) = (x, A_2y)$. $\square$

In the above discussion, it has been observed that, if $x \in V$ and $\|x\| = 1$, then $x \square x$ is the orthogonal projection onto the one-dimensional subspace span$\{x\}$. Recall that $P \in \mathcal{L}(V, V)$ is an orthogonal projection if P is a projection (i.e., $P^2 = P$) and if $\mathfrak{N}(P) = (\mathfrak{R}(P))^{\perp}$. The next result characterizes orthogonal projections as those projections that are self-adjoint.

Proposition 1.17. If $P \in \mathcal{L}(V, V)$, the following are equivalent:

(i) P is an orthogonal projection.
(ii) $P^2 = P = P'$.

Proof. If (ii) holds, then P is a projection and P is self-adjoint. By Proposition 1.13, $\mathfrak{N}(P) = (\mathfrak{R}(P'))^{\perp} = (\mathfrak{R}(P))^{\perp}$ since $P = P'$. Thus P is an orthogonal projection. Conversely, if (i) holds, then $P^2 = P$ since P is a projection. We must show that if P is a projection and $\mathfrak{N}(P) = (\mathfrak{R}(P))^{\perp}$, then $P = P'$. Since $V = \mathfrak{R}(P) \oplus \mathfrak{N}(P)$, consider $x, y \in V$ and write $x = x_1 + x_2$, $y = y_1 + y_2$ with $x_1, y_1 \in \mathfrak{R}(P)$ and $x_2, y_2 \in \mathfrak{N}(P) = (\mathfrak{R}(P))^{\perp}$. Using the fact that P is the identity on $\mathfrak{R}(P)$, compute as follows:

$$\begin{aligned}(P'x, y) &= (x, Py) = (x_1 + x_2, Py_1) = (x_1, y_1) = (Px_1, y_1)\\ &= (P(x_1 + x_2), y_1 + y_2) = (Px, y).\end{aligned}$$

Since P' is the unique linear transformation that satisfies $(x, Py) = (P'x, y)$, we have $P = P'$. □

It is sometimes convenient to represent an orthogonal projection in terms of outer products. If P is the orthogonal projection onto M, let $\{x_1, \ldots, x_k\}$ be an orthonormal basis for M in $(V, (\cdot, \cdot))$. Set $A = \Sigma x_i \Box x_i$ so A is self-adjoint. If $x \in M$, then $x = \Sigma(x, x_i)x_i$ and $Ax = (\Sigma x_i \Box x_i)x = \Sigma(x, x_i)x_i = x$. If $x \in M^\perp$, then $Ax = 0$. Since A agrees with P on M and $M^\perp$, $A = P = \Sigma x_i \Box x_i$. Thus all orthogonal projections are sums of rank one orthogonal projections (given by outer products) and different terms in the sum are orthogonal to each other (i.e., $(x_i \Box x_i)(x_j \Box x_j) = 0$ if $i \neq j$). Generalizing this a little bit, two orthogonal projections P_1 and P_2 are called *orthogonal* if $P_1 P_2 = 0$. It is not hard to show that P_1 and P_2 are orthogonal to each other iff the range of P_1 and the range of P_2 are orthogonal to each other, as subspaces. The next result shows that a sum of orthogonal projections is an orthogonal projection iff each pair of summands is orthogonal.

Proposition 1.18. Let $P_1, \ldots, P_k$ be orthogonal projections on $(V, (\cdot, \cdot))$. Then $P = P_1 + \cdots + P_k$ is an orthogonal projection iff $P_i P_j = 0$ for $i \neq j$.

Proof. See Halmos (1958, Section 76). □

We now turn to a discussion of orthogonal linear transformations on an inner product space $(V, (\cdot, \cdot))$. Basically, an orthogonal transformation is one that preserves the geometric structure (distance and angles) of the inner product. A variety of characterizations of orthogonal transformations is possible.

Proposition 1.19. If $(V, (\cdot, \cdot))$ is a finite dimensional inner product space and if $A \in \mathcal{L}(V, V)$, then the following are equivalent:

(i) $(Ax, Ay) = (x, y)$ for all $x, y \in V$.
(ii) $\|Ax\| = \|x\|$ for all $x \in V$.
(iii) $AA' = A'A = I$.
(iv) If $\{x_1, \ldots, x_n\}$ is an orthonormal basis for $(V, (\cdot, \cdot))$, then $\{Ax_1, \ldots, Ax_n\}$ is also an orthonormal basis for $(V, (\cdot, \cdot))$.

Proof. Recall that (i) is our definition of an orthogonal transformation. We prove that (i) implies (ii), (ii) implies (iii), (iii) implies (i), and then show that (i) implies (iv) and (iv) implies (ii). That (i) implies (ii) is clear since

$\|Ax\|^2 = (Ax, Ax)$. For (ii) implies (iii), $(x, x) = (Ax, Ax) = (x, A'Ax)$ implies that $A'A = I$ since $A'A$ and I are self-adjoint (see Proposition 1.16). But, by the uniqueness of inverses, this shows that $A' = A^{-1}$ so $I = AA^{-1} = AA'$ and (iii) holds. Assuming (iii), we have $(x, y) = (x, A'Ay) = (Ax, Ay)$ and (i) holds. If (i) holds and $\{x_1, \dots, x_n\}$ is an orthonormal basis for $(V, (\cdot, \cdot))$, then $\delta_{ij} = (x_i, x_j) = (Ax_i, Ax_j)$, which implies that $\{Ax_1, \dots, Ax_n\}$ is an orthonormal basis. Now, assume (iv) holds. For $x \in V$, we have $x = \Sigma(x, x_i)x_i$ and $\|x\|^2 = \Sigma(x, x_i)^2$. Thus

$$\begin{aligned}\|Ax\|^2 = (Ax, Ax) &= \left(\sum_i (x, x_i)Ax_i, \sum_j (x, x_j)Ax_j\right)\\ &= \sum_i \sum_j (x, x_i)(x, x_j)(Ax_i, Ax_j) = \sum_i \sum_j (x, x_i)(x, x_j)\delta_{ij}\\ &= \sum_i (x, x_i)^2 = \|x\|^2.\end{aligned}$$

Therefore (ii) holds. □

Some immediate consequences of the preceding proposition are: if A is orthogonal, so is $A^{-1} = A'$ and if A_1 and A_2 are orthogonal, then A_1A_2 is orthogonal. Let $\mathcal{O}(V)$ denote all the orthogonal transformations on the inner product space $(V, (\cdot, \cdot))$. Then $\mathcal{O}(V)$ is closed under inverses, $I \in \mathcal{O}(V)$, and $\mathcal{O}(V)$ is closed under products of linear transformations. In other words, $\mathcal{O}(V)$ is a group of linear transformations on $(V, (\cdot, \cdot))$ and $\mathcal{O}(V)$ is called the *orthogonal group* of $(V, (\cdot, \cdot))$. This and many other groups of linear transformations are studied in later chapters.

One characterization of orthogonal transformations on $(V, (\cdot, \cdot))$ is that they map orthonormal bases into orthonormal bases. Thus given two orthonormal bases, there exists a unique orthogonal transformation that maps one basis onto the other. This leads to the following question. Suppose $\{x_1, \dots, x_k\}$ and $\{y_1, \dots, y_k\}$ are two finite sets of vectors in $(V(\cdot, \cdot))$. Under what conditions will there exist an orthogonal transformation A such that $Ax_i = y_i$ for $i = 1, \dots, k$? If such an $A \in \mathcal{O}(V)$ exists, then $(x_i, x_j) = (Ax_i, Ax_j) = (y_i, y_j)$ for all $i, j = 1, \dots, k$. That this condition is also sufficient for the existence of an $A \in \mathcal{O}(V)$ that maps x_i to y_i, $i = 1, \dots, k$, is the content of the next result.

Proposition 1.20. Let $\{x_1, \dots, x_k\}$ and $\{y_1, \dots, y_k\}$ be finite sets in $(V, (\cdot, \cdot))$. The following are equivalent:

(i) $(x_i, x_j) = (y_i, y_j)$ for $i, j = 1, \dots, k$.

(ii) There exists an $A \in \mathcal{O}(V)$ such that $Ax_i = y_i$ for $i = 1, \dots, k$.

Proof. That (ii) implies (i) is clear, so assume that (i) holds. Let $M = \text{span}\{x_1, \dots, x_k\}$. The idea of the proof is to define A on M using linearity and then extend the definition of A to V using linearity again. Of course, it must be verified that all this makes sense and that the A so defined is orthogonal. The details of this, which are primarily computational, follow. First, by (i), $\Sigma\alpha_i x_i = 0$ iff $\Sigma\alpha_i y_i = 0$ since $(\Sigma\alpha_i x_i, \Sigma\alpha_j x_j) = \Sigma\Sigma\alpha_i\alpha_j(x_i, x_j) = \Sigma\Sigma\alpha_i\alpha_j(y_i, y_j) = (\Sigma\alpha_i y_i, \Sigma\alpha_j y_j)$. Let $N = \text{span}\{y_1, \dots, y_k\}$ and define B on M to N by $B(\Sigma\alpha_i x_i) \equiv \Sigma\alpha_i y_i$. B is well defined since $\Sigma\alpha_i x_i = \Sigma\beta_i x_i$ implies that $\Sigma\alpha_i y_i = \Sigma\beta_i y_i$ and the linearity of B on M is easy to check. Since B maps M onto N, $\dim(N) \leqslant \dim(M)$. But if $B(\Sigma\alpha_i x_i) = 0$, then $\Sigma\alpha_i y_i = 0$, so $\Sigma\alpha_i x_i = 0$. Therefore the null space of B is $\{0\} \subseteq M$ and $\dim(M) = \dim(N)$. Let $M^\perp$ and $N^\perp$ be the orthogonal complements of M and N, respectively, and let $\{u_1, \dots, u_s\}$ and $\{v_1, \dots, v_s\}$ be orthonormal bases for $M^\perp$ and $N^\perp$, respectively. Extend the definition of B to V by first defining $B(u_i) = v_i$ for $i = 1, \dots, s$ and then extend by linearity. Let A be the linear transformation so defined. We now claim that $\|Aw\|^2 = \|w\|^2$ for all $w \in V$. To see this write $w = w_1 + w_2$ where $w_1 \in M$ and $w_2 \in M^\perp$. Then $Aw_1 \in N$ and $Aw_2 \in N^\perp$. Thus $\|Aw\|^2 = \|Aw_1 + Aw_2\|^2 = \|Aw_1\|^2 + \|Aw_2\|^2$. But $w_1 = \Sigma\alpha_i x_i$ for some scalars α_i. Thus

$$\begin{aligned}\|Aw_1\|^2 &= \left(A\left(\sum\alpha_i x_i\right), A\left(\sum\alpha_j x_j\right)\right) = \sum\sum\alpha_i\alpha_j(Ax_i, Ax_j)\\ &= \sum\sum\alpha_i\alpha_j(y_i, y_j) = \sum\sum\alpha_i\alpha_j(x_i, x_j)\\ &= \left(\sum\alpha_i x_i, \sum\alpha_j x_j\right) = \|w_1\|^2.\end{aligned}$$

Similarly, $\|Aw_2\|^2 = \|w_2\|^2$. Since $\|w\|^2 = \|w_1\|^2 + \|w_2\|^2$, the claim that $\|Aw\|^2 = \|w\|^2$ is established. By Proposition 1.19, A is orthogonal. □

◆ **Example 1.6.** Consider the real vector space R^n with the standard basis and the usual inner product. Also, let $\mathcal{L}_{n,n}$ be the real vector space of all $n \times n$ real matrices. Thus each element of $\mathcal{L}_{n,n}$ determines a linear transformation on R^n and vice versa. More precisely, if A is a linear transformation on R^n to R^n and $[A]$ denotes the matrix of A in the standard basis on both the range and domain of A, then $[Ax] = [A]x$ for $x \in R^n$. Here, $[Ax] \in R^n$ is the vector of coordinates of Ax in the standard basis and $[A]x$ means the matrix $[A] = \{a_{ij}\}$ times the coordinate vector $x \in R^n$. Conversely, if $[A] \in \mathcal{L}_{n,n}$ and we define a linear transformation A by $Ax = [A]x$, then the matrix of A is $[A]$. It is easy to show that if A is a linear transformation on R^n to R^n with the standard inner product, then $[A'] = [A]'$ where A' denotes the adjoint of A and $[A]'$ denotes the transpose of the matrix $[A]$. Now, we are in a position to relate the notions of self-adjointness and skew symmetry of linear transformations to properties of matrices. Proofs of the following two assertions are straightforward and are left to the reader. Let A be a linear transformation on R^n to R^n with matrix $[A]$.

(i) A is self-adjoint iff $[A] = [A]'$.
(ii) A is skew-symmetric iff $[A]' = -[A]$.

Elements of $\mathcal{L}_{n,n}$ that satisfy $B = B'$ are usually called *symmetric* matrices, while the term *skew-symmetric* is used if $B' = -B$, $B \in \mathcal{L}_{n,n}$. Also, the matrix B is called *positive definite* if $x'Bx > 0$ for all $x \in R^n$, $x \neq 0$. Of course $x'Bx$ is just the standard inner product of x with Bx. Clearly, B is positive definite iff the linear transformation it defines is positive definite.

If A is an orthogonal transformation on R^n to R^n, then $[A]$ must satisfy $[A][A]' = [A]'[A] = I_n$ where I_n is the $n \times n$ identity matrix. Thus a matrix $B \in \mathcal{L}_{n,n}$ is called *orthogonal* if $BB' = B'B = I_n$. An interesting geometric interpretation of the condition $BB' = B'B = I_n$ follows. If $B = \{b_{ij}\}$, the vectors $b_j \in R^n$ with coordinates b_{ij}, $i = 1, \ldots, n$, are the *column vectors* of B and the vectors $c_i \in R^n$ with coordinates b_{ij}, $j = 1, \ldots, n$, are the *row vectors* of B. The matrix BB' has elements $c_i'c_j$ and the condition $BB' = I_n$ means that $c_i'c_j = \delta_{ij}$—that is, the vectors $c_1, \ldots, c_n$ form an orthonormal basis for R^n in the usual inner product. Similarly, the condition $B'B = I_n$ holds iff the vectors $b_1, \ldots, b_n$ form an orthonormal basis for R^n. Hence a matrix B is orthogonal iff both its rows and columns determine an orthonormal basis for R^n with the standard inner product. ◆

1.4. THE CAUCHY–SCHWARZ INEQUALITY

The form of the Cauchy–Schwarz Inequality given here is general enough to be applicable to both finite and infinite dimensional vector spaces. The examples below illustrate that the generality is needed to treat some standard situations that arise in analysis and in the study of random variables. In a finite dimensional inner product space $(V, (\cdot, \cdot))$, the inequality established in this section shows that $|(x, y)| \leqslant \|x\|\|y\|$ where $\|x\|^2 = (x, x)$. Thus $-1 \leqslant (x, y)/\|x\|\|y\| \leqslant 1$ and the quantity $(x, y)/\|x\|\|y\|$ is defined to be the cosine of the angle between the vectors x and y. A variety of applications of the Cauchy–Schwarz Inequality arise in later chapters. We now proceed with the technical discussion.

Suppose that V is a real vector space, not necessarily finite dimensional. Let $[\cdot, \cdot]$ denote a non-negative definite symmetric bilinear function on $V \times V$—that is, $[\cdot, \cdot]$ is a real-valued function on $V \times V$ that satisfies (i) $[x, y] = [y, x]$, (ii) $[\alpha_1 x_1 + \alpha_2 x_2, y] = \alpha_1[x_1, y] + \alpha_2[x_2, y]$, and (iii) $[x, x] \geqslant 0$. It is clear that (i) and (ii) imply that $[x, \alpha_1 y_1 + \alpha_2 y_2] = \alpha_1[x, y_1] + \alpha_2[x, y_2]$. The Cauchy–Schwarz Inequality states that $[x, y]^2 \leqslant [x, x][y, y]$. We also give necessary and sufficient conditions for equality to hold in this inequality. First, a preliminary result.

Proposition 1.21. Let $M = \{x | [x, x] = 0\}$. Then M is a subspace of V.

Proof. If $x \in M$ and $\alpha \in R$, then $[\alpha x, \alpha x] = \alpha^2[x, x] = 0$ so $\alpha x \in M$. Thus we must show that if $x_1, x_2 \in M$, then $x_1 + x_2 \in M$. For $\alpha \in R$, $0 \leqslant [x_1 + \alpha x_2, x_1 + \alpha x_2] = [x_1, x_1] + 2\alpha[x_1, x_2] + \alpha^2[x_2, x_2] = 2\alpha[x_1, x_2]$ since $x_1, x_2 \in M$. But if $2\alpha[x_1, x_2] \geqslant 0$ for all $\alpha \in R$, $[x_1, x_2] = 0$, and this implies that $0 = [x_1 + \alpha x_2, x_1 + \alpha x_2]$ for all $\alpha \in R$ by the above equality. Therefore, $x_1 + \alpha x_2 \in M$ for all α when $x_1, x_2 \in M$ and thus M is a subspace. □

Theorem 1.1. (Cauchy–Schwarz Inequality). Let $[\cdot, \cdot]$ be a non-negative definite symmetric bilinear function on $V \times V$ and set $M = \{x | [x, x] = 0\}$. Then:

(i) $[x, y]^2 \leqslant [x, x][y, y]$ for $x, y \in V$.
(ii) $[x, y]^2 = [x, x][y, y]$ iff $\alpha x + \beta y \in M$ for some real α and β not both zero.

Proof. To prove (i), we consider two cases. If $x \in M$, then $0 \leqslant [y + \alpha x, y + \alpha x] = [y, y] + 2\alpha[x, y]$ for all $\alpha \in R$, so $[x, y] = 0$ and (i) holds. Similarly, if $y \in M$, (i) holds. If $x \notin M$ and $y \notin M$, let $x_1 = x/[x, x]^{1/2}$ and let $y_1 = y/[y, y]^{1/2}$. Then we must show that $|[x_1, y_1]| \leqslant 1$. This follows from the two inequalities

$$0 \leqslant [x_1 - y_1, x_1 - y_1] = 2 - 2[x_1, y_1]$$

and

$$0 \leqslant [x_1 + y_1, x_1 + y_1] = 2 + 2[x_1, y_1].$$

The proof of (i) is now complete.

To prove (ii), first assume that $[x, y]^2 = [x, x][y, y]$. If either $x \in M$ or $y \in M$, then $\alpha x + \beta y \in M$ for some α, β not both zero. Thus consider $x \notin M$ and $y \notin M$. An examination of the proof of (i) shows that we can have equality in (i) iff either $0 = [x_1 - y_1, x_1 - y_1]$ or $0 = [x_1 + y_1, x_1 + y_1]$ and, in either case, this implies that $\alpha x + \beta y \in M$ for some real α, β not both zero. Now, assume $\alpha x + \beta y \in M$ for some real α, β not both zero. If $\alpha = 0$ or $\beta = 0$ or $x \in M$ or $y \in M$, we clearly have equality in (i). For the case when $\alpha\beta \neq 0$, $x \notin M$, and $y \notin M$, our assumption implies that $x_1 + \gamma y_1 \in M$ for some $\gamma \neq 0$, since M is a subspace. Thus there is a real $\gamma \neq 0$ such that $0 = [x_1 + \gamma y_1, x_1 + \gamma y_1] = 1 + 2\gamma[x_1, y_1] + \gamma^2$. The equation for the roots of a quadratic shows that this can hold only if $|[x_1, y_1]| = 1$. Hence equality in (i) holds. □

◆ **Example 1.7.** Let $(V, (\cdot, \cdot))$ be a finite dimensional inner product space and suppose A is a non-negative definite linear transformation on V to V. Then $[x, y] \equiv (x, Ay)$ is a non-negative definite symmetric bilinear function. The set $M = \{x | (x, Ax) = 0\}$ is equal to $\mathfrak{N}(A)$—this follows easily from Theorem 1.1(i). Theorem 1.1 shows that $(x, Ay)^2 \leqslant (x, Ax)(y, Ay)$ and provides conditions for equality. In particular, when A is nonsingular, $M = \{0\}$ and equality holds iff x and y are linearly dependent. Of course, if $A = I$, then we have $(x, y)^2 \leqslant \|x\|^2\|y\|^2$, which is one classical form of the Cauchy–Schwarz Inequality. ◆

◆ **Example 1.8.** In this example, take V to be the set of all continuous real-valued functions defined on a closed bounded interval, say a to b, of the real line. It is easily verified that

$$[x_1, x_2] \equiv \int_a^b x_1(t) x_2(t)\, dt$$

is symmetric, bilinear, and non-negative definite. Also $[x, x] > 0$ unless $x = 0$ since x is continuous. Hence $M = \{0\}$. The Cauchy–Schwarz Inequality yields

$$\left(\int_a^b x_1(t) x_2(t)\, dt\right)^2 \leqslant \int_a^b x_1^2(t)\, dt \int_a^b x_2^2(t)\, dt.$$ ◆

◆ **Example 1.9.** The following example has its origins in the study of the covariance between two real-valued random variables. Consider a probability space $(\Omega, \mathcal{F}, P_0)$ where Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P_0 is a probability measure on $\mathcal{F}$. A *random variable* X is a real-valued function defined on Ω such that the inverse image of each Borel set in R is an element of $\mathcal{F}$; symbolically, $X^{-1}(B) \in \mathcal{F}$ for each Borel set B of R. Sums and products of random variables are random variables and the constant functions on Ω are random variables. If X is a random variable such that $\int |X(\omega)| P_0(d\omega) < +\infty$, then X is *integrable* and we write $\mathcal{E}X$ for $\int X(\omega) P_0(d\omega)$.

Now, let V be the collection of all real-valued random variables X, such that $\mathcal{E}X^2 < +\infty$. It is clear that if $X \in V$, then $\alpha X \in V$ for all real α. Since $(X_1 + X_2)^2 \leqslant 2(X_1^2 + X_2^2)$, if X_1 and X_2 are in V, then $X_1 + X_2$ is in V. Thus V is a real vector space with addition being the pointwise addition of random variables and scalar multiplication being pointwise multiplication of random variables by scalars. For $X_1, X_2 \in V$, the inequality $|X_1 X_2| \leqslant X_1^2 + X_2^2$ implies that $X_1 X_2$ is integrable. In particular, setting $X_2 = 1$, X_1 is integrable. Define $[\cdot, \cdot]$ on $V \times V$ by $[X_1, X_2] = \mathcal{E}(X_1 X_2)$. That $[\cdot, \cdot]$ is symmetric and bilinear is clear. Since $[X_1, X_1] = \mathcal{E}X_1^2 \geqslant 0$, $[\cdot, \cdot]$ is non-negative definite. The Cauchy–Schwarz Inequality yields $(\mathcal{E}X_1 X_2)^2 \leqslant \mathcal{E}X_1^2 \mathcal{E}X_2^2$, and setting $X_2 = 1$, this gives $(\mathcal{E}X_1)^2 \leqslant \mathcal{E}X_1^2$. Of course, this is just a verification that the variance of a random variable is non-negative. For future use, let $\text{var}(X_1) \equiv \mathcal{E}X_1^2 - (\mathcal{E}X_1)^2$. To discuss conditions for equality in the Cauchy–Schwarz Inequality, the subspace $M = \{X | [X, X] = 0\}$ needs to be described. Since $[X, X] = \mathcal{E}X^2$, $X \in M$ iff X is zero, except on set of P_0 measure zero—that is, $X = 0$ a.e. (P_0). Therefore, $(\mathcal{E}X_1 X_2)^2 = \mathcal{E}X_1^2 \mathcal{E}X_2^2$ iff $\alpha X_1 + \beta X_2 = 0$ a.e. (P_0) for some real α, β not both zero. In particular, $\text{var}(X_1) = 0$ iff $X_1 - \mathcal{E}X_1 = 0$ a.e. (P_0).

A somewhat more interesting non-negative definite symmetric bilinear function on $V \times V$ is

$$\text{cov}\{X_1, X_2\} \equiv \mathcal{E}X_1 X_2 - \mathcal{E}X_1 \mathcal{E}X_2,$$

and is called the *covariance* between X_1 and X_2. Symmetry is clear and bilinearity is easily checked. Since $\text{cov}\{X_1, X_1\} = \mathcal{E}X_1^2 - (\mathcal{E}X_1)^2 = \text{var}(X_1)$, $\text{cov}\{\cdot, \cdot\}$ is non-negative definite and $M_1 = \{X | \text{cov}\{X, X\} = 0\}$ is just the set of random variables in V that have variance zero. For this case, the Cauchy–Schwarz Inequality is

$$(\text{cov}\{X_1, X_2\})^2 \leqslant \text{var}(X_1)\text{var}(X_2).$$

Equality holds iff there exist α, β, not both zero, such that $\text{var}(\alpha X_1 + \beta X_2) = 0$; or equivalently, $\alpha(X_1 - \mathcal{E}X_1) + \beta(X_2 - \mathcal{E}X_2) = 0$ a.e. (P_0) for some α, β not both zero. The properties of $\text{cov}\{\cdot, \cdot\}$ given here are used in the next chapter to define the covariance of a random vector. ◆

1.5. THE SPACE $\mathcal{L}(V, W)$

When $(V, (\cdot, \cdot))$ is an inner product space, the adjoint of a linear transformation in $\mathcal{L}(V, V)$ was introduced in Section 1.3 and used to define some special linear transformations in $\mathcal{L}(V, V)$. Here, some of the notions discussed in relation to $\mathcal{L}(V, V)$ are extended to the case of linear transformations in $\mathcal{L}(V, W)$ where $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$ are two inner product spaces. In particular, adjoints and outer products are defined, bilinear functions on $V \times W$ are characterized, and Kronecker products are introduced. Of course, all the results in this section apply to $\mathcal{L}(V, V)$ by taking $(W, [\cdot, \cdot]) = (V, (\cdot, \cdot))$ and the reader should take particular notice of this special case. There is one point that needs some clarification. Given $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$, the adjoint of $A \in \mathcal{L}(V, W)$, to be defined below, depends on both the inner products $(\cdot, \cdot)$ and $[\cdot, \cdot]$. However, in the previous discussion of adjoints in $\mathcal{L}(V, V)$, it was assumed that the inner product was the same on both the range and the domain of the linear transformation (i.e., V is the domain and range). Whenever we discuss adjoints of $A \in \mathcal{L}(V, V)$ it is assumed that only one inner product is involved, unless the contrary is explicitly stated—that is, when specializing results from $\mathcal{L}(V, W)$ to $\mathcal{L}(V, V)$, we take $W = V$ and $[\cdot, \cdot] = (\cdot, \cdot)$.

The first order of business is to define the adjoint of $A \in \mathcal{L}(V, W)$ where $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$ are inner product spaces. For a fixed $w \in W$, $[w, Ax]$ is a linear function of $x \in V$ and, by Proposition 1.10, there exists a unique vector $y(w) \in V$ such that $[w, Ax] = (y(w), x)$ for all $x \in V$. It is easy to verify that $y(\alpha_1 w_1 + \alpha_2 w_2) = \alpha_1 y(w_1) + \alpha_2 y(w_2)$. Hence $y(\cdot)$ determines a linear transformation on W to V, say A', which satisfies $[w, Ax] = (A'w, x)$ for all $w \in W$ and $x \in V$.

Definition 1.19. Given inner product spaces $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$, if $A \in \mathcal{L}(V, W)$, the unique linear transformation $A' \in \mathcal{L}(W, V)$ that satisfies $[w, Ax] = (A'w, x)$ for all $w \in W$ and $x \in V$ is called the *adjoint* of A.

The existence and uniqueness of A' was demonstrated in the discussion preceeding Definition 1.19. It is not hard to show that $(A+B)' = A' + B'$, $(A')' = A$, and $(\alpha A)' = \alpha A'$. In the present context, Proposition 1.13 becomes Proposition 1.22.

Proposition 1.22. Suppose $A \in \mathcal{L}(V, W)$. Then:

(i) $\mathcal{R}(A) = (\mathfrak{N}(A'))^{\perp}$.
(ii) $\mathcal{R}(A) = \mathcal{R}(AA')$.
(iii) $\mathfrak{N}(A) = \mathfrak{N}(A'A)$.
(iv) $r(A) = r(A')$.

Proof. The proof here is essentially the same as that given for Proposition 1.13 and is left to the reader. □

The notion of an outer product has a natural extension to $\mathcal{L}(V, W)$.

Definition 1.20. For $x \in (V, (\cdot, \cdot))$ and $w \in (W, [\cdot, \cdot])$, the *outer product*, $w \square x$ is that linear transformation in $\mathcal{L}(V, W)$ given by $(w \square x)(y) \equiv (x, y)w$ for all $y \in V$.

If $w = 0$ or $x = 0$, then $w \square x = 0$. When both w and x are not zero, then $w \square x$ has rank one, $\mathcal{R}(w \square x) = \text{span}\{w\}$, and $\mathfrak{N}(w \square x) = (\text{span}\{x\})^{\perp}$. Also, a minor modification of the proof of Proposition 1.14 shows that, if $A \in \mathcal{L}(V, W)$, then $r(A) = 1$ iff $A = w \square x$ for some nonzero w and x.

Proposition 1.23. The outer product has the following properties:

(i) $(\alpha_1 w_1 + \alpha_2 w_2) \square x = \alpha_1 w_1 \square x + \alpha_2 w_2 \square x$.
(ii) $w \square (\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 w \square x_1 + \alpha_2 w \square x_2$.
(iii) $(w \square x)' = x \square w \in \mathcal{L}(W, V)$.

If $(V_1, (\cdot, \cdot)_1)$, $(V_2, (\cdot, \cdot)_2)$, and $(V_3, (\cdot, \cdot)_3)$ are inner product spaces with $x_1 \in V_1$, $x_2, y_2 \in V_2$, and $y_3 \in V_3$, then

(iv) $(y_3 \square y_2)(x_2 \square x_1) = (x_2, y_2)_2 y_3 \square x_1 \in \mathcal{L}(V_1, V_3)$.

Proof. Assertions (i), (ii), and (iii) follow easily. For (iv), consider $x \in V_1$. Then $(x_2 \square x_1)x = (x_1, x)_1 x_2$, so $(y_3 \square y_2)(x_2 \square x_1)x = (x_1, x)_1 (y_3 \square y_2)x_2 = (x_1, x)_1 (y_2, x_2)_2 y_3 \in V_3$. However, $(x_2, y_2)_2 (y_3 \square x_1)x = (x_2, y_2)_2 (x_1, x)_1 y_3$. Thus (iv) holds. □

There is a natural way to construct an inner product on $\mathcal{L}(V, W)$ from inner products on V and W. This construction and its relation to outer products are described in the next proposition.

Proposition 1.24. Let $\{x_1, \ldots, x_m\}$ be an orthonormal basis for $(V, (\cdot, \cdot))$ and let $\{w_1, \ldots, w_n\}$ be an orthonormal basis for $(W, [\cdot, \cdot])$. Then:

(i) $\{w_i \square x_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is a basis for $\mathcal{L}(V, W)$.

Let $a_{ij} = [w_i, Ax_j]$. Then:

(ii) $A = \Sigma\Sigma a_{ij} w_i \square x_j$ and the matrix of A is $[A] = \{a_{ij}\}$ in the given bases.

If $A = \Sigma\Sigma a_{ij} w_i \square x_j$ and $B = \Sigma\Sigma b_{ij} w_i \square x_j$, define $\langle A, B\rangle \equiv \Sigma\Sigma a_{ij} b_{ij}$. Then:

(iii) $\langle \cdot, \cdot \rangle$ is an inner product on $\mathcal{L}(V, W)$ and $\{w_i \square x_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is an orthonormal basis for $(\mathcal{L}(V, W), \langle \cdot, \cdot \rangle)$.

Proof. Since $\dim(\mathcal{L}(V, W)) = mn$, to prove (i) it suffices to prove (ii). Let $B = \Sigma\Sigma a_{ij} w_i \square x_j$. Then

$$[w_k, Bx_l] = \sum_i \sum_j a_{ij} \left[w_k, (w_i \square x_j) x_l\right] = \sum_i \sum_j a_{ij} \delta_{ik} \delta_{jl} = a_{kl},$$

so $[w_i, Bx_j] = [w_i, Ax_j]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Therefore, $[w, Bx] = [w, Ax]$ for all $w \in W$ and $x \in V$, which implies that $[w, (B - A)x] = 0$. Choosing $w = (B - A)x$, we see that $(B - A)x = 0$ for all $x \in V$ and, therefore, $B = A$. To show that the matrix of A is $[A] = \{a_{ij}\}$, recall that the matrix of A consists of the scalars b_{kj} defined by $Ax_j = \Sigma_k b_{kj} w_k$. The inner product of w_i with each side of this equation is

$$a_{ij} = \left[w_i, Ax_j\right] = \sum_k b_{kj} [w_i, w_k] = b_{ij}$$

and the proof of (ii) is complete.

For (iii), $\langle \cdot, \cdot \rangle$ is clearly symmetric and bilinear. Since $\langle A, A\rangle = \Sigma\Sigma a_{ij}^2$, the positivity of $\langle \cdot, \cdot \rangle$ follows. That $\{w_i \square x_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is an orthonormal basis for $(\mathcal{L}(V, W), \langle \cdot, \cdot \rangle)$ follows immediately from the definition of $\langle \cdot, \cdot \rangle$. □

A few words are in order concerning the inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{L}(V, W)$. Since $\{w_i \square x_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is an orthonormal basis, we know that if $A \in \mathcal{L}(V, W)$, then

$$A = \sum\sum \langle A, w_i \square x_j \rangle w_i \square x_j,$$

since this is the unique expansion of a vector in any orthonormal basis. However, $A = \Sigma\Sigma[w_i, Ax_j]w_i \square x_j$ by (ii) of Proposition 1.24. Thus $\langle A, w_i \square x_j \rangle = [w_i, Ax_j]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Since both sides of this relation are linear in w_i and x_j, we have $\langle A, w \square x \rangle = [w, Ax]$ for all $w \in W$ and $x \in V$. In particular, if $A = \tilde{w} \square \tilde{x}$, then

$$\langle \tilde{w} \square \tilde{x}, w \square x \rangle = \left[w, (\tilde{w} \square \tilde{x})x\right] = \left[w, (\tilde{x}, x)\tilde{w}\right] = [w, \tilde{w}](x, \tilde{x}).$$

This relation has some interesting implications.

Proposition 1.25. The inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{L}(V, W)$ satisfies

(i) $\langle \tilde{w} \square \tilde{x}, w \square x \rangle = [\tilde{w}, w](\tilde{x}, x)$

for all $\tilde{w}, w \in W$ and $\tilde{x}, x \in V$, and $\langle \cdot, \cdot \rangle$ is the unique inner product with this property. Further, if $\{z_1, \ldots, z_n\}$ and $\{y_1, \ldots, y_m\}$ are any orthonormal bases for W and V, respectively, then $\{z_i \square y_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is an orthonormal basis for $(\mathcal{L}(V, W), \langle \cdot, \cdot \rangle)$.

Proof. Equation (i) has been verified. If $\{\cdot, \cdot\}$ is another inner product on $\mathcal{L}(V, W)$ that satisfies (i), then

$$\{w_{i_1} \square x_{j_1}, w_{i_2} \square x_{j_2}\} = \langle w_{i_1} \square x_{j_1}, w_{i_2} \square x_{j_2} \rangle$$

for all $i_1, i_2 = 1, \ldots, n$ and $j_1, j_2 = 1, \ldots, m$ where $\{x_1, \ldots, x_m\}$ and $\{w_1, \ldots, w_n\}$ are the orthonormal bases used to define $\langle \cdot, \cdot \rangle$. Using (i) of Proposition 1.24 and the bilinearity of inner products, this implies that $\{A, B\} = \langle A, B \rangle$ for all $A, B \in \mathcal{L}(V, W)$. Therefore, the two inner products are the same. The verification that $\{z_i \square y_j | i = 1, \ldots, n,\ j = 1, \ldots, m\}$ is an orthonormal basis follows easily from (i). □

The result of Proposition 1.25 is a formal statement of the fact that $\langle \cdot, \cdot \rangle$ does not depend on the particular orthonormal bases used to define it, but $\langle \cdot, \cdot \rangle$ is determined by the inner products on V and W. Whenever V and W are inner product spaces, the symbol $\langle \cdot, \cdot \rangle$ always means the inner product on $\mathcal{L}(V, W)$ as defined above.

◆ **Example 1.10.** Consider $V = R^m$ and $W = R^n$ with the usual inner products and the standard bases. Thus we have the inner product $\langle \cdot , \cdot \rangle$ on $\mathcal{L}_{m,n}$—the linear space of $n \times m$ real matrices. For $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ in $\mathcal{L}_{m,n}$,

$$\langle A, B \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij}.$$

If $C = AB' : n \times n$, then

$$c_{ii} = \sum_j a_{ij} b_{ij}, \qquad i = 1, \dots, n,$$

so $\langle A, B \rangle = \Sigma c_{ii}$. In other words, $\langle A, B \rangle$ is just the sum of the diagonal elements of the $n \times n$ matrix AB'. This observation leads to the definition of the trace of any square matrix. If $C: k \times k$ is a real matrix, the *trace of* C, denoted by $\operatorname{tr} C$, is the sum of the diagonal elements of C. The identity $\langle A, B \rangle = \langle B, A \rangle$ shows that $\operatorname{tr} AB' = \operatorname{tr} B'A$ for all $A, B \in \mathcal{L}_{m,n}$. In the present example, it is clear that $w \square x = wx'$ for $x \in R^m$ and $w \in R^n$, so $w \square x$ is just the $n \times 1$ matrix w times the $1 \times m$ matrix x'. Also, the identity in Proposition 1.25 is a reflection of the fact that

$$\operatorname{tr} \tilde{w}\tilde{x}'xw' = \tilde{w}'w\tilde{x}'x$$

for $w, \tilde{w} \in R^n$ and $x, \tilde{x} \in R^m$. ◆

If $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$ are inner product spaces and $A \in \mathcal{L}(V, W)$, then $[Ax, w]$ is linear in x for fixed w and linear in w for fixed x. This observation leads to the following definition.

Definition 1.21. A function f defined on $V \times W$ to R is called *bilinear* if:

(i) $f(\alpha_1 x_1 + \alpha_2 x_2, w) = \alpha_1 f(x_1, w) + \alpha_2 f(x_2, w)$.
(ii) $f(x, \alpha_1 w_1 + \alpha_2 w_2) = \alpha_1 f(x, w_1) + \alpha_2 f(x, w_2)$.

These conditions apply for scalars α_1 and α_2; $x, x_1, x_2 \in V$ and $w, w_1, w_2 \in W$.

Our next result shows there is a natural one-to-one correspondence between bilinear functions and $\mathcal{L}(V, W)$.

Proposition 1.26. If f is a bilinear function on $V \times W$ to R, then there exists an $A \in \mathcal{L}(V, W)$ such that $f(x, w) = [Ax, w]$ for all $x \in V$ and $w \in W$. Conversely, each $A \in \mathcal{L}(V, W)$ determines the bilinear function $[Ax, w]$ on $V \times W$.

Proof. Let $\{x_1, \dots, x_m\}$ be an orthonormal basis for $(V, (\cdot, \cdot))$ and $\{w_1, \dots, w_n\}$ be an orthonormal basis for $(W, [\cdot, \cdot])$. Set $a_{ij} = f(x_j, w_i)$ for $i = 1, \dots, n$ and $j = 1, \dots, m$ and let $A = \Sigma\Sigma a_{ij} w_i \square x_j$. By Proposition 1.24, we have

$$a_{ij} = \left[Ax_j, w_i\right] = f(x_j, w_i).$$

The bilinearity of f and of $[Ax, w]$ implies $[Ax, w] = f(x, w)$ for all $x \in V$ and $w \in W$. The converse is obvious. □

Thus far, we have seen that $\mathcal{L}(V, W)$ is a real vector space and that, if V and W have inner products $(\cdot, \cdot)$ and $[\cdot, \cdot]$, respectively, then $\mathcal{L}(V, W)$ has a natural inner product determined by $(\cdot, \cdot)$ and $[\cdot, \cdot]$. Since $\mathcal{L}(V, W)$ is a vector space, there are linear transformations on $\mathcal{L}(V, W)$ to other vector spaces and there is not much more to say in general. However, $\mathcal{L}(V, W)$ is built from outer products and it is natural to ask if there are special linear transformations on $\mathcal{L}(V, W)$ that transform outer products into outer products. For example, if $A \in \mathcal{L}(V, V)$ and $B \in \mathcal{L}(W, W)$, suppose we define $B \otimes A$ on $\mathcal{L}(V, W)$ by $(B \otimes A)C = BCA'$ where A' denotes the transpose of $A \in \mathcal{L}(V, V)$. Clearly, $B \otimes A$ is a linear transformation. If $C = w \square x$, then $(B \otimes A)(w \square x) = B(w \square x)A' \in \mathcal{L}(V, W)$. But for $v \in V$,

$$\begin{aligned}(B(w \square x)A')v &= B(w \square x)(A'v) = B((x, A'v)w)\\ &= (Ax, v)Bw = ((Bw) \square (Ax))v.\end{aligned}$$

This calculation shows that $(B \otimes A)(w \square x) = (Bw) \square (Ax)$, so outer products get mapped into outer products by $B \otimes A$. Generalizing this a bit, we have the following definition.

Definition 1.22. Let $(V_1, (\cdot, \cdot)_1)$, $(V_2, (\cdot, \cdot)_2)$, $(W_1, [\cdot, \cdot]_1)$, and $(W_2, [\cdot, \cdot]_2)$ be inner product spaces. For $A \in \mathcal{L}(V_1, V_2)$ and $B \in \mathcal{L}(W_1, W_2)$, the *Kronecker product* of B and A, denoted by $B \otimes A$, is a linear transformation on $\mathcal{L}(V_1, W_1)$ to $\mathcal{L}(V_2, W_2)$, defined by

$$(B \otimes A)C \equiv BCA'$$

for all $C \in \mathcal{L}(V_1, W_1)$.

In most applications of Kronecker products, $V_1 = V_2$ and $W_1 = W_2$, so $B \otimes A$ is a linear transformation on $\mathcal{L}(V_1, W_1)$ to $\mathcal{L}(V_1, W_1)$. It is not easy to say in a few words why the transpose of A should appear in the definition of the Kronecker product, but the result below should convince the reader that the definition is the "right" one. Of course, by A', we mean the linear transformation on V_2 to V_1, which satisfies $(x_2, Ax_1)_2 = (A'x_2, x_1)_1$ for $x_1 \in V_1$ and $x_2 \in V_2$.

Proposition 1.27. In the notation of Definition 1.22,

(i) $(B \otimes A)(w_1 \square v_1) = (Bw_1)\square(Av_1) \in \mathcal{L}(V_2, W_2)$.

Also,

(ii) $(B \otimes A)' = B' \otimes A'$,

where $(B \otimes A)'$ denotes the transpose of the linear transformation $B \otimes A$ on $(\mathcal{L}(V_1, W_1), \langle \cdot, \cdot \rangle_1)$ to $(\mathcal{L}(V_2, W_2), \langle \cdot, \cdot \rangle_2)$.

Proof. To verify (i), for $v_2 \in V_2$, compute as follows:

$$[(B \otimes A)(w_1 \square v_1)](v_2) = B(w_1 \square v_1)A'v_2 = B(v_1, A'v_2)_2 w_1$$
$$= (Av_1, v_2)_1 Bw_1 = [(Bw_1)\square(Av_1)](v_2).$$

Since this holds for all $v_2 \in V_2$, assertion (i) holds. The proof of (ii) requires we show that $B' \otimes A'$ satisfies the defining equation of the adjoint—that is, for $C_1 \in \mathcal{L}(V_1, W_1)$ and $C_2 \in \mathcal{L}(V_2, W_2)$,

$$\langle C_2, (B \otimes A)C_1 \rangle_2 = \langle (B' \otimes A')C_2, C_1 \rangle_1.$$

Since outer products generate $\mathcal{L}(V_1, W_1)$, it is enough to show the above holds for $C_1 = w_1 \square x_1$ with $w_1 \in W_1$ and $x_1 \in V_1$. But, by (i) and the definition of transpose,

$$\langle C_2, (B \otimes A)(w_1 \square x_1) \rangle_2 = \langle C_2, Bw_1 \square Ax_1 \rangle_2 = [C_2 Ax_1, Bw_1]_2$$
$$= [B'C_2 Ax_1, w_1]_1 = \langle B'C_2 A, w_1 \square x_1 \rangle_1$$
$$= \langle (B' \otimes A')C_2, w_1 \square x_1 \rangle_1,$$

and this completes the proof of (ii). □

We now turn to the case when $A \in \mathcal{L}(V, V)$ and $B \in \mathcal{L}(W, W)$ so $B \otimes A$ is a linear transformation on $\mathcal{L}(V, W)$ to $\mathcal{L}(V, W)$. First note that if A is self-adjoint relative to the inner product on V and B is self-adjoint relative to the inner product on W, then Proposition 1.27 shows that $B \otimes A$ is self-adjoint relative to the natural induced inner product on $\mathcal{L}(V, W)$.

Proposition 1.28. For $A_i \in \mathcal{L}(V, V)$, $i = 1, 2$, and $B_i \in \mathcal{L}(W, W)$, $i = 1, 2$, we have:

(i) $(B_1 \otimes A_1)(B_2 \otimes A_2) = (B_1 B_2) \otimes (A_1 A_2)$.
(ii) If A_1^{-1} and B_1^{-1} exist, then $(B_1 \otimes A_1)^{-1} = B_1^{-1} \otimes A_1^{-1}$.
(iii) If A_1 and B_1 are orthogonal projections, then $B_1 \otimes A_1$ is an orthogonal projection.

Proof. The proof of (i) goes as follows: For $C \in \mathcal{L}(V, W)$,

$$(B_1 \otimes A_1)(B_2 \otimes A_2)C = (B_1 \otimes A_1)(B_2 C A_2') = B_1 B_2 C A_2' A_1'$$
$$= B_1 B_2 C (A_1 A_2)' = ((B_1 B_2) \otimes (A_1 A_2))C.$$

Now, (ii) follows immediately from (i). For (iii), it needs to be shown that $(B_1 \otimes A_1)^2 = B_1 \otimes A_1 = (B_1 \otimes A_1)'$. The second equality has been verified. The first follows from (i) and the fact that $B_1^2 = B_1$ and $A_1^2 = A_1$. □

Other properties of Kronecker products are given as the need arises. One issue to think about is this: if $C \in \mathcal{L}(V, W)$ and $B \in \mathcal{L}(W, W)$, then BC can be thought of as the product of the two linear transformations B and C. However, BC can also be interpreted as $(B \otimes I)C$, $I \in \mathcal{L}(V, V)$—that is, BC is the value of the linear transformation $B \otimes I$ at C. Of course, the particular situation determines the appropriate way to think about BC.

Linear isometries are the final subject of discussion in this section, and are a natural generalization of orthogonal transformations on $(V, (\cdot, \cdot))$. Consider finite dimensional inner product spaces V and W with inner products $(\cdot, \cdot)$ and $[\cdot, \cdot]$ and assume that $\dim V \leqslant \dim W$. The reason for this assumption is made clear in a moment.

Definition 1.23. A linear transformation $A \in \mathcal{L}(V, W)$ is a *linear isometry* if $(v_1, v_2) = [Av_1, Av_2]$ for all $v_1, v_2 \in V$.

If A is a linear isometry and $v \in V$, $v \neq 0$, then $0 < (v, v) = [Av, Av]$. This implies that $\mathfrak{N}(A) = \{0\}$, so necessarily $\dim V \leqslant \dim W$. When $W = V$

and $[\cdot, \cdot] = (\cdot, \cdot)$, then linear isometries are simply orthogonal transformations. As with orthogonal transformations, a number of equivalent descriptions of linear isometries are available.

Proposition 1.29. For $A \in \mathcal{L}(V, W)$ ($\dim V \leqslant \dim W$), the following are equivalent:

(i) A is a linear isometry.
(ii) $A'A = I \in \mathcal{L}(V, V)$.
(iii) $[Av, Av] = (v, v)$, $v \in V$.

Proof. The proof is similar to the proof of Proposition 1.19 and is left to the reader. □

The next proposition is an analog of Proposition 1.20 that covers linear isometries and that has a number of applications.

Proposition 1.30. Let $v_1, \ldots, v_k$ be vectors in $(V, (\cdot, \cdot))$, let $w_1, \ldots, w_k$ be vectors in $(W, [\cdot, \cdot])$, and assume $\dim V \leqslant \dim W$. There exists a linear isometry $A \in \mathcal{L}(V, W)$ such that $Av_i = w_i$, $i = 1, \ldots, k$, iff $(v_i, v_j) = [w_i, w_j]$ for $i, j = 1, \ldots, k$.

Proof. The proof is a minor modification of that given for Proposition 1.20 and the details are left to the reader. □

Proposition 1.31. Suppose $A \in \mathcal{L}(V, W_1)$ and $B \in \mathcal{L}(V, W_2)$ where $\dim W_2 \leqslant \dim W_1$, and $(\cdot, \cdot)$, $[\cdot, \cdot]_1$, and $[\cdot, \cdot]_2$ are inner products on V, W_1, and W_2. Then $A'A = B'B$ iff there exists a linear isometry $\Psi \in \mathcal{L}(W_2, W_1)$ such that $A = \Psi B$.

Proof. If $A = \Psi B$, then $A'A = B'\Psi'\Psi B = B'B$, since $\Psi'\Psi = I \in \mathcal{L}(W_2, W_2)$. Conversely, suppose $A'A = B'B$ and let $\{v_1, \ldots, v_m\}$ be a basis for V. With $x_i = Av_i \in W_1$ and $y_i = Bv_i \in W_2$, $i = 1, \ldots, m$, we have $[x_i, x_j]_1 = [Av_i, Av_j]_1 = (v_i, A'Av_j) = (v_i, B'Bv_j) = [Bv_i, Bv_j]_2 = [y_i, y_j]_2$ for $i, j = 1, \ldots, m$. Applying Proposition 1.30, there exists a linear isometry $\Psi \in \mathcal{L}(W_2, W_1)$ such that $\Psi y_i = x_i$ for $i = 1, \ldots, m$. Therefore, $\Psi Bv_i = Av_i$ for $i = 1, \ldots, m$ and, since $\{v_1, \ldots, v_m\}$ is a basis for V, $\Psi B = A$. □

◆ **Example 1.11.** Take $V = R^m$ and $W = R^n$ with the usual inner products and assume $m \leqslant n$. Then a matrix $A = \{a_{ij}\}: n \times m$ is a linear isometry iff $A'A = I_m$ where I_m is the $m \times m$ identity matrix. If $a_1, \ldots, a_m$ denote the columns of the matrix A, then $A'A$ is just the $m \times m$ matrix with elements $a_i'a_j$, $i, j = 1, \ldots, m$. Thus the condition $A'A = I_m$ means that $a_i'a_j = \delta_{ij}$ so A is a linear isometry on R^m to R^n iff the columns of A are an orthonormal set of vectors in R^n. Now, let $\mathfrak{F}_{m,n}$ be the set of all $n \times m$ real matrices that are linear isometries—that is, $A \in \mathfrak{F}_{m,n}$ iff $A'A = I_m$. The set $\mathfrak{F}_{m,n}$ is sometimes called the space of *m-frames in* R^n as the columns of A form an m-dimensional orthonormal "frame" in R^n. When $m = 1$, $\mathfrak{F}_{1,n}$ is just the set of vectors in R^n of length one, and when $m = n$, $\mathfrak{F}_{n,n}$ is the set of all $n \times n$ orthogonal matrices. We have much more to say about $\mathfrak{F}_{m,n}$ in later chapters.

An immediate application of Proposition 1.31 shows that, if $A : n_1 \times m$ and $B : n_2 \times m$ are real matrices with $n_2 \leqslant n_1$, then $A'A = B'B$ iff $A = \Psi B$ where $\Psi : n_1 \times n_2$ satisfies $\Psi'\Psi = I_{n_2}$. In particular, when $n_1 = n_2$, $A'A = B'B$ iff there exists an orthogonal matrix $\Psi : n_1 \times n_1$ such that $A = \Psi B$. ◆

1.6. DETERMINANTS AND EIGENVALUES

At this point in our discussion, we are forced, by mathematical necessity, to introduce complex numbers and complex matrices. Eigenvalues are defined as the roots of a certain polynomial and, to insure the existence of roots, complex numbers arise. This section begins with complex matrices, determinants, and their basic properties. After defining eigenvalues, the properties of the eigenvalues of linear transformations on real vector spaces are described.

In what follows, $\mathbb{C}$ denotes the field of complex numbers and the symbol i is reserved for $\sqrt{-1}$. If $\alpha \in \mathbb{C}$, say $\alpha = a + ib$, then $\bar{\alpha} = a - ib$ is the complex conjugate of α. Let $\mathbb{C}^n$ be the set of all n-tuples (henceforth called vectors) of complex numbers—that is, $x \in \mathbb{C}^n$ iff

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}; \qquad x_j \in \mathbb{C}, \quad j = 1, \ldots, n.$$

The number x_j is called the jth coordinate of x, $j = 1, \ldots, n$. For $x, y \in \mathbb{C}^n$, $x + y$ is defined to be the vector with coordinates $x_j + y_j$, $j = 1, \ldots, n$, and for $\alpha \in \mathbb{C}$, αx is the vector with coordinates αx_j, $j = 1, \ldots, n$. Replacing R by $\mathbb{C}$ in Definition 1.1, we see that $\mathbb{C}^n$ satisfies all the axioms of a vector space where scalars are now taken to be complex numbers, rather than real numbers. More generally, if we replace R by $\mathbb{C}$ in (II) of Definition 1.1, we have the definition of a *complex vector space*. All of the definitions, results, and proofs in Sections 1.1 and 1.2 are valid, without change, for complex vector spaces. In particular, $\mathbb{C}^n$ is an n-dimensional complex vector space and the *standard basis* for $\mathbb{C}^n$ is $\{\varepsilon_1, \ldots, \varepsilon_n\}$ where ε_j has its jth coordinate equal to one and the remaining coordinates are zero.

As with real matrices, an $m \times n$ array $A = \{a_{jk}\}$ for $j = 1, \ldots, m$, and $k = 1, \ldots, n$ where $a_{jk} \in \mathbb{C}$ is called an $m \times n$ complex matrix. If $A = \{a_{jk}\}: m \times n$ and $B = \{b_{kl}\}: n \times p$ are complex matrices, then $C = AB$ is the $m \times p$ complex matrix with with entries $c_{jl} = \sum_k a_{jk} b_{kl}$ for $j = 1, \ldots, m$ and $l = 1, \ldots, p$. The matrix C is called the *product* of A and B (in that order). In particular, when $p = 1$, the matrix B is $n \times 1$ so B is an element of $\mathbb{C}^n$. Thus if $x \in \mathbb{C}^n$ (x now plays the role of B) and $A : m \times n$ is a complex matrix, $Ax \in \mathbb{C}^m$. Clearly, each $A : m \times n$ determines a linear transformation on $\mathbb{C}^n$ to $\mathbb{C}^m$ via the definition of Ax for $x \in \mathbb{C}^n$. For an $m \times n$ complex matrix $A = \{a_{jk}\}$, the *conjugate transpose* of A, denoted by A^*, is the $n \times m$ matrix $A^* = \{\bar{a}_{kj}\}$, $k = 1, \ldots, n, j = 1, \ldots, m$, where $\bar{a}_{kj}$ is the complex conjugate of $a_{kj} \in \mathbb{C}$. In particular, if $x \in \mathbb{C}^n$, x^* denotes the *conjugate transpose* of x. The following relation is easily verified:

$$\overline{y^*Ax} = x^*A^*y$$

where $y \in \mathbb{C}^m$, $x \in \mathbb{C}^n$, and A is an $m \times n$ complex matrix. Of course, the bar over y^*Ax denotes the complex conjugate of $y^*Ax \in \mathbb{C}$.

With the preliminaries out of the way, we now want to define determinant functions. Let $\mathcal{C}_n$ denote the set of all $n \times n$ complex matrices so $\mathcal{C}_n$ is an n^2-dimensional complex vector space. If $A \in \mathcal{C}_n$, write $A = (a_1, a_2, \ldots, a_n)$ where a_j is the jth column of A.

Definition 1.24. A function D defined on $\mathcal{C}_n$ and taking values in $\mathbb{C}$ is called a *determinant function* if

(i) $D(A) = D(a_1, \ldots, a_n)$ is linear in each column vector a_j when the other columns are held fixed. That is,

$$D(a_1, \ldots, \alpha a_j + \beta b_j, \ldots, a_n) = \alpha D(a_1, \ldots, a_j, \ldots, a_n) + \beta D(a_1, \ldots, b_j, \ldots, a_n)$$

for $\alpha, \beta \in \mathbb{C}$.

(ii) For any two indices j and k, $j < k$,

$$D(a_1, \ldots, a_j, \ldots, a_k, \ldots, a_n) = -D(a_1, \ldots, a_k, \ldots, a_j, \ldots, a_n).$$

Functions D on $\mathcal{C}_n$ to $\mathbb{C}$ that satisfy (i) are called *n-linear* since they are linear in each of the n vectors $a_1,\ldots, a_n$ when the remaining ones are held fixed. If D is n-linear and satisfies (ii), D is sometimes called an *alternating* n-linear function, since $D(A)$ changes sign if two columns of A are interchanged. The basic result that relates all determinant functions is the following.

Proposition 1.32. The set of determinant functions is a one-dimensional complex vector space. If D is a determinant function and $D \not\equiv 0$, then $D(I) \neq 0$ where I is the $n \times n$ identity matrix in $\mathcal{C}_n$.

Proof. We briefly outline the proof of this proposition since the proof is instructive and yields the classical formula defining the determinant of an $n \times n$ matrix. Suppose $D(A) = D(a_1,\ldots, a_n)$ is a determinant function. For each $k = 1,\ldots, n$, $a_k = \sum_j a_{jk}\varepsilon_j$ where $\{\varepsilon_1,\ldots, \varepsilon_n\}$ is the standard basis for $\mathbb{C}_n$ and $A = \{a_{jk}\}: n \times n$. Since D is n-linear and $a_1 = \sum a_{j1}\varepsilon_j$,

$$D(a_1,\ldots, a_n) = \sum_j a_{j1} D(\varepsilon_j, a_2,\ldots, a_n).$$

Applying this same argument for $a_2 = \sum a_{j2}\varepsilon_j$,

$$D(a_1,\ldots, a_n) = \sum_{j_1}\sum_{j_2} a_{j_1 1}a_{j_2 2} D(\varepsilon_{j_1}, \varepsilon_{j_2}, a_3,\ldots, a_n).$$

Continuing in the obvious way,

$$D(a_1,\ldots, a_n) = \sum_{j_1,\ldots,j_n} a_{j_1 1}a_{j_2 2}\cdots a_{j_n n} D(\varepsilon_{j_1}, \varepsilon_{j_2},\ldots, \varepsilon_{j_n})$$

where the summation extends over all $j_1,\ldots, j_n$ with $1 \leqslant j_l \leqslant n$ for $l = 1,\ldots, n$. The above formula shows that a determinant function is determined by the n^n numbers $D(\varepsilon_{j_1},\ldots, \varepsilon_{j_n})$ for $1 \leqslant j_l \leqslant n$, and this fact followed solely from the assumption that D is n-linear. But since D is alternating, it is clear that, if two columns of A are the same, then $D(A) = 0$. In particular, if two indices j_l and j_k are the same, then $D(\varepsilon_{j_1},\ldots, \varepsilon_{j_n}) = 0$. Thus the summation above extends only over those indices where $j_1,\ldots, j_n$ are all distinct. In other words, the summation extends over all permutations of the set $\{1, 2,\ldots, n\}$. If π denotes a permutation of $1, 2,\ldots, n$, then

$$D(a_1,\ldots, a_n) = \sum_\pi a_{\pi(1)1}\cdots a_{\pi(n)n} D(\varepsilon_{\pi(1)},\ldots, \varepsilon_{\pi(n)})$$

where the summation now extends over all $n!$ permutations. But for a fixed permutation $\pi(1),\dots,\pi(n)$ of $1,\dots,n$, there is a sequence of pairwise interchanges of the elements of $\pi(1),\dots,\pi(n)$, which results in the order $1,2,\dots,n$. In fact there are many such sequences of interchanges, but the number of interchanges is always odd or always even (see Hoffman and Kunze, 1971, Section 5.3). Using this, let $\text{sgn}(\pi)=1$ if the number of interchanges required to put $\pi(1),\dots,\pi(n)$ into the order $1,2,\dots,n$ is even and let $\text{sgn}(\pi)=-1$ otherwise. Now, since D is alternating, it is clear that

$$D(\varepsilon_{\pi(1)},\dots,\varepsilon_{\pi(n)})=\text{sgn}(\pi)D(\varepsilon_1,\dots,\varepsilon_n).$$

Therefore, we have arrived at the formula $D(a_1,\dots,a_n)=D(I)\Sigma_\pi\text{sgn}(\pi)a_{\pi(1)1}\cdots a_{\pi(n)n}$ since $D(I)=D(\varepsilon_1,\dots,\varepsilon_n)$. It is routine to verify that, for any complex number α, the function defined by

$$D_\alpha(a_1,\dots,a_n)\equiv\alpha\sum_\pi\text{sgn}(\pi)a_{\pi(1)1}\cdots a_{\pi(n)n}$$

is a determinant function and the argument given above shows that every determinant function is a D_α for some $\alpha\in\mathbb{C}$. This completes the proof; for more details, the reader is referred to Hoffman and Kunze (1971, Chapter 5). □

Definition 1.25. If $A\in\mathcal{C}_n$, the *determinant* of A, denoted by $\det(A)$ (or $\det A$), is defined to be $D_1(A)$ where D_1 is the unique determinant function with $D_1(I)=1$.

The proof of Proposition 1.32 gives the formula for $\det(A)$, but that is not of much concern to us. The properties of $\det(\cdot)$ given below are most easily established using the fact that $\det(\cdot)$ is an alternating n-linear function of the columns of A.

Proposition 1.33. For $A, B\in\mathcal{C}_n$:

(i) $\det(AB)=\det A\det B$.
(ii) $\det A^*=\overline{\det A}$.
(iii) $\det A\neq 0$ iff the columns of A are linear independent vectors in the complex vector space $\mathbb{C}^n$.

If $A_{11}:n_1\times n_1$, $A_{12}:n_1\times n_2$, $A_{21}:n_2\times n_1$, and $A_{22}:n_2\times n_2$ are complex matrices, then:

(iv) $\det\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} = \det\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} = \det A_{11} \det A_{22}$.

(v) If A is a real matrix, then $\det(A)$ is real and $\det(A) = 0$ iff the columns of A are linearly dependent vectors over the real vector space R^n.

Proof. The proofs of these assertions can be found in Hoffman and Kunze (1971, Chapter 5). □

These properties of $\det(\cdot)$ have a number of useful and interesting implications. If A has columns $a_1, \dots, a_n$, then the range of the linear transformation determined by A is just $\text{span}\{a_1, \dots, a_n\}$. Thus A is invertible iff $\text{span}\{a_1, \dots, a_n\} = \mathbb{C}^n$ iff $\det A \neq 0$. If $\det A \neq 0$, then $1 = \det AA^{-1} = \det A \det A^{-1}$, so $\det A^{-1} = 1/\det A$. Consider $B_{11} : n_1 \times n_1$, $B_{12} : n_1 \times n_2$, $B_{21} : n_2 \times n_1$, and $B_{22} : n_2 \times n_2$— complex matrices. Then it is easy to verify the identity:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

where A_{11}, A_{12}, A_{21}, and A_{22} are defined in Proposition 1.33. This tells us how to multiply the two $(n_1 + n_2) \times (n_1 + n_2)$ complex matrices in terms of their blocks. Of course, such matrices are called partitioned matrices.

Proposition 1.34. Let A be a complex matrix, partitioned as above. If $\det A_{11} \neq 0$, then:

(i) $\det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det A_{11} \det(A_{22} - A_{21}A_{11}^{-1}A_{12})$.

If $\det A_{22} \neq 0$, then:

(ii) $\det\begin{pmatrix} A_{11} & A_{12} \\ A_{22} & A_{22} \end{pmatrix} = \det A_{22} \det(A_{11} - A_{12}A_{22}^{-1}A_{21})$.

Proof. For (i), first note that

$$\det\begin{pmatrix} I_{n_1} & -A_{11}^{-1}A_{12} \\ 0 & I_{n_2} \end{pmatrix} = 1,$$

by Proposition 1.33, (iv). Therefore, by (i) of Proposition 1.33,

$$\det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} I_{n_1} & -A_{11}^{-1}A_{12} \\ 0 & I_{n_2} \end{pmatrix}$$

$$= \det\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix}.$$

$$= \det A_{11}\det\left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right).$$

The proof of (ii) is similar. □

Proposition 1.35. Let $A : n \times m$ and $B : m \times n$ be complex matrices. Then

$$\det(I_n + AB) = \det(I_m + BA).$$

Proof. Apply the previous proposition to

$$\begin{pmatrix} I_n & -A \\ B & I_m \end{pmatrix}.$$

□

We now turn to a discussion of the eigenvalues of an $n \times n$ complex matrix. The definition of an eigenvalue is motivated by the following considerations. Let $A \in \mathcal{C}_n$. To analyze the linear transformation determined by A, we would like to find a basis $x_1, \ldots, x_n$ of $\mathbb{C}_n$ such that $Ax_j = \lambda_j x_j$, $j = 1, \ldots, n$, where $\lambda_j \in \mathbb{C}$. If this were possible, then the matrix of the linear transformation in the basis $\{x_1, \ldots, x_n\}$ would simply be

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

where the elements not indicated are zero. Of course, this says that the linear transformation is λ_j times the identity transformation when restricted to span$\{x_j\}$. Unfortunately, it is not possible to find such a basis for each linear transformation. However, the numbers $\lambda_1, \ldots, \lambda_n$, which are called eigenvalues after we have an appropriate definition, can be interpreted in another way. Given $\lambda \in \mathbb{C}$, $Ax = \lambda x$ for some nonzero vector x iff $(A - \lambda I)x = 0$, and this is equivalent to saying that $A - \lambda I$ is a singular matrix, that is, $\det(A - \lambda I) = 0$. In other words, $A - \lambda I$ is singular iff there exists $x \neq 0$ such that $Ax = \lambda x$. However, using the formula for $\det(\cdot)$, a bit of calculation shows that

$$\det(A - \lambda I) = (-1)^n \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1 \lambda + \alpha_0$$

where $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ are complex numbers. Thus $\det(A - \lambda I)$ is a polynomial of degree n in the complex variable λ, and it has n roots (counting multiplicities). This leads to the following definition.

Definition 1.26. Let $A \in \mathcal{C}_n$ and set

$$p(\lambda) = \det(A - \lambda I).$$

Then nth degree polynomial p is called the *characteristic polynomial* of A and the n roots of the polynomial (counting multiplicities) are called the *eigenvalues* of A.

If $p(\lambda) = \det(A - \lambda I)$ has roots $\lambda_1, \ldots, \lambda_n$, then it is clear that

$$p(\lambda) = \prod_{j=1}^{n} (\lambda_j - \lambda)$$

since the right-hand side of the above equation is an nth degree polynomial with roots $\lambda_1, \ldots, \lambda_n$ and the coefficient of λ^n is $(-1)^n$. In particular,

$$p(0) = \prod_{1}^{n} \lambda_j = \det(A)$$

so the determinant of A is the product of its eigenvalues.

There is a particular case when the characteristic polynomial of A can be computed explicitly. If $A \in \mathcal{C}_n$, $A = \{a_{jk}\}$ is called *lower triangular* if $a_{jk} = 0$ when $k > j$. Thus A is lower triangular if all the elements above the diagonal of A are zero. An application of Proposition 1.33 (iv) shows that when A is lower triangular, then

$$\det(A) = \prod_{1}^{n} a_{jj}.$$

But when A is lower triangular with diagonal elements $a_{jj}, j = 1, \ldots, n$, then $A - \lambda I$ is lower triangular with diagonal elements $(a_{jj} - \lambda)$, $j = 1, \ldots, n$.

Thus

$$p(\lambda) = \det(A - \lambda I) = \prod_1^n (a_{jj} - \lambda),$$

so A has eigenvalues $a_{11}, \ldots, a_{nn}$.

Before returning to real vector spaces, we first establish the existence of eigenvectors (to be defined below).

Proposition 1.36. If λ is an eigenvalue of $A \in \mathcal{C}_n$, then there exists a nonzero vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$.

Proof. Since λ is an eigenvalue of A, the matrix $A - \lambda I$ is singular, so the dimension of the range of $A - \lambda I$ is less than n. Thus the dimension of the null space of $A - \lambda I$ is greater than 0. Hence there is a nonzero vector in the null space of $A - \lambda I$, say x, and $(A - \lambda I)x = 0$. □

Definition 1.27. If $A \in \mathcal{C}_n$, a nonzero vector $x \in C^n$ is called an *eigenvector* of A if there exists a complex number $\lambda \in \mathbb{C}$ such that $Ax = \lambda x$.

If $x \neq 0$ is an eigenvector of A and $Ax = \lambda x$, then $(A - \lambda I)x = 0$ so $A - \lambda I$ is singular. Therefore, λ must be an eigenvalue for A. Conversely, if $\lambda \in \mathbb{C}$ is an eigenvalue, Proposition 1.36 shows there is an eigenvector x such that $Ax = \lambda x$.

Now, suppose V is an n-dimensional real vector space and B is a linear transformation on V to V. We want to define the characteristic polynomial, and hence the eigenvalues of B. Let $\{v_1, \ldots, v_n\}$ be a basis for V so the matrix of B is $[B] = \{b_{jk}\}$ where the b_{jk}'s satisfy $Bv_k = \Sigma_j b_{jk} v_j$. The characteristic polynomial of $[B]$ is

$$p(\lambda) = \det([B] - \lambda I)$$

where I is the $n \times n$ identity matrix and $\lambda \in \mathbb{C}$. If we could show that $p(\lambda)$ does not depend on the particular basis for V, then we would have a reasonable definition of the characteristic polynomial of B.

Proposition 1.37. Suppose $\{v_1, \ldots, v_n\}$ and $\{y_1, \ldots, y_n\}$ are bases for the real vector space V, and let $B \in \mathcal{L}(V, V)$. Let $[B] = \{b_{jk}\}$ be the matrix of B in the basis $\{v_1, \ldots, v_n\}$ and let $[B]_1 = \{a_{jk}\}$ be the matrix of B in the basis $\{y_1, \ldots, y_n\}$. Then there exists a nonsingular real matrix $C = \{c_{jk}\}$ such that

$$[B]_1 = C^{-1}[B]C.$$

Proof. The numbers a_{jk} are uniquely determined by the relations

$$By_k = \sum_j a_{jk} y_j, \qquad k = 1, \dots, n.$$

Define the linear transformation C_1 on V to V by $C_1 v_j = y_j$, $j = 1, \dots, n$. Then C_1 is nonsingular since C_1 maps a basis onto a basis. Therefore,

$$BC_1 v_k = \sum_j a_{jk} C_1 v_j = C_1 \Big(\sum_j a_{jk} v_j \Big)$$

and this yields

$$C_1^{-1} B C_1 v_k = \sum_j a_{jk} v_j.$$

Thus the matrix of $C_1^{-1} B C_1$ in the basis $\{v_1, \dots, v_n\}$ is $\{a_{jk}\}$. From Proposition 1.5, we have

$$[B]_1 = \{a_{jk}\} = [C_1^{-1} B C_1] = [C_1^{-1}][B][C_1] = [C_1]^{-1}[B][C_1]$$

where $[C_1]$ is the matrix of C_1 in the basis $\{v_1, \dots, v_n\}$. Setting $C = [C_1]$, the conclusion follows. □

The above proposition implies that

$$\begin{aligned} p(\lambda) &= \det([B] - \lambda I) = \det\big(C^{-1}([B] - \lambda I)C\big) \\ &= \det\big(C^{-1}[B]C - \lambda I\big) = \det([B]_1 - \lambda I). \end{aligned}$$

Thus $p(\lambda)$ does not depend on the particular basis we use to represent B, and, therefore, we call p the characteristic polynomial of the linear transformation B. The suggestive notation

$$p(\lambda) = \det(B - \lambda I)$$

is often used. Notice that Proposition 1.37 also shows that it makes sense to define $\det(B)$ for $B \in \mathcal{L}(V, V)$ as the value of $\det[B]$ in any basis, since the value does not depend on the basis. Of course, the roots of the polynomial $p(\lambda) = \det(B - \lambda I)$ are called the eigenvalues of the linear transformation B. Even though $[B]$ is a real matrix in any basis for V, some or all of the eigenvalues of B may be complex numbers. Proposition 1.37 also allows us to define the trace of $A \in \mathcal{L}(V, V)$. If $\{v_1, \ldots, v_n\}$ is a basis for V, let $\operatorname{tr} A \equiv \operatorname{tr}[A]$ where $[A]$ is the matrix of A in the given basis. For any nonsingular matrix C,

$$\operatorname{tr}[A] = \operatorname{tr} CC^{-1}[A] = \operatorname{tr} C^{-1}[A]C,$$

which shows that our definition of $\operatorname{tr} A$ does not depend on the particular basis chosen.

The next result summarizes the properties of eigenvalues for linear transformations on a real inner product space.

Proposition 1.38. Suppose $(V, (\cdot, \cdot))$ is a finite dimensional real inner product space and let $A \in \mathcal{L}(V, V)$.

(i) If $\lambda \in \mathbb{C}$ is an eigenvalue of A, then $\bar{\lambda}$ is an eigenvalue of A.
(ii) If A is symmetric, the eigenvalues of A are real
(iii) If A is skew-symmetric, then the eigenvalues of A are pure imaginary
(iv) If A is orthogonal and λ is an eigenvalue of A, then $\lambda\bar{\lambda} = 1$.

Proof. If $A \in \mathcal{L}(V, V)$, then the characteristic polynomial of A is

$$p(\lambda) = \det([A] - \lambda I), \qquad \lambda \in \mathbb{C},$$

where $[A]$ is the matrix of A in a basis for V. An examination of the formula for $\det(\cdot)$ shows that

$$p(\lambda) = (-1)^n \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0$$

where $\alpha_0, \ldots, \alpha_{n-1}$ are real numbers since $[A]$ is a real matrix Thus if $p(\lambda) = 0$, then $p(\bar{\lambda}) = \overline{p(\lambda)} = 0$ so whenever $p(\lambda) = 0$, $p(\bar{\lambda}) = 0$. This establishes assertion (i).

For (ii), let λ be an eigenvalue of A, and let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for $(V, (\cdot, \cdot))$. Thus the matrix of A, say $[A]$, is a real symmetric matrix and $[A] - \lambda I$ is singular as a matrix acting on $\mathbb{C}^n$. By Proposition 1.36, there exists a nonzero vector $x \in \mathbb{C}^n$ such that $[A]x = \lambda x$. Thus $x^*[A]x = \lambda x^* x$. But since $[A]$ is real and symmetric,

$$\overline{x^*[A]x} = x^*[A]^*x = x^*[A]x = \overline{\lambda x^* x} = \bar{\lambda} x^* x.$$

Thus $\bar{\lambda} x^* x = \lambda x^* x$ and, since $x \neq 0$, $\bar{\lambda} = \lambda$ so λ is real.

To prove (iii), again let $[A]$ be the matrix of A in the orthonormal basis $\{v_1, \ldots, v_n\}$ so $[A]' = [A]^* = -[A]$. If λ is an eigenvalue of A, then there exists $x \in \mathbb{C}^n$, $x \neq 0$, such that $[A]x = \lambda x$. Thus $x^*[A]x = \lambda x^* x$ and

$$\bar{\lambda} x^* x = \overline{x^*[A]x} = -x^*[A]x = -\lambda x^* x.$$

Since $x \neq 0$, $\bar{\lambda} = -\lambda$, which implies that $\lambda = ib$ for some real number b—that is, λ is pure imaginary and this proves (iii).

If A is orthogonal, then $[A]$ is an $n \times n$ orthogonal matrix in the orthonormal basis $\{v_1, \ldots, v_n\}$. Again, if λ is an eigenvalue of A, then $[A]x = \lambda x$ for some $x \in \mathbb{C}^n$, $x \neq 0$. Thus $\bar{\lambda} x^* = x^*[A]^* = x^*[A]'$ since $[A]$ is a real matrix. Therefore

$$\lambda \bar{\lambda} x^* x = x^*[A]'[A]x = x^* x$$

as $[A]'[A] = I$. Hence $\lambda\bar{\lambda} = 1$ and the proof of Proposition 1.38 is complete. □

It has just been shown that if $(V, (\cdot, \cdot))$ is a finite dimensional vector space and if $A \in \mathcal{L}(V, V)$ is self-adjoint, then the eigenvalues of A are real. The spectral theorem, to be established in the next section, provides much more useful information about self-adjoint transformations. For example, one application of the spectral theorem shows that a self-adjoint transformation is positive definite iff all its eigenvalues are positive.

If $A \in \mathcal{L}(V, W)$ and $B \in \mathcal{L}(W, V)$, the next result compares the eigenvalues of $AB \in \mathcal{L}(W, W)$ with those of $BA \in \mathcal{L}(V, V)$.

Proposition 1.39. The nonzero eigenvalues of AB are the same as the nonzero eigenvalues of BA, including multiplicities. If $W = V$, AB and BA have the same eigenvalues and multiplicities.

Proof. Let $m = \dim V$ and $n = \dim W$. The characteristic polynomial of BA is

$$p_1(\lambda) = \det(BA - \lambda I_m).$$

Now, for $\lambda \neq 0$, compute as follows:

$$\begin{aligned}\det(BA - \lambda I_m) &= \det(-\lambda)\left(\frac{BA}{-\lambda} + I_m\right) \\ &= (-\lambda)^m \det\left(\frac{BA}{-\lambda} + I_m\right) = (-\lambda)^m \det\left(\frac{AB}{-\lambda} + I_n\right) \\ &= (-\lambda)^m \det\left(\frac{1}{-\lambda}\right)(AB - \lambda I_n) = \frac{(-\lambda)^m}{(-\lambda)^n} \det(AB - \lambda I_n).\end{aligned}$$

Therefore, the characteristic polynomial of AB, say $p_2(\lambda) = \det(AB - \lambda I_n)$, is related to $p_1(\lambda)$ by

$$p_1(\lambda) = \frac{(-\lambda)^m}{(-\lambda)^n} p_2(\lambda), \qquad \lambda \in \mathbb{C}, \quad \lambda \neq 0.$$

Both of the assertions follow from this relationship. □

1.7. THE SPECTRAL THEOREM

The spectral theorem for self-adjoint linear transformations on a finite dimensional real inner product space provides a basic theoretical tool not only for understanding self-adjoint transformations but also for establishing a variety of useful facts about general linear transformations. The form of the spectral theorem given below is slightly weaker than that given in Halmos (1958, see Section 79), but it suffices for most of our purposes. Applications of this result include a necessary and sufficient condition that a self-adjoint transformation be positive definite and a demonstration that positive definite transformations possess square roots. The singular value decomposition theorem, which follows from the spectral theorem, provides a useful decomposition result for linear transformations on one inner product space to another. This section ends with a description of the relationship between the singular value decomposition theorem and angles between two subspaces of an inner product space.

Let $(V, (\cdot, \cdot))$ be a finite dimensional real inner product space. The spectral theorem follows from the two results below. If $A \in \mathcal{L}(V, V)$ and M is subspace of V, M is called *invariant* under A if $A(M) = \{Ax | x \in M\} \subseteq M$.

Proposition 1.40. Suppose $A \in \mathcal{L}(V, V)$ is self-adjoint and let M be a subspace of V. If $A(M) \subseteq M$, then $A(M^\perp) \subseteq M^\perp$.

Proof. Suppose $v \in A(M^\perp)$. It must be shown that $(v, x) = 0$ for all $x \in M$. Since $v \in A(M^\perp)$, $v = Av_1$ for $v_1 \in M^\perp$. Therefore,

$$(v, x) = (Av_1, x) = (v_1, Ax) = 0$$

since A is self-adjoint and $x \in M$ implies $Ax \in M$ by assumption. □

Proposition 1.41. Suppose $A \in \mathcal{L}(V, V)$ is self-adjoint and λ is an eigenvalue of A. Then there exists a $v \in V$, $v \neq 0$, such that $Av = \lambda v$.

Proof. Since A is self-adjoint, the eigenvalues of A are real. Let $\{v_1, \dots, v_n\}$ be a basis for V and let $[A]$ be the matrix of A in this basis. By Proposition 1.36, there exists a nonzero vector $z \in \mathbb{C}^n$ such that $[A]z = \lambda z$. Write $z = z_1 + iz_2$ where $z_1 \in R^n$ is the real part of z and $z_2 \in R^n$ is the imaginary part of z. Since $[A]$ is real and λ is real, we have $[A]z_1 = \lambda z_1$ and $[A]z_2 = \lambda z_2$. But, z_1 and z_2 cannot both be zero as $z \neq 0$. For definiteness, say $z_1 \neq 0$ and let $v \in V$ be the vector whose coordinates in basis $\{v_1, \dots, v_n\}$ are z_1. Then $v \neq 0$ and $[A][v] = \lambda[v]$. Therefore $Av = \lambda v$. □

Theorem 1.2 (Spectral Theorem). If $A \in \mathcal{L}(V, V)$ is self-adjoint, then there exists an orthonormal basis $\{x_1, \dots, x_n\}$ for V and real numbers $\lambda_1, \dots, \lambda_n$ such that

$$A = \sum_1^n \lambda_i x_i \square x_i.$$

Further, $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and $Ax_i = \lambda_i x_i$, $i = 1, \dots, n$.

Proof. The proof of the first assertion is by induction on dimension. For $n = 1$, the result is obvious. Assume the result is true for integers $1, 2, \dots, n - 1$ and consider $A \in \mathcal{L}(V, V)$, which is self-adjoint on the inner product space $(V, (\cdot, \cdot))$, $n = \dim V$. Let λ be an eigenvalue of A. By Proposition 1.41, there exists $v \in V$, $v \neq 0$, such that $Av = \lambda v$. Set $x_n = v/\|v\|$ and $\lambda_n = \lambda$. Then $Ax_n = \lambda_n x_n$. With $M = \text{span}\{x_n\}$, it is clear that $A(M) \subseteq M$ so $A(M^\perp) \subseteq M^\perp$ by Proposition 1.40. However, if we let A_1 be the restriction of A to the $(n-1)$-dimensional inner product space $(M^\perp, (\cdot, \cdot))$, then A_1 is clearly self-adjoint. By the induction hypothesis there is an orthonormal basis $\{x_1, \dots, x_{n-1}\}$ for $M^\perp$ and real numbers $\lambda_1, \dots, \lambda_{n-1}$ such that

$$A_1 = \sum_1^{n-1} \lambda_i x_i \square x_i.$$

It is clear that $\{x_1, \dots, x_n\}$ is an orthonormal basis for V and we claim that

$$A = \sum_1^n \lambda_i x_i \square x_i.$$

To see this, consider $v_0 \in V$ and write $v_0 = v_1 + v_2$ with $v_1 \in M$ and $v_2 \in M^\perp$. Then

$$Av_0 = Av_1 + Av_2 = \lambda_n v_1 + A_1 v_2 = \lambda_n v_1 + \sum_1^{n-1} \lambda_i (x_i \square x_i) v_2.$$

However,

$$\left(\sum_1^n \lambda_i x_i \square x_i\right)(v_1 + v_2) = \lambda_n (v_1, x_n) x_n + \sum_1^{n-1} \lambda_i (x_i \square x_i) v_2$$

since $v_1 \in M$ and $v_2 \in M^{\perp}$. But $(v_1, x_n)x_n = v_1$ since $v_1 \in \text{span}\{x_n\}$. Therefore $A = \sum_1^n \lambda_i x_i \square x_i$, which establishes the first assertion.

For the second assertion, if $A = \sum_1^n \lambda_i x_i \square x_i$ where $\{x_1, \ldots, x_n\}$ is an orthonormal basis for $(V, (\cdot, \cdot))$, then

$$Ax_j = \sum_i \lambda_i (x_i \square x_i) x_j = \sum_i \lambda_i (x_i, x_j) x_i = \lambda_j x_j.$$

Thus the matrix of A, say $[A]$, in this basis has diagonal elements $\lambda_1, \ldots, \lambda_n$ and all other elements of $[A]$ are zero. Therefore the characteristic polynomial of A is

$$p(\lambda) = \det([A] - \lambda I) = \prod_1^n (\lambda_i - \lambda),$$

which has roots $\lambda_1, \ldots, \lambda_n$. The proof of the spectral theorem is complete. □

When $A = \sum \lambda_i x_i \square x_i$, then A is particularly easy to understand. Namely, A is λ_i times the identity transformation when restricted to $\text{span}\{x_i\}$. Also, if $x \in V$, then $x = \sum (x_i, x) x_i$ so $Ax = \sum \lambda_i (x_i, x) x_i$. In the case when A is an orthogonal projection onto the subspace M, we know that $A = \sum_1^k x_i \square x_i$ where $k = \dim M$ and $\{x_1, \ldots, x_k\}$ is an orthonormal basis for M. Thus A has eigenvalues of zero and one, and one occurs with multiplicity $k = \dim M$. Conversely, the spectral theorem implies that, if A is self-adjoint and has only zero and one as eigenvalues, then A is an orthogonal projection onto a subspace of dimensional equal to the multiplicity of the eigenvalue one.

We now begin to reap the benefits of the spectral theorem.

Proposition 1.42. If $A \in \mathcal{L}(V, V)$, then A is positive definite iff all the eigenvalues of A are strictly positive. Also, A is positive semidefinite iff the eigenvalues of A are non-negative.

Proof. Write A in spectral form:

$$A = \sum_1^n \lambda_i x_i \square x_i$$

where $\{x_1,\ldots, x_n\}$ is an orthonormal basis for $(V,(\cdot,\cdot))$. Then $(x, Ax) = \Sigma\lambda_i(x_i, x)^2$. If $\lambda_i > 0$ for $i = 1,\ldots, n$, then $x \neq 0$ implies that $\Sigma\lambda_i(x_i, x)^2 > 0$ and A is positive definite. Conversely, if A is positive definite, set $x = x_j$ and we have $0 < (x_j, Ax_j) = \lambda_j$. Thus all the eigenvalues of A are strictly positive. The other assertion is proved similarly. □

The representation of A in spectral form suggests a way to define various functions of A. If $A = \Sigma\lambda_i x_i \Box x_i$, then

$$A^2 = \Big(\sum_i \lambda_i x_i \Box x_i\Big)\Big(\sum_j \lambda_j x_j \Box x_j\Big) = \sum_i\sum_j \lambda_i\lambda_j (x_i \Box x_i)(x_j \Box x_j)$$

$$= \sum_i\sum_j \lambda_i\lambda_j (x_i, x_j) x_i \Box x_j = \sum_i \lambda_i^2 x_i \Box x_i.$$

More generally, if k is a positive integer, a bit of calculation shows that

$$A^k = \sum_i \lambda_i^k x_i \Box x_i, \qquad k = 1, 2, \ldots.$$

For $k = 0$, we adopt the convention that $A^0 = I$ since $\Sigma x_i \Box x_i = I$. Now if p is any polynomial on R, the above equation forces us to define $p(A)$ by

$$p(A) = \sum_i p(\lambda_i) x_i \Box x_i.$$

This suggests that, if f is any real-valued function that is defined at $\lambda_1,\ldots, \lambda_n$, we should define $f(A)$ by

$$f(A) = \sum_i f(\lambda_i) x_i \Box x_i.$$

Adopting this suggestive definition shows that if $\lambda_1,\ldots, \lambda_n$ are the eigenvalues of A, then $f(\lambda_1),\ldots, f(\lambda_n)$ are the eigenvalues of $f(A)$. In particular, if $\lambda_i \neq 0$ for all i and $f(t) = t^{-1}$, $t \neq 0$, then it is clear that $f(A) = A^{-1}$. Another useful choice for f is given in the following proposition.

Proposition 1.43. If $A \in \mathcal{L}(V, V)$ is positive semidefinite, then there exists a $B \in \mathcal{L}(V, V)$ that is positive semidefinite and satisfies $B^2 = A$.

Proof. Choose $f(t) = t^{1/2}$, and let

$$B \equiv f(A) = \sum_1^n \lambda_i^{1/2} x_i \Box x_i.$$

The square root is well defined since $\lambda_i \geqslant 0$ for $i = 1, \ldots, n$ as A is positive semidefinite. Since B has non-negative eigenvalues, B is positive definite. That $B^2 = A$ is clear. □

There is a technical problem with our definition of $f(A)$ that is caused by the nonuniqueness of the representation

$$A = \sum_1^n \lambda_i x_i \square x_i$$

for self-adjoint transformations. For example, if the first n_1 λ_i's are equal and the last $n - n_1$ λ_i's are equal, then

$$A = \lambda_1\left(\sum_1^{n_1} x_i \square x_i\right) + \lambda_n\left(\sum_{n_1+1}^{n} x_i \square x_i\right).$$

However, $\Sigma_1^{n_1} x_i \square x_i$ is the orthogonal projection onto $M_1 \equiv \text{span}(x_1, \ldots, x_{n_1}\}$. If $y_1, \ldots, y_n$ is any other orthonormal basis for $(V, (\cdot, \cdot))$ such that $\text{span}\{x_1, \ldots, x_{n_1}\} = \text{span}\{y_1, \ldots, y_{n_1}\}$, it is clear that

$$A = \lambda_1 \sum_1^{n_1} y_i \square y_i + \lambda_n \sum_{n_1+1}^{n} y_i \square y_i = \sum_1^n \lambda_i y_i \square y_i.$$

Obviously, $\lambda_1, \ldots, \lambda_n$ are uniquely defined as the eigenvalues for A (counting multiplicities), but the orthonormal basis $\{x_1, \ldots, x_n\}$ providing the spectral form for A is not unique. It is therefore necessary to verify that the definition of $f(A)$ does not depend on the particular orthonormal basis in the representation for A or to provide an alternative representation for A. It is this latter alternative that we follow. The result below is also called the spectral theorem.

Theorem 1.2a (Spectral Theorem). Suppose A is a self-adjoint linear transformation on V to V where $n = \dim V$. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ be the distinct eigenvalues of A and let n_i be the multiplicity of λ_i, $i = 1, \ldots, r$. Then there exists orthogonal projections $P_1, \ldots, P_r$ with $P_i P_j = 0$ for $i \neq j$, $n_i = \text{rank}(P_i)$, and $\Sigma_1^n P_i = I$ such that

$$A = \sum_1^r \lambda_i P_i.$$

Further, this decomposition is unique in the following sense. If $\mu_1 > \cdots >$

μ_k and $Q_1, \dots, Q_k$ are orthogonal projections such that $Q_i Q_j = 0$ for $i \neq j$, $\Sigma Q_i = I$, and

$$A = \sum_1^k \mu_i Q_i,$$

then $k = r$, $\mu_i = \lambda_i$, and $Q_i = P_i$ for $i = 1, \dots, k$.

Proof. The first assertion follows immediately from the spectral representation given in Theorem 1.2. For a proof of the uniqueness assertion, see Halmos (1958, Section 79). □

Now, our definition of $f(A)$ is

$$f(A) = \sum_1^r f(\lambda_i) P_i$$

when $A = \Sigma_1^r \lambda_i P_i$. Of course, it is assumed that f is defined at $\lambda_1, \dots, \lambda_r$. This is exactly the same definition as before, but the problem about the nonuniqueness of the representation of A has disappeared. One application of the uniqueness part of the above theorem is that the positive semidefinite square root given in Proposition 1.43 is unique. The proof of this is left to the reader (see Halmos, 1958, Section 82).

Other functions of self-adjoint linear transformations come up later and we consider them as the need arises. Another application of the spectral theorem solves an interesting extremal problem. To motivate this problem, suppose A is self-adjoint on $(V, (\cdot, \cdot))$ with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Thus $A = \Sigma \lambda_i x_i \square x_i$ where $\{x_1, \dots, x_n\}$ is an orthonormal basis for V. For $x \in V$ and $\|x\| = 1$, we ask how large (x, Ax) can be. To answer this, write $(x, Ax) = \Sigma \lambda_i (x, (x_i \square x_i) x) = \Sigma \lambda_i (x, x_i)^2$, and note that $0 \leqslant (x, x_i)^2$ and $1 = \|x\|^2 = \Sigma_1^n (x_i, x)^2$. Therefore, $\lambda_1 \geqslant \Sigma \lambda_i (x, x_i)^2$ with equality for $x = x_1$. The conclusion is

$$\sup_{x, \|x\| = 1} (x, Ax) = \lambda_1$$

where λ_1 is the largest eigenvalue of A. This result also shows that $\lambda_1(A)$—the largest eigenvalue of the self-adjoint transformation A—is a convex function of A. In other words, if A_1 and A_2 are self-adjoint and $\alpha \in [0, 1]$, then $\lambda_1(\alpha A_1 + (1 - \alpha) A_2) \leqslant \alpha \lambda_1(A_1) + (1 - \alpha) \lambda_1(A_2)$. To prove this, first notice that for each $x \in V$, (x, Ax) is a linear, and hence convex, function of A. Since the supremum of a family of convex functions is a convex function, it follows that

$$\lambda_1(A) = \sup_{x,\|x\|=1} (x, Ax)$$

is a convex function defined on the real linear space of self-adjoint linear transformations. An interesting generalization of this is the following.

Proposition 1.44. Consider a self-adjoint transformation A defined on the n-dimensional inner product space $(V, (\cdot, \cdot))$ and let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the ordered eigenvalues of A. For $1 \leqslant k \leqslant n$, let $\mathfrak{B}_k$ be the collection of all k-tuples $\{v_1, \ldots, v_k\}$ such that $\{v_1, \ldots, v_k\}$ is an orthonormal set in $(V, (\cdot, \cdot))$. Then

$$\sup_{\{v_1,\ldots,v_k\}\in\mathfrak{B}_k} \sum_1^k (v_i, Av_i) = \sum_1^k \lambda_i.$$

Proof. Recall that $\langle \cdot , \cdot \rangle$ is the inner product on $\mathcal{L}(V, V)$ induced by the inner product $(\cdot, \cdot)$ on V, and $(x, Ax) = \langle x \square x, A \rangle$ for $x \in V$. Thus

$$\sum_1^k (v_i, Av_i) = \sum_1^k \langle v_i \square v_i, A \rangle = \left\langle \sum_1^k v_i \square v_i, A \right\rangle.$$

Write A in spectral form, $A = \Sigma_1^n \lambda_1 x_i \square x_i$. For $\{v_1, \ldots, v_k\} \in \mathfrak{B}_k$, $P_k = \Sigma_1^k v_i \square v_i$ is the orthogonal projection onto span$\{v_1, \ldots, v_k\}$. Thus for $\{v_1, \ldots, v_k\} \in \mathfrak{B}_k$,

$$\left\langle \sum_i^k v_i \square v_i, A \right\rangle = \left\langle P_k, \sum_1^n \lambda_i x_i \square x_i \right\rangle$$

$$= \sum_1^n \langle P_k, \lambda_i x_i \square x_i \rangle = \sum_1^n \lambda_i (x_i, P_k x_i).$$

Since P_k is an orthogonal projection and $\|x_i\| = 1$, $i = 1, \ldots, n$, $0 \leqslant (x_i, P_k x_i) \leqslant 1$. Also,

$$\sum_1^n (x_i, P_k x_i) = \sum_1^n \langle P_k, x_i \square x_i \rangle = \left\langle P_k, \sum_1^n x_i \square x_i \right\rangle = \langle P_k, I \rangle$$

because $\Sigma_1^n x_i \square x_i = I \in \mathcal{L}(V, V)$. But $P_k = \Sigma_1^k v_i \square v_i$, so

$$\langle P_k, I \rangle = \sum_1^k \langle v_i \square v_i, I \rangle = \sum_1^k (v_i, v_i) = k.$$

Therefore, the real numbers $\alpha_i = (x_i, P_k x_i)$, $i = 1,\dots, n$, satisfy $0 \leqslant \alpha_i \leqslant 1$ and $\Sigma_1^n \alpha_i = k$. A moment's reflection shows that, for any numbers $\alpha_1,\dots, \alpha_n$ satisfying these conditions, we have

$$\sum_1^n \lambda_i \alpha_i \leqslant \sum_1^k \lambda_i$$

since $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. Therefore,

$$\left\langle \sum_1^k v_i \square v_i, A \right\rangle \leqslant \sum_1^k \lambda_i$$

for $\{v_1,\dots, v_k\} \in \mathfrak{B}_k$. However, setting $v_i = x_i$, $i = 1,\dots, k$, yields equality in the above inequality. □

For $A \in \mathfrak{L}(V, V)$, which is self-adjoint, define $\text{tr}_k A = \Sigma_1^k \lambda_i$ where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ are the ordered eigenvalues of A. The symbol $\text{tr}_k A$ is read "trace sub-k of A." Since $\langle \Sigma_1^k v_i \square v_i, A\rangle$ is a linear function of A and $\text{tr}_k A$ is the supremum over all $\{v_1,\dots, v_k\} \in \mathfrak{B}_k$, it follows that $\text{tr}_k A$ is a convex function of A. Of course, when $k = n$, $\text{tr}_k A$ is just the trace of A.

For completeness, a statement of the spectral theorem for $n \times n$ symmetric matrices is in order.

Proposition 1.45. Suppose A is an $n \times n$ real symmetric matrix. Then there exists an $n \times n$ orthogonal matrix Γ and an $n \times n$ diagonal matrix D such that $A = \Gamma D \Gamma'$. The columns of Γ are the eigenvectors of A and the diagonal elements of D, say $\lambda_1,\dots, \lambda_n$, are the eigenvalues of A.

Proof. This is nothing more than a disguised version of the spectral theorem. To see this, write

$$A = \sum_1^n \lambda_i x_i x_i'$$

where $x_i \in R^n$, $\lambda_i \in R$, and $\{x_1,\dots, x_n\}$ is an orthonormal basis for R^n with the usual inner product (here $x_i \square x_i$ is $x_i x_i'$ since we have the usual inner product on R^n). Let Γ have columns $x_1,\dots, x_n$ and let D have diagonal elements $\lambda_1,\dots, \lambda_n$. Then a straightforward computation shows that

$$\sum_1^n \lambda_i x_i x_i' = \Gamma D \Gamma'.$$

The remaining assertions follow immediately from the spectral theorem. □

Our final application of the spectral theorem in this chapter deals with a representation theorem for a linear transformation $A \in \mathcal{L}(V, W)$ where $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$ are finite dimensional inner product spaces. In this context, eigenvalues and eigenvectors of A make no sense, but something can be salvaged by considering $A'A \in \mathcal{L}(V, V)$. First, $A'A$ is non-negative definite and $\mathfrak{N}(A'A) = \mathfrak{N}(A)$. Let $k = \text{rank}(A) = \text{rank}(A'A)$ and let $\lambda_1 \geqslant \cdots \geqslant \lambda_k > 0$ be the nonzero eigenvalues of $A'A$. There must be exactly k positive eigenvalues of $A'A$ as $\text{rank}(A) = k$. The spectral theorem shows that

$$A'A = \sum_{1}^{k} \lambda_i x_i \,\square\, x_i$$

where $\{x_1, \ldots, x_n\}$ is an orthonormal basis for V and $A'Ax_i = \lambda_i x_i$ for $i = 1, \ldots, k$, $A'Ax_i = 0$ for $i = k + 1, \ldots, n$. Therefore, $\mathfrak{N}(A) = \mathfrak{N}(A'A) = (\text{span}\{x_1, \ldots, x_k\})^{\perp}$.

Proposition 1.46. In the notation above, let $w_i = (1/\sqrt{\lambda_i})Ax_i$ for $i = 1, \ldots, k$. Then $\{w_1, \ldots, w_k\}$ is an orthonormal basis for $\mathfrak{R}(A) \subseteq W$ and $A = \sum_1^k \sqrt{\lambda_i}\, w_i \,\square\, x_i$.

Proof. Since $\dim \mathfrak{R}(A) = k$, $\{w_1, \ldots, w_k\}$ is a basis for $\mathfrak{R}(A)$ if $\{w_1, \ldots, w_k\}$ is an orthonormal set. But

$$\begin{aligned} [w_i, w_j] &= (\lambda_i \lambda_j)^{-1/2}[Ax_i, Ax_j] = (\lambda_i \lambda_j)^{-1/2}(x_i, A'Ax_j) \\ &= (\lambda_i \lambda_j)^{-1/2}\lambda_j (x_i, x_j) = \delta_{ij} \end{aligned}$$

and the first assertion holds. To show $A = \sum_1^k \sqrt{\lambda_i}\, w_i \,\square\, x_i$, we verify the two linear transformations agree on the basis $\{x_1, \ldots, x_n\}$. For $1 \leqslant j \leqslant k$, $Ax_j = \sqrt{\lambda_j}\, w_j$ by definition and

$$\left(\sum_1^k \sqrt{\lambda_i}\, w_i \,\square\, x_i \right) x_j = \sum_1^k \sqrt{\lambda_i}\, (x_i, x_j) w_i = \sqrt{\lambda_j}\, w_j.$$

For $k + 1 \leqslant j \leqslant n$, $Ax_j = 0$ since $\mathfrak{N}(A) = (\text{span}\{x_1, \ldots, x_k\})^{\perp}$. Also

$$\left(\sum_1^k \sqrt{\lambda_i}\, w_i \,\square\, x_i \right) x_j = \sum_1^k \sqrt{\lambda_i}\, (x_i, x_j) w_i = 0$$

as $j > k$. $\square$

Some immediate consequences of the above representation are (i) $AA' = \sum_1^k \lambda_i w_i \square w_i$, (ii) $A' = \sum_1^k \sqrt{\lambda_i}\, x_i \square w_i$ and $A'w_i = \sqrt{\lambda_i}\, x_i$ for $i = 1, \ldots, k$. In summary, we have the following.

Theorem 1.3 (Singular Value Decomposition Theorem). Given $A \in \mathcal{L}(V, W)$ of rank k, there exist orthonormal vectors $x_1, \ldots, x_k$ in V and $w_1, \ldots, w_k$ in W and positive numbers $\mu_1, \ldots, \mu_k$ such that

$$A = \sum_1^k \mu_i w_i \square x_i.$$

Also, $\mathcal{R}(A) = \text{span}\{w_1, \ldots, w_k\}$, $\mathcal{N}(A) = (\text{span}\{x_1, \ldots, x_k\})^\perp$, $Ax_i = \mu_i w_i$, $i = 1, \ldots, k$, $A' = \sum_1^k \mu_i x_i \square w_i$, $A'A = \sum_1^k \mu_i^2 x_i \square x_i$, $AA' = \sum_1^k \mu_i^2 w_i \square w_i$. The numbers $\mu_1^2, \ldots, \mu_k^2$ are the positive eigenvalues of both AA' and $A'A$.

For matrices, this result takes the following form.

Proposition 1.47. If A is a real $n \times m$ matrix of rank k, then there exist matrices $\Gamma : n \times k$, $D : k \times k$, and $\Psi : k \times m$ that satisfy $\Gamma'\Gamma = I_k$, $\Psi\Psi' = I_k$, D is a diagonal matrix with positive diagonal elements, and

$$A = \Gamma D \Psi.$$

Proof. Take $V = R^m$, $W = R^n$ and apply Theorem 1.3 to get

$$A = \sum_1^k \mu_i w_i x_i'$$

where $x_1, \ldots, x_k$ are orthonormal in R^m, $w_1, \ldots, w_k$ are orthonormal in R^n, and $\mu_i > 0$, $i = 1, \ldots, k$. Let Γ have columns $w_1, \ldots, w_k$, let Ψ have rows $x_1', \ldots, x_k'$, and let D be diagonal with diagonal elements $\mu_1, \ldots, \mu_k$. An easy calculation shows that

$$\sum_1^k \mu_i w_i x_i' = \Gamma D \Psi. \qquad \square$$

In the case that $A \in \mathcal{L}(V, V)$ with rank k, Theorem 1.3 shows that there exist orthonormal sets $\{x_1, \ldots, x_k\}$ and $\{w_1, \ldots, w_k\}$ of V such that

$$A = \sum_1^k \mu_i w_i \square x_i$$

where $\mu_i > 0$, $i = 1, \ldots, k$. Also, $\mathcal{R}(A) = \text{span}\{w_1, \ldots, w_k\}$ and $\mathcal{N}(A) =$

$(\text{span}\{x_1,\ldots,x_k\})^\perp$. Now, consider two subspaces M_1 and M_2 of the inner product space $(V,(\cdot,\cdot))$ and let P_1 and P_2 be the orthogonal projections onto M_1 and M_2. In what follows, the geometrical relationship between the two subspaces (measured in terms of angles, which are defined below) is related to the singular value decomposition of the linear transformation $P_1P_2 \in \mathcal{L}(V,V)$. It is clear that $\mathcal{R}(P_2P_1) \subseteq M_2$ and $\mathcal{N}(P_2P_1) \supseteq M_1^\perp$. Let $k = \text{rank}(P_2P_1)$ so $k \leqslant \dim(M_i)$, $i = 1,2$. Theorem 1.3 implies that

$$P_2P_1 = \sum_1^k \mu_i w_i \,\square\, x_i$$

where $\mu_i > 0$, $i = 1,\ldots,k$, $\mathcal{R}(P_2P_1) = \text{span}\{w_1,\ldots,w_k\} \subseteq M_2$, and $(\mathcal{N}(P_2P_1))^\perp = \text{span}\{x_1,\ldots,x_k\} \subseteq M_1$. Also, $\{w_1,\ldots,w_k\}$ and $\{x_1,\ldots,x_k\}$ are orthonormal sets. Since $P_2P_1x_j = \mu_j w_j$ and $(P_2P_1)'P_2P_1 = P_1P_2P_2P_1 = P_1P_2P_1 = \sum_1^k \mu_i^2 x_i \,\square\, x_i$, we have

$$\mu_j(x_i, w_j) = \left(x_i, P_2P_1x_j\right) = \left(P_1x_i, P_2P_1x_j\right) = \left(x_i, P_1P_2P_1x_j\right)$$

$$= \left(x_i, \left(\sum_1^k \mu_l^2 x_l \,\square\, x_l\right)x_j\right) = \mu_j^2(x_i, x_j) = \delta_{ij}\mu_j^2.$$

Therefore, for $i, j = 1,\ldots,k$,

$$(x_i, w_j) = \delta_{ij}\mu_j$$

since $\mu_j > 0$. Furthermore, if $x \in M_1 \cap (\text{span}\{x_1,\ldots,x_k\})^\perp$ and $w \in M_2$, then $(x,w) = (P_1x, P_2w) = (P_2P_1x, w) = 0$ since $P_2P_1x = 0$. Similarly, if $w \in M_2 \cap (\text{span}\{w_1,\ldots,w_k\})^\perp$ and $x \in M_1$, then $(x,w) = 0$.

The above discussion yields the following proposition.

Proposition 1.48. Suppose M_1 and M_2 are subspaces of $(V,(\cdot,\cdot))$ and let P_1 and P_2 be the orthogonal projections onto M_1 and M_2. If $k = \text{rank}(P_2P_1)$, then there exist orthonormal sets $\{x_1,\ldots,x_k\} \subseteq M_1$, $\{w_1,\ldots,w_k\} \subseteq M_2$ and positive numbers $\mu_1 \geqslant \cdots \geqslant \mu_k$ such that:

(i) $P_2P_1 = \sum_1^k \mu_i w_i \,\square\, x_i$.
(ii) $P_1P_2P_1 = \sum_1^k \mu_i^2 x_i \,\square\, x_i$.
(iii) $P_2P_1P_2 = \sum_1^k \mu_i^2 w_i \,\square\, w_i$.
(iv) $0 < \mu_j \leqslant 1$ and $(x_i, w_j) = \delta_{ij}\mu_j$ for $i, j = 1,\ldots,k$.
(v) If $x \in M_1 \cap (\text{span}\{x_1,\ldots,x_k\})^\perp$ and $w \in M_2$, then $(x,w) = 0$. If $w \in M_2 \cap (\text{span}\{w_1,\ldots,w_k\})^\perp$ and $x \in M_1$, then $(x,w) = 0$.

Proof. Assertions (i), (ii), (iii), and (v) have been verified as has the relationship $(x_i, w_j) = \delta_{ij}\mu_j$. Since $0 < \mu_j = (x_j, w_j)$, the Cauchy–Schwarz Inequality yields $(x_j, w_j) \leqslant \|x_j\|\,\|w_j\| = 1$. □

In Proposition 1.48, if $k = \operatorname{rank} P_2 P_1 = 0$, then M_1 and M_2 are orthogonal to each other and $P_1 P_2 = P_2 P_1 = 0$. The next result provides the framework in which to relate the numbers $\mu_1 \geqslant \cdots \geqslant \mu_k$ to angles.

Proposition 1.49. In the notation of Proposition 1.48, let $M_{11} = M_1$, $M_{21} = M_2$,

$$M_{1i} = (\operatorname{span}\{x_1, \dots, x_{i-1}\})^{\perp} \cap M_1,$$

and

$$M_{2i} = (\operatorname{span}\{w_1, \dots, w_{i-1}\})^{\perp} \cap M_2$$

for $i = 2, \dots, k+1$. Also, for $i = 1, \dots, k$, let

$$D_{1i} = \{x | x \in M_{1i}, \|x\| = 1\}$$

and

$$D_{2i} = \{w | w \in M_{2i}, \|w\| = 1\}.$$

Then

$$\sup_{x \in D_{1i}} \sup_{w \in D_{2i}} (x, w) = (x_i, w_i) = \mu_i$$

for $i = 1, \dots, k$. Also, $M_{1(k+1)} \perp M_2$ and $M_{2(k+1)} \perp M_1$.

Proof. Since $x_i \in D_{1i}$ and $w_i \in D_{2i}$, the iterated supremum is at least (x_i, w_i) and $(x_i, w_i) = \mu_i$ by (iv) of Proposition 1.48. Thus it suffices to show that for each $x \in D_{1i}$ and $w \in D_{2i}$, we have the inequality $(x, w) \leqslant \mu_i$. However, for $x \in D_{1i}$ and $w \in D_{2i}$,

$$(x, w) = (P_1 x, P_2 w) = (P_2 P_1 x, w) \leqslant \|P_2 P_1 x\|\,\|w\| = \|P_2 P_1 x\|$$

since $\|w\| = 1$ as $w \in D_{1i}$. Thus

$$(x, w) \leqslant \|P_2 P_1 x\| = (P_2 P_1 x, P_2 P_1 x)^{1/2} = (P_1 P_2 P_1 x, x)^{1/2}$$

$$= \left[\sum_{j=1}^{k} \mu_j^2 \big((x_j \square x_j) x, x\big)\right]^{1/2} = \left[\sum_{j=1}^{k} \mu_j^2 (x_j, x)^2\right]^{1/2}.$$

Since $x \in D_{1i}$, $(x, x_j) = 0$ for $j = 1, \dots, i-1$. Also, the numbers $a_j \equiv$

$(x_j, x)^2$ satisfy $0 \leqslant a_j \leqslant 1$ and $\Sigma_1^k a_j \leqslant 1$ as $\|x\| = 1$. Therefore,

$$(x, w) \leqslant \left[\sum_{j=1}^{k} \mu_j^2 (x_j, x)^2\right]^{1/2} = \left[\sum_{j=i}^{k} a_j \mu_j^2\right]^{1/2} \leqslant \left(\mu_i^2\right)^{1/2} = \mu_i.$$

The last inequality follows from the fact that $\mu_1 \geqslant \cdots \geqslant \mu_k > 0$ and the conditions on the a_j's. Hence,

$$\sup_{x \in D_{1i}} \sup_{w \in D_{2i}} (x, w) = (x_i, w_i) = \mu_i,$$

and the first assertion holds. The second assertion is simply a restatement of (v) of Proposition 1.48. □

Definition 1.28. Let M_1 and M_2 be subspaces of $(V, (\cdot, \cdot))$. Given the numbers $\mu_1 \geqslant \cdots \geqslant \mu_k > 0$, whose existence is guaranteed by Proposition 1.48, define $\theta_i \in [0, \pi/2)$ by

$$\cos \theta_i = \mu_i, \qquad i = 1, \dots, k.$$

Let $t = \min\{\dim M_1, \dim M_2\}$ and set $\theta_i = \pi/2$ for $i = k+1, \dots, t$. The numbers $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_t$ are called the *ordered angles* between M_1 and M_2.

The following discussion is intended to provide motivation, explanation, and a geometric interpretation of the above definition. Recall that if y_1 and y_2 are two vectors in $(V, (\cdot, \cdot))$ of length 1, then the cosine of the angle between y_1 and y_2 is defined by $\cos \theta = (y_1, y_2)$ where $0 \leqslant \theta \leqslant \pi$. However, if we want to define the angle between the two lines span$\{y_1\}$ and span$\{y_2\}$, then a choice must be made between two angles that are complements of each other. The convention adopted here is to choose the angle in $[0, \pi/2]$. Thus the cosine of the angle between span$\{y_1\}$ and span$\{y_2\}$ is just $|(y_1, y_2)|$. To show this agrees with the definition above, we have $M_i = \text{span}\{y_i\}$ and $P_i = y_i \square y_i$ is the orthogonal projection onto M_i, $i = 1, 2$. The rank of $P_2 P_1$ is either zero or one and the rank is zero iff $y_1 \perp y_2$. If $y_1 \perp y_2$, then the angle between M_1 and M_2 is $\pi/2$, which agrees with Definition 1.28. When the rank of $P_2 P_1$ is one, $P_1 P_2 P_1 = (y_1, y_2)^2 y_1 \square y_1$ whose only nonzero eigenvalue is $(y_1, y_2)^2$. Thus $\mu_1^2 = (y_1, y_2)^2$ so $\mu_1 = |(y_1, y_2)| = \cos \theta_1$, and again we have agreement with Definition 1.28.

Now consider the case when $M_1 = \text{span}\{y_1\}$, $\|y_1\| = 1$, and M_2 is an arbitrary subspace of $(V, (\cdot, \cdot))$. Geometrically, it is clear that the angle between M_1 and M_2 is just the angle between M_1 and the orthogonal projection of M_1 onto M_2, say $M_2^* = \text{span}\{P_2 y_1\}$ where P_2 is the orthogonal projection onto M_2. Thus the cosine of the angle between M_1 and M_2 is

$$\cos\theta = \left|\left(y_1, \frac{P_2 y_1}{\|P_2 y_1\|}\right)\right| = \|P_2 y_1\|.$$

If $P_2 y_1 = 0$, then $M_1 \perp M_2$ and $\cos\theta = 0$ so $\theta = \pi/2$ in agreement with Definition 1.28. When $P_2 y_1 \neq 0$, then $P_1 P_2 P_1 = (y_1, P_2 y_1) y_1 \square y_1$, whose only nonzero eigenvalue is $(y_1, P_2 y_1) = (P_2 y_1, P_2 y_1) = \|P_2 y_1\|^2 = \mu_1^2$. Therefore, $\mu_1 = \|P_2 y_1\|$ and again we have agreement with Definition 1.28.

In the general case when $\dim(M_i) > 1$ for $i = 1, 2$, it is not entirely clear how we should define the angles between M_1 and M_2. However, the following considerations should provide some justification for Definition 1.28. First, if $x \in M_1$ and $w \in M_2$, $\|x\| = \|w\| = 1$. The cosine of the angle between span x and span w is $|(x, w)|$. Thus the largest cosine of any angle (equivalently, the smallest angle in $[0, \pi/2]$) between a one-dimensional subspace of M_1 and a one-dimensional subspace of M_2 is

$$\sup_{x \in D_{11}} \sup_{w \in D_{21}} |(x, w)| = \sup_{x \in D_{11}} \sup_{w \in D_{21}} (x, w).$$

The sets D_{11} and D_{21} are defined in Proposition 1.49. By Proposition 1.49, this iterated supremum is μ_1 and is achieved for $x = x_1 \in D_{11}$ and $w = w_1 \in D_{21}$. Thus the cosine of the angle between $\text{span}\{x_1\}$ and $\text{span}\{w_1\}$ is μ_1. Now, remove $\text{span}\{x_1\}$ from M_1 to get $M_{12} = (\text{span}\{x_1\})^\perp \cap M_1$ and remove $\text{span}\{w_1\}$ from M_2 to get $M_{22} = (\text{span}\{w_1\})^\perp \cap M_2$. The second largest cosine of any angle between M_1 and M_2 is defined to be the largest cosine of any angle between M_{12} and M_{22} and is given by

$$\sup_{x \in D_{12}} \sup_{w \in D_{22}} (x, w) = (x_2, w_2) = \mu_2.$$

Next $\text{span}\{x_2\}$ is removed from M_{12} and $\text{span}\{w_2\}$ is removed from M_{22}, yielding M_{13} and M_{23}. The third largest cosine of any angle between M_1 and M_2 is defined to be the largest cosine of any angle between M_{13} and M_{23}, and so on. After k steps, we are left with $M_{1(k+1)}$ and $M_{2(k+1)}$, which are orthogonal to each other. Thus the remaining angles are $\pi/2$. The above is precisely the content of Definition 1.28, given the results of Propositions 1.48 and 1.49.

The statistical interpretation of the angles between subspaces is given in a later chapter. In a statistical context, the cosines of these angles are called canonical correlation coefficients and are a measure of the affine dependence between the random vectors.

PROBLEMS

All vector spaces are finite dimensional unless specified otherwise.

1. Let V_{n+1} be the set of all nth degree polynomials (in the real variable t) with real coefficients. With the usual definition of addition and scalar multiplication, prove that V_{n+1} is an $(n+1)$-dimensional real vector space.

2. For $A \in \mathcal{L}(V, W)$, suppose that M is any subspace of V such that $M \oplus \mathfrak{N}(A) = V$.
   (i) Show that $\mathfrak{R}(A) = A(M)$ where $A(M) = \{w | w = Ax$ for some $x \in M\}$.
   (ii) If $x_1, \ldots, x_k$ is any linearly independent set in V such that $\text{span}\{x_1, \ldots, x_k\} \cap \mathfrak{N}(A) = \{0\}$, prove that $Ax_1, \ldots, Ax_k$ is linearly independent.

3. For $A \in \mathcal{L}(V, W)$, fix $w_0 \in W$ and consider the linear equation $Ax = w_0$. If $w_0 \notin \mathfrak{R}(A)$, there is no solution to this equation. If $w_0 \in \mathfrak{R}(A)$, let x_0 be any solution so $Ax_0 = w_0$. Prove that $\mathfrak{N}(A) + x_0$ is the set of all solutions to $Ax = w_0$.

4. For the direct sum space $V_1 \oplus V_2$, suppose $A_{ij} \in \mathcal{L}(V_j, V_i)$ and let

$$T_1 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

be defined by

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \{v_1, v_2\} = \{A_{11}v_1 + A_{12}v_2, A_{21}v_1 + A_{22}v_2\}$$

for $\{v_1, v_2\} \in V_1 \oplus V_2$.
   (i) Prove that T_1 is a linear transformation.
   (ii) Conversely, prove that every $T_1 \in \mathcal{L}(V_1 \oplus V_2, V_1 \oplus V_2)$ has such a representation.
   (iii) If

$$T = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

and

$$U = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

prove that the representation of TU is

$$\begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

5. Let $x_1, \ldots, x_r, x_{r+1}$ be vectors in V with $x_1, \ldots, x_r$ being linearly independent. For $w_1, \ldots, w_r, w_{r+1}$ in W, give a necessary and sufficient condition for the existence of an $A \in \mathcal{L}(V, W)$ that satisfies $Ax_i = w_i$, $i = 1, \ldots, r + 1$.

6. Suppose $A \in \mathcal{L}(V, V)$ satisfies $A^2 = cA$ where $c \neq 0$. Find a constant k so that $B = kA$ is a projection.

7. Suppose A is an $m \times n$ matrix with columns $a_1, \ldots, a_n$ and B is an $n \times k$ matrix with rows $b'_1, \ldots, b'_n$. Show that $AB = \Sigma_1^n a_i b'_i$.

8. Let $x_1, \ldots, x_k$ be vectors in R^n, set $M = \text{span}\{x_1, \ldots, x_k\}$, and let A be the $n \times k$ matrix with columns $x_1, \ldots, x_k$ so $A \in \mathcal{L}(R^k, R^n)$.

(i) Show $M = \mathcal{R}(A)$.

(ii) Show $\dim(M) = \text{rank}(A'A)$.

9. For linearly independent $x_1, \ldots, x_k$ in $(V, (\cdot, \cdot))$, let $y_1, \ldots, y_k$ be the vectors obtained by applying the Gram–Schmidt (G-S) Process to $x_1, \ldots, x_k$. Show that if $z_i = Ax_i$, $i = 1, \ldots, k$, where $A \in \mathcal{O}(V)$, then the vectors obtained by the G-S Process from $z_1, \ldots, z_k$ are $Ay_1, \ldots, Ay_k$. (In other words, the G-S Process commutes with orthogonal transformations.)

10. In $(V, (\cdot, \cdot))$, let $x_1, \ldots, x_k$ be vectors with $x_1 \neq 0$. Form $y_1^1, \ldots, y_k^1$ by $y_1^1 = x_1/\|x_1\|$ and $y_i^1 = x_i - (x_i, y_1^1)y_1^1$, $i = 2, \ldots, k$:

(i) Show $\text{span}\{x_1, \ldots, x_r\} = \text{span}\{y_1^1, \ldots, y_r^1\}$ for $r = 1, 2, \ldots, k$.

(ii) Show $y_1^1 \perp \text{span}\{y_2^1, \ldots, y_k^1\}$ so $\text{span}\{y_1^1, \ldots, y_r^1\} = \text{span}\{y_1^1\} \oplus \text{span}\{y_2^1, \ldots, y_r^1\}$ for $r = 2, \ldots, k$.

(iii) Now, form $y_2^2, \ldots, y_k^2$ from $y_2^1, \ldots, y_k^1$ as the y^1's were formed from the x's (reordering if necessary to achieve $y_2^1 \neq 0$). Show $\text{span}\{x_1, \ldots, x_k\} = \text{span}\{y_1^1\} \oplus \text{span}\{y_2^2\} \oplus \text{span}\{y_3^2, \ldots, y_k^2\}$.

(iv) Let $m = \dim(\text{span}\{x_1, \ldots, x_k\})$. Show that after applying the above procedure m times, we get an orthonormal basis $y_1^1, y_2^2, \ldots, y_m^m$ for $\text{span}\{x_1, \ldots, x_k\}$.

(v) If $x_1, \dots, x_k$ are linearly independent, show that $\text{span}\{x_1, \dots, x_r\} = \text{span}\{y_1^1, y_2^2, \dots, y_r^r\}$ for $r = 1, \dots, k$.

11. Let $x_1, \dots, x_m$ be a basis for $(V, (\cdot, \cdot))$ and $w_1, \dots, w_n$ be a basis for $(W, [\cdot, \cdot])$. For $A, B \in \mathcal{L}(V, W)$, show that $[Ax_i, w_j] = [Bx_i, w_j]$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ implies that $A = B$.

12. For $x_i \in (V, (\cdot, \cdot))$ and $y_i \in (W, [\cdot, \cdot])$, $i = 1, 2$, suppose that $x_1 \square y_1 = x_2 \square y_2 \neq 0$. Prove that $x_1 = cx_2$ for some scalar $c \neq 0$ and then $y_1 = c^{-1}y_2$.

13. Given two inner products on V, say $(\cdot, \cdot)$ and $[\cdot, \cdot]$, show that there exist positive constants c_1 and c_2 such that $c_1[x, x] \leqslant (x, x) \leqslant c_2[x, x]$, $x \in V$. Using this, show that for any open ball in $(V, (\cdot, \cdot))$, say $B = \{x | (x, x)^{1/2} < \alpha\}$, there exist open balls in $(V, [\cdot, \cdot])$, say $B_i = \{x | [x, x]^{1/2} < \beta_i\}$, $i = 1, 2$, such that $B_1 \subseteq B \subseteq B_2$.

14. In $(V, (\cdot, \cdot))$, prove that $\|x + y\| \leqslant \|x\| + \|y\|$. Using this, prove that $h(x) = \|x\|$ is a convex function.

15. For positive integers I and J, consider the IJ-dimensional real vector space, V, of all real-valued functions defined on $\{1, 2, \dots, I\} \times \{1, 2, \dots, J\}$. Denote the value of $y \in V$ at (i, j) by y_{ij}. The inner product on V is taken to be $(y, \tilde{y}) = \Sigma\Sigma y_{ij}\tilde{y}_{ij}$. The symbol $1 \in V$ denotes the vector all of whose coordinates are one.

(i) Define A on V to V by $Ay = \bar{y}_{..}1$ where $\bar{y}_{..} = (IJ)^{-1}\Sigma\Sigma y_{ij}$. Show that A is the orthogonal projection onto span$\{1\}$.

(ii) Define linear transformations B_1, B_2, and B_3 on V by

$$(B_1 y)_{ij} = \bar{y}_{i.} - \bar{y}_{..}$$

$$(B_2 y)_{ij} = \bar{y}_{.j} - \bar{y}_{..}$$

$$(B_3 y)_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

where

$$\bar{y}_{i.} = J^{-1}\sum_j y_{ij}$$

and

$$\bar{y}_{.j} = I^{-1}\sum_i y_{ij}.$$

Show that B_1, B_2, and B_3 are orthogonal projections and the following holds:

$$AB_k = 0, \qquad k = 1, 2, 3$$

$$B_1B_2 = B_1B_3 = B_2B_3 = 0$$

$$(A + B_1 + B_2 + B_3)y = y, \qquad y \in V.$$

(iii) Show that

$$\|y\|^2 = \|Ay\|^2 + \|B_1y\|^2 + \|B_2y\|^2 + \|B_3y\|^2.$$

16. For $\Gamma \in \mathcal{O}(V)$ and M a subspace of V, suppose that $\Gamma(M) \subseteq M$. Prove that $\Gamma(M^\perp) \subseteq M^\perp$.

17. Given a subspace M of $(V, (\cdot, \cdot))$, show the following are equivalent:
(i) $|(x, y)| \leqslant c\|x\|$ for all $x \in M$.
(ii) $\|P_My\| \leqslant c$.
Here c is a fixed positive constant and P_M is the orthogonal projection onto M.

18. In $(V, (\cdot, \cdot))$, suppose A and B are positive semidefinite. For $C, D \in \mathcal{L}(V, V)$ prove that $(\operatorname{tr} ACBD')^2 \leqslant \operatorname{tr} ACBC' \operatorname{tr} ADBD'$.

19. Show that $\mathbb{C}^n$ is a $2n$-dimensional real vector space.

20. Let A be an $n \times n$ real matrix. Prove:
(i) If λ_0 is a real eigenvalue of A, then there exists a corresponding real eigenvector.
(ii) If λ_0 is an eigenvalue that is not real, then any corresponding eigenvector cannot be real or pure imaginary.

21. In an n-dimensional space $(V, (\cdot, \cdot))$, suppose P is a rank r orthogonal projection. For $\alpha, \beta \in R$, let $A = \alpha P + \beta(I - P)$. Find eigenvalues, eigenvectors, and the characteristic polynomial of A. Show that A is positive definite iff $\alpha > 0$ and $\beta > 0$. What is A^{-1} when it exists?

22. Suppose A and B are self-adjoint and $A - B \geqslant 0$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and $\mu_1 \geqslant \cdots \geqslant \mu_n$ be the eigenvalues of A and B. Show that $\lambda_i \geqslant \mu_i$, $i = 1, \ldots, n$.

23. If $S, T \in \mathcal{L}(V, V)$ and $S > 0$, $T \geqslant 0$, prove that $\langle S, T \rangle = 0$ implies $T = 0$.

24. For $A \in (\mathcal{L}(V, V), \langle \cdot, \cdot \rangle)$, show that $\langle A, I \rangle = \operatorname{tr} A$.

25. Suppose A and B in $\mathcal{L}(V, V)$ are self-adjoint and write $A \geqslant B$ to mean $A - B \geqslant 0$.

(i) If $A \geqslant B$, show that $CAC' \geqslant CBC'$ for all $C \in \mathcal{L}(V, V)$.

(ii) Show $I \geqslant A$ iff all the eigenvalues of A are less than or equal to one.

(iii) Assume $A > 0$, $B > 0$, and $A \geqslant B$. Is $A^{1/2} \geqslant B^{1/2}$? Is $A^2 \geqslant B^2$?

26. If P is an orthogonal projection, show that $\operatorname{tr} P$ is the rank of P.

27. Let $x_1, \ldots, x_n$ be an orthonormal basis for $(V, (\cdot, \cdot))$ and consider the vector space $(\mathcal{L}(V, V), \langle \cdot , \cdot \rangle)$. Let M be the subspace of $\mathcal{L}(V, V)$ consisting of all self-adjoint linear transformations and let N be the subspace of all skew symmetric linear transformations. Prove:

(i) $\{x_i \square x_j + x_j \square x_i | i \leqslant j\}$ is an orthogonal basis for M.

(ii) $\{x_i \square x_j - x_j \square x_i | i < j\}$ is an orthogonal basis for N.

(iii) M is orthogonal to N and $M \oplus N = \mathcal{L}(V, V)$.

(iv) The orthogonal projection onto M is $A \to (A + A')/2$, $A \in \mathcal{L}(V, V)$.

28. Consider $\mathcal{L}_{n,n}$ with the usual inner product $\langle A, B \rangle = \operatorname{tr} AB'$, and let $\mathcal{S}_n$ be the subspace of symmetric matrices. Then $(\mathcal{S}_n, \langle \cdot , \cdot \rangle)$ is an inner product space. Show $\dim \mathcal{S}_n = n(n+1)/2$ and for $S, T \in \mathcal{S}_n$, $\langle S, T \rangle = \Sigma_i s_{ii} t_{ii} + 2\Sigma\Sigma_{i<j} s_{ij} t_{ij}$.

29. For $A \in \mathcal{L}(V, W)$, one definition of the norm of A is

$$|||A||| = \sup_{\|v\|=1} \|Av\|$$

where $\| \cdot \|$ is the given norm on W.

(i) Show that $|||A|||$ is the square root of the largest eigenvalue of $A'A$.

(ii) Show that $|||\alpha A||| = |\alpha| \, |||A|||$, $\alpha \in R$ and $|||A + B||| \leqslant |||A||| + |||B|||$.

30. In the inner product spaces $(V, (\cdot, \cdot))$ and $(W, [\cdot, \cdot])$, consider $A \in \mathcal{L}(V, V)$ and $B \in \mathcal{L}(W, W)$, which are both self-adjoint. Write these in spectral form as

$$A = \sum_1^m \lambda_i x_i \square x_i$$

$$B = \sum_1^n \mu_j w_j \square w_j.$$

(Note: The symbol $\square$ has a different meaning in these two equations since the definition of $\square$ depends on the inner product.) Of course, $x_1,\ldots, x_m [w_1,\ldots, w_n]$ is an orthonormal basis for $(V,(\cdot,\cdot))[(W,[\cdot,\cdot])]$. Also, $\{x_i \square w_j | i = 1,\ldots, m,\ j = 1,\ldots, n\}$ is an orthonormal basis for $(\mathcal{L}(W,V), \langle \cdot , \cdot \rangle)$, and $A \otimes B$ is a linear transformation on $\mathcal{L}(W,V)$ to $\mathcal{L}(W,V)$.

(i) Show that $(A \otimes B)(x_i \square w_j) = \lambda_i \mu_j (x_i \square w_j)$ so $\lambda_i \mu_j$ is an eigenvalue of $A \otimes B$.

(ii) Show that $A \otimes B = \Sigma\Sigma\lambda_i \mu_j (x_i \square w_j) \tilde{\square} (x_i \square w_j)$ and this is a spectral decomposition for $A \otimes B$. What are the eigenvalues and corresponding eigenvectors for $A \otimes B$?

(iii) If A and B are positive definite (semidefinite), show that $A \otimes B$ is positive definite (semidefinite).

(iv) Show that $\operatorname{tr} A \otimes B = (\operatorname{tr} A)(\operatorname{tr} B)$ and $\det A \otimes B = (\det A)^n (\det B)^m$.

31. Let $x_1,\ldots, x_p$ be linearly independent vectors in R^n, set $M = \operatorname{span}\{x_1,\ldots, x_p\}$, and let $A : n \times p$ have columns $x_1,\ldots, x_p$. Thus $\mathcal{R}(A) = M$ and $A'A$ is positive definite.

(i) Show that $\psi = A(A'A)^{-1/2}$ is a linear isometry whose columns form an orthonormal basis for M. Here, $(A'A)^{-1/2}$ denotes the inverse of the positive definite square root of $A'A$.

(ii) Show that $\psi\psi' = A(A'A)^{-1}A'$ is the orthogonal projection on to M.

32. Consider two subspaces, M_1 and M_2, of R^n with bases $x_1,\ldots, x_q$ and $y_1,\ldots, y_r$. Let $A(B)$ have columns $x_1,\ldots,x_q$ $(y_1,\ldots, y_r)$. Then $P_1 = A(A'A)^{-1}A'$ and $P_2 = B(B'B)^{-1}B'$ are the orthogonal projections onto M_1 and M_2, respectively. The cosines of the angles between M_1 and M_2 can be obtained by computing the nonzero eigenvalues of $P_1 P_2 P_1$. Show that these are the same as the nonzero eigenvalues of

$$(A'A)^{-1}A'B(B'B)^{-1}B'A : q \times q$$

and of

$$(B'B)^{-1}B'A(A'A)^{-1}A'B : r \times r.$$

33. In R^4, set $x_1' = (1,0,0,0)$, $x_2' = (0,1,0,0)$, $y_1' = (1,1,1,1)$, and $y_2' = (1,-1,1,-1)$. Find the cosines of the angles between $M_1 = \operatorname{span}\{x_1, x_2\}$ and $M_2 = \operatorname{span}\{y_1, y_2\}$.

34. For two subspaces M_1 and M_2 of $(V,(\cdot,\cdot))$, argue that the angles between M_1 and M_2 are the same as the angles between $\Gamma(M_1)$ and $\Gamma(M_2)$ for any $\Gamma \in \mathcal{O}(V)$.

35. This problem has to do with the vector space V of Example 1.9 and V may be infinite dimensional. The results in this problem are not used in the sequel. Write $X_1 \simeq X_2$ if $X_1 = X_2$ a.e. (P_0) for X_1 and X_2 in V. It is easy to verify $\simeq$ is an equivalence relation on V. Let $M = \{X | X \in V, X = 0 \text{ a.e. } (P_0)\}$ so $X_1 \simeq X_2$ iff $X_1 - X_2 \in M$. Let L^2 be the set of equivalence classes in V.

(i) Show that L^2 is a real vector space with the obvious definition of addition and scalar multiplication.

Define $(\cdot, \cdot)$ on L^2 by $(y_1, y_2) = \mathcal{E} X_1 X_2$ where X_i is an element of the equivalence class y_i, $i = 1, 2$.

(ii) Show that $(\cdot, \cdot)$ is well defined and is an inner product on L^2.

Now, let $\mathcal{F}_0$ be a sub σ-algebra of $\mathcal{F}$. For $y \in L^2$, let Py denote the conditional expectation given $\mathcal{F}_0$ of any element in y.

(iii) Show that P is well defined and is a linear transformation on L^2 to L^2.

Let N be the set of equivalence classes of all $\mathcal{F}_0$ measurable functions in V. Clearly, N is a subspace of L^2.

(iv) Show that $P^2 = P$, P is the identity on N, and $\mathcal{R}(P) = N$. Also show that P is self-adjoint—that is $(y_1, Py_2) = (Py_1, y_2)$.

Would you say that P is the orthogonal projection onto N?

NOTES AND REFERENCES

1. The first half of this chapter follows Halmos (1968) very closely. After this, the material was selected primarily for its use in later chapters. The material on outer products and Kronecker products follows the author's tastes more than anything else.

2. The detailed discussion of angles between subspaces resulted from unsuccessful attempts to find a source that meshed with the treatment of canonical correlations given in Chapter 10. A different development can be found in Dempster (1969, Chapter 5).

3. Besides Halmos (1958) and Hoffman and Kunze (1971), I have found the book by Noble and Daniel (1977) useful for standard material on linear algebra.

4. Rao (1973, Chapter 1) gives many useful linear algebra facts not discussed here.