Rob Clifton

Introductory Notes on the Mathematics Needed for Quantum Theory

(1996)

CONTENTS

1

VECTOR SPACES

Recall the picture of a vector as a directed line segment $\nearrow$ with its direction indicated by an arrow attached to one end. Focus on the collection of all vectors in the plane that start at a fixed origin O but can point in any direction and have any length. Figure 1.1 illustrates two simple operations that can be performed on these vectors. We could alter the length and/or orientation of any given vector $\overrightarrow{OP}$ by multiplying it by a factor k, which will produce a new vector $\overrightarrow{OQ}$ lying in the same line (the figure illustrates the case $k < -1$). Alternatively, we could add two vectors together using the parallelogram law, obtaining a third vector $\overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{OB}$. In either case, the resulting vector once again points out from the origin. Thus, the set of all vectors at the origin is 'closed' under multiplication by a factor and under vector addition. These two closure properties are characteristic of a set that forms a vector space.

A vector space can be instantiated in myriad ways, not just geometrically in terms of vectors pointing in different directions in space subject to expansion and addition via the parallelogram law. In fact, a vector space need not instantiate any of the other familiar structure possessed by spatial vectors, such as the fact that there is a well-defined angle between any two of them or that each has a precisely defined length. We leave investigation of such additional structure until the next chapter.

1.1 Definition

Formally, a **real vector space** or **vector space over the real numbers** R, consists of a set V of objects $|v\rangle$, $|w\rangle, \ldots$ called vectors, a mapping $+$ that assigns to any two vectors $|v\rangle$ and $|w\rangle$ a third vector $|v\rangle + |w\rangle$, called their sum, and a mapping $\times$ that assigns to any given vector $|v\rangle$ and real number r another vector $r \times |v\rangle$, written $r|v\rangle$, which is the vector that results from multiplying $|v\rangle$ by r. In addition, these sum and product operations on vectors are required to satisfy certain properties.

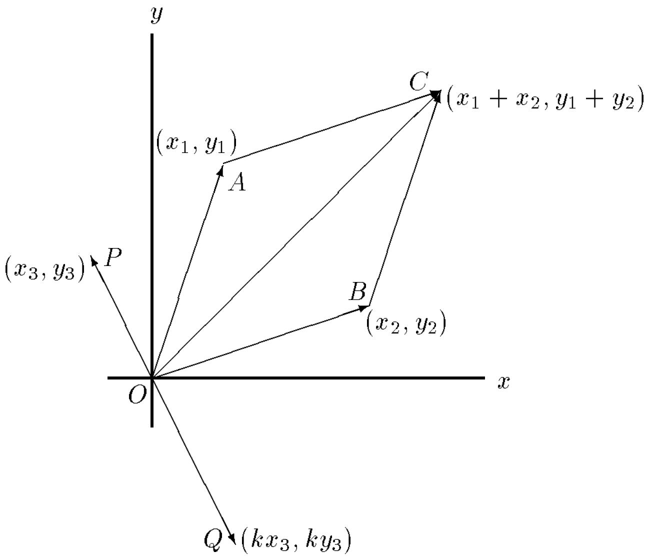

FIG. 1.1. Vector space operations

The sum of two vectors must be commutative and associative. There must be an identity vector, called the zero vector, $|0\rangle$, with the property that for any $|v\rangle \in V$, $|v\rangle + |0\rangle = |v\rangle$ (from which it follows that the zero vector is unique). And every vector $|v\rangle$ must have an inverse, which is written $-|v\rangle$, satisfying $|v\rangle + (-|v\rangle) = |0\rangle$ (from which it follows that inverses are unique). Subtraction is then defined by $|v\rangle - |w\rangle \overset{\text{def}}{=} |v\rangle + (-|w\rangle)$.

The product of a vector by a real number must possess the following properties (for all $r, r' \in R$ and $|v\rangle, |v'\rangle \in V$):

$$1|v\rangle = |v\rangle, \tag{1.1}$$

$$(r + r')|v\rangle = r|v\rangle + r'|v\rangle, \tag{1.2}$$

$$r(|v\rangle + |v'\rangle) = r|v\rangle + r|v'\rangle, \tag{1.3}$$

$$r(r'|v\rangle) = (rr')|v\rangle. \tag{1.4}$$

Notice that there is never a need for a notational distinction between real number addition and vector addition because the addition mapping at issue is always clear from the nature of the summands. Thus in (1.2) we have real numbers on the left and vectors on the right, avoiding any possibility of confusion. For the same reason, no special notation is needed to distinguish between the product of two real numbers and the product of a vector by a real number. Indeed, for both products we have done the usual thing and suppressed the × sign (cf. (1.4)).

Replacing 'real' everywhere above by 'complex', we obtain the definition of a **complex vector space**, or **vector space over the complex numbers** C. Often it will not be important whether we are dealing with a real or complex vector space, in which case we shall refer generically to the set over which the vector space is defined as a set of 'numbers' and denote it by the symbol K (with elements k, k', etc.). Indeed, we shall often just refer to a vector space V, leaving out reference to K altogether.

The following are immediate consequences of the definition of a vector space (for arbitrary $k \in K$ and $|v\rangle \in V$):

$$0|v\rangle = |0\rangle, \tag{1.5}$$

$$k|0\rangle = |0\rangle, \tag{1.6}$$

$$(-k)|v\rangle = k(-|v\rangle) = -(k|v\rangle), \tag{1.7}$$

$$k|v\rangle = |0\rangle \;\Rightarrow\; k = 0 \text{ or } |v\rangle = |0\rangle. \tag{1.8}$$

For example, (1.5) is proved as follows:

$$0|v\rangle = 0|v\rangle + 0|v\rangle - 0|v\rangle = (0+0)|v\rangle - 0|v\rangle = 0|v\rangle - 0|v\rangle = |0\rangle. \tag{1.9}$$

We leave the proofs of (1.6)—(1.8) to the reader.

1.2 Examples

We start with a simple but abstract example. Let S be any set. Take the set of vectors to be the set of all number-valued functions on S with **finite support**, i.e., all those functions that take nonzero values on at most finitely many elements of S. Defining the sum of two such functions f and f' to be the (finite support) function with action (for all $s \in S$)

$$(f + f')(s) = f(s) + f'(s) \tag{1.10}$$

and kf to be the (finite support) function with action (for all $k \in K$ and $s \in S$)

$$(kf)(s) = k(f(s)), \tag{1.11}$$

all the necessary properties of a vector space are satisfied. This construction is known as the **free vector space on the set** S.

Now take S to be the set $\{1, \ldots, n\}$. Evidently the free vector space on this set consists of all possible n-tuples of numbers. Writing each n-tuple as a $1 \times n$ column matrix, (1.10) reduces to the standard matrix addition rule

$$\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} + \begin{pmatrix} k'_1 \\ \vdots \\ k'_n \end{pmatrix} = \begin{pmatrix} k_1 + k'_1 \\ \vdots \\ k_n + k'_n \end{pmatrix} \tag{1.12}$$

and (1.11) reduces to the standard rule for multiplying a matrix by a number

$$k \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = \begin{pmatrix} kk_1 \\ \vdots \\ kk_n \end{pmatrix}. \tag{1.13}$$

We shall call this the vector space of column matrices and denote it by ℓ^n. But when we need to focus on the case $K = R$ or $K = C$, we shall follow standard notation and denote ℓ^n as R^n or C^n. R^2 is none other than the space of vectors in the plane with which we began the previous section. Associating with each vector the coordinates of its tip relative to the origin and some fixed coordinate axes, we can write the vector's coordinates as a 1×2 column matrix. (1.12) is then equivalent to the parallelogram addition law and (1.13) to changing the length of the vector by a factor k (cf. the coordinates in figure 1.1).

Next, take S to be the set of all natural numbers N. The free vector space in this case, which we denote by ℓ^N, is the set of all column matrices with a countably infinite number of entries but only finitely many of them nonzero. Of course there is nothing stopping us from dropping this finite support requirement and considering the set of *all* countably infinite column matrices endowed with the infinite analogues of the vector space operations in (1.12) and (1.13). We shall denote that space by ℓ.

1.3 Subspaces and Spans

Let V be any vector space and W be any nonempty subset of V. W is called a **subspace** of V if, upon restricting the operations $+$ and $\times$ so that they act only upon elements of W, it is a vector space in its own right. Equivalently, W is a subspace exactly when (for any $k, k' \in K$)

$$|w\rangle, |w'\rangle \in W \Rightarrow \; k|w\rangle + k'|w'\rangle \in W; \tag{1.14}$$

for all that is needed for W to be a subspace of V is that W be closed under $+$ and $\times$ and then these operations will automatically possess the necessary vector space properties (since they are merely restrictions to W of the vector space operations in V). Note that since W is nonempty, this equivalent definition of subspacehood entails that $|0\rangle \in W$. Indeed, a trivial example of a subspace is the zero subspace consisting of just the single vector $\{0\}$. (To avoid the inelegant notation '$\{|0\rangle\}$', we shall always drop vector brackets inside set brackets when it is clear that the elements of the set are vectors.) Less trivial examples of subspaces are lines—or 'rays'—and planes through the origin in R^3. Also, ℓ^N is a subspace of ℓ; however neither space contains ℓ^n as a subspace.

As the reader may easily verify, any intersection of subspaces of V is also a subspace of V (whereas the union of two subspaces is not a subspace, unless of course one is contained in the other). So we can define the **subspace generated** by a subset $S \subseteq V$ to be the intersection of all the subspaces of V that contain S, which is evidently the smallest subspace of V containing S. Thus, in R^3, any two distinct lines through the origin generate a plane, and a plane together with any vector not contained in the plane generate the whole space.

A vector $|v\rangle \in V$ is said to be a **linear combination** of elements in a subset $S \subseteq V$ if it can be expanded as

$$|v\rangle = \sum_{i=1}^{n} k_i |s_i\rangle \tag{1.15}$$

in terms of a finite subset $\{s_i\}_{i=1}^n \subseteq S$ with a finite set of numbers $\{k_i\}_{i=1}^n$ as **expansion coefficients**. (Note that one considers only finite sums. Later on we shall need infinite vector sums, but such sums cannot be defined until our vector spaces are taken to possess some additional 'topological' structure.) The **span** of a set $S \subseteq V$ is the set of all vectors in V that are linear combinations of elements in S. Thus, the span of S is just another term for the subspace S generates. We shall denote the span of the union of two *subspaces* U and W by $U + W$, because it is just the set of all vectors in V that can be written as a sum of a vector from U and a vector from W.

1.4 The Lattice of Subspaces

The binary operations of intersection $\cap$ and span $+$ on pairs of subspaces endow the set of all subspaces of any vector space with the structure of a 'lattice'.

To see what a lattice is, we first need another definition. A **partially ordered set**—or **poset**—is a set S on which there is a binary relation $\leq$, read 'less than or equal to', that satisfies (for all $a, b, c \in S$):

$$\textbf{reflexivity}\text{: } a \leq a, \tag{1.16}$$

$$\textbf{antisymmetry}\text{: } a \leq b,\ b \leq a\ \Rightarrow\ a = b, \tag{1.17}$$

$$\textbf{transitivity}\text{: } a \leq b,\ b \leq c\ \Rightarrow\ a \leq c. \tag{1.18}$$

The reason for the term 'partial' is that we are allowing that certain pairs of elements in S may not be ordered with respect to each other. For example, if we order the set of all subsets of any given set T by defining (for any $A, B \subseteq T$) $A \leq B$ just in case $A \subseteq B$, then this ordering—which is called **ordering the subsets of T by inclusion**—is necessarily partial, because for disjoint or partly overlapping subsets of T, neither is contained in the other. On the other hand, the set of real numbers with $\leq$ given its usual meaning is a poset that is **totally ordered**.

A **lattice** $\mathcal{L}$ is simply a poset in which each pair of elements $a, b \in \mathcal{L}$ possess both a join and a meet. The **join** (or least upper bound) of a and b, written $a \vee b$, is the least element in $\mathcal{L}$ greater than or equal to both a and b, and the **meet** (or greatest lower bound) of a and b, written $a \wedge b$, is the greatest element less than or equal to both a and b. Not all posets need be lattices (try to think of an example of one that is not), but when a meet or join of two elements does exist, it must be unique (by a simple argument using antisymmetry), rendering unambiguous the denotations '$a \vee b$' and '$a \wedge b$'. Returning, then, to the set of all subspaces of a vector space, order them by inclusion. Because $U + W$ is the smallest subspace containing both U and W, $U + W$ is the least upper bound of U and W. And since the intersection of U and W is the largest subspace contained in them both, $U \cap W$ is their greatest lower bound. So the poset of subspaces of a vector space V, ordered by inclusion, is indeed a lattice, call it $\mathcal{L}(V)$, with $\vee$ given by $+$ and $\wedge$ by $\cap$.

$\mathcal{L}(V)$ has some additional properties that not all lattices need have. It has a (necessarily unique) **maximum element** 1 and a (necessarily unique) **minimum element** 0, so, for all $a \in \mathcal{L}$, $0 \leq a \leq 1$. In $\mathcal{L}(V)$, 1 is the whole space V and 0 the zero subspace. $\mathcal{L}(V)$ is also **complete**, meaning that every subset of $\mathcal{L}(V)$ has both a join and a meet, because the span or intersection of an arbitrary collection of subspaces is also a subspace. Finally, $\mathcal{L}(V)$ is an **atomic** lattice, meaning that every nonzero element in the lattice contains a minimal nonzero element. Such an element is called an **atom** of the lattice, and the precise definition is: a is an atom in $\mathcal{L}$ if and only if $a \neq 0$ and, for all $b \in \mathcal{L}$, $b \leq a \Rightarrow b = a$ or $b = 0$. If U is any nonzero subspace in $\mathcal{L}(V)$, then an atom contained in U is obtained by taking the subspace generated by any nonzero vector contained in U.

For subspaces of a vector space, intersection does not distribute over sum, nor does sum distribute over intersection. That is, if A, B and C are subspaces, then in general

$$A \cap (B + C) \neq (A \cap B) + (A \cap C), \tag{1.19}$$

$$A + (B \cap C) \neq (A + B) \cap (A + C), \tag{1.20}$$

though equalities will hold for certain choices of the subspaces. (A single choice for A, B, and C in R^2 suffices to establish both inequalities above. What is it?) When the meet in a lattice fails to distribute over the join (or vice-versa), as in the case of $\mathcal{L}(V)$, the lattice is called **nondistributive**.

A **sublattice** of a lattice $\mathcal{L}$ is any subset of $\mathcal{L}$ that forms a lattice in its own right (under the partial ordering inherited from $\mathcal{L}$) or, equivalently, any subset closed under the operations of meet and join. Since the intersection of any collection of sublattices is itself a sublattice, we can define the **sublattice generated** by a subset $S \subseteq \mathcal{L}$ to be the intersection of all sublattices of $\mathcal{L}$ that contain S. And since the generated sublattice is, by definition, the smallest sublattice containing S, it is what you get when you close S under the meet and join operations of $\mathcal{L}$. A simple example is the sublattice of $\mathcal{L}(R^3)$ generated by any three distinct rays. This sublattice consists of the rays themselves (automatically), the plane(s) in which they lie (closing under $\vee$), the zero subspace (closing under $\wedge$), and the whole space (closing under $\vee$ again).

1.5 Linear Independence and Bases

A subset $T \subseteq V$ is called a **linearly independent** set of vectors if for any finite set of distinct vectors $\{t_i\}_{i=1}^n \subseteq T$ and any finite set of numbers $\{k_i\}_{i=1}^n$,

$$\sum_{i=1}^n k_i|t_i\rangle = |0\rangle \;\Rightarrow\; k_i = 0 \text{ for } i = 1, \ldots, n. \tag{1.21}$$

In particular, T cannot be linearly independent if it contains the zero vector, since $1|0\rangle = |0\rangle$. If a subset $T \subseteq V$ fails to be linearly independent, evidently at least one element of T must be a linear combination of other elements in T. In that case, T is called **linearly dependent**.

An example of a linearly independent set is the set of all 'blip' functions $\{f_s\}_{s \in S}$ in the free vector space over S, which are defined by

$$f_s(x) = \begin{cases} 1 \text{ if } x = s, \\ 0 \text{ if } x \neq s. \end{cases} \tag{1.22}$$

Another example is the sequence of matrices in ℓ^N:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix}, \ldots \tag{1.23}$$

Any vector in the span of a linearly independent set is a *unique* linear combination of members of that set. The proof consists of nothing more than assuming that there is a vector expandable in more than one way, i.e., with different coefficients, in terms of vectors in the spanning set, and then deducing from their linear independence that the different coefficients must, after all, coincide.

A subset $B \subseteq V$ is called a **basis** for V if B is linearly independent and V is in its span. So by the previous paragraph, any vector's expansion coefficients in a given basis are unique. Equivalently, B is a basis for V if and only if it is linearly independent and not a proper subset of any other linearly independent subset of V, i.e., if B is a **maximally independent** subset of V. For if U were *not* maximally independent, there would be a vector not linearly dependent upon the vectors in B, so that B could not possibly span V. On the other hand, assuming B *is* maximally independent, there can be no vectors in V outside B's span; for if there were such a vector $|v\rangle$, $B \cup \{v\}$ would form a linearly independent set properly containing B.

In R^3 any three noncoplanar vectors form a basis. The blip functions on a set span the free vector space over it, and therefore constitute a basis. Finally, the matrices in (1.23) are a basis for ℓ^N (how do they span it?) but *not* for ℓ. Indeed, no set of vectors drawn entirely from the subspace ℓ^N of ℓ could possibly form a basis for the latter, because no column matrix without finite support can be written as a finite linear combination of matrices *with* finite support.

Every vector space has a basis (excluding, of course, the zero vector space consisting of $\{0\}$ alone). The argument goes as follows. Start with a single (nonzero) vector $|v\rangle \in V$, which trivially forms a linearly independent set on its own. If there is a vector in V, say $|w\rangle$, not in the span of $\{v\}$, add it to that set to produce a new linearly independent set $\{v, w\}$. (And if there is no such $|w\rangle$, $\{v\}$ must already be a basis for V.) If $\{v, w\}$ still fails to span V, add to that set another vector not in its span so that the augmented set is again linearly independent. Continue adding vectors from V into the set in this way until you finally get a maximally independent set, i.e., a basis for V.

But what if infinitely many vectors have to be added before a basis for V is reached (if ever)? In that case, the argument must employ **Zorn's lemma**. This 'lemma' (which would be better described as an axiom, since it can be neither proved nor disproved) states that: if every totally ordered subset of a poset is bounded above, then the poset has a (not necessarily unique) maximal element. Here, a subset A of a poset S is said to be **bounded above** if there is an element $s \in S$ such that $a \leq s$ for every $a \in A$, and an element $s \in S$ is called a **maximal element** if, for any $s' \in S$, $s \leq s' \Rightarrow s = s'$. We can now finish the argument of the previous paragraph in the infinite case as follows.

Consider the poset L of all linearly independent subsets of V, ordered by inclusion. (Note that L is *not* the poset $\mathcal{L}(V)$, whose elements are subspaces.) If it can be shown that L satisfies the hypothesis of Zorn's lemma, we can use that lemma to conclude that V contains a maximally independent subset. So let $\{S_\lambda\}_{\lambda\in\Lambda}$ be any totally ordered subset of L parameterized by some index set Λ; so, whenever $\lambda \leq \lambda'$, we have $S_\lambda \subseteq S_{\lambda'}$, with each S_λ a set of linearly independent vectors in V. Then the set $\bigcup_{\lambda\in\Lambda} S_\lambda$ must also be linearly independent. For any (finite) linear dependency could only occur if it occurred within one of the sets S_λ, which are linearly independent. Thus we see that $\bigcup_{\lambda\in\Lambda} S_\lambda \in L$. And since this union contains each S_λ, it bounds the collection $\{S_\lambda\}_{\lambda\in\Lambda}$ from above so that Zorn's lemma indeed applies to L.

1.6 Dimension

The most important characteristic of a vector space is its dimension. In order to define dimension, we first need assurance that all bases of a vector space have the same 'cardinality'. We start with the simplest case when all the bases of the vector space are finite and show that they must have the same number of elements.

Suppose $W = \{w_1, \ldots, w_m\}$ and $U = \{u_1, \ldots, u_n\}$ are two bases for V. We need to show that $m = n$. Because U spans V, adding the first vector in W to U yields a new sequence of vectors $\{w_1, u_1, \ldots, u_n\}$ that still span V but must now be linearly dependent. This means that there must be at least one vector in this sequence that is a linear combination of the vectors that precede it. That vector cannot be $|w_1\rangle$, so it must be some vector from the set U, say $|u_i\rangle$. Since $|u_i\rangle$ is a linear combination of other vectors in the sequence, it may be deleted so that the remaining sequence $\{w_1, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n\}$ continues to span V.

Next, add the *second* vector in W to the beginning of the previous sequence to obtain

$$\{w_1, w_2, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n\} \tag{1.24}$$

which (again) still spans V but must be linearly dependent. So there is (again) a vector in this sequence that is a linear combination of preceding vectors. That vector (still) cannot be $|w_1\rangle$, but neither can it be $|w_2\rangle$ because W is linearly independent. So some $|u_j\rangle$ is a linear combination of preceding vectors in the sequence, and we are again free to delete it so that the remaining sequence

$$\{w_1, w_2, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n\} \tag{1.25}$$

continues to span V.

Continue by repeating this argument, at each stage adding a vector from W to the beginning of the sequence and deleting a vector in U from the end. If $m > n$, then n iterations of the argument will eventually yield the conclusion that the *proper* subset $\{w_1, \ldots, w_n\}$ of W spans V, which contradicts the linear independence of W. Interchanging the roles played by the sets U and W throughout the argument, $n > m$ implies a contradiction with the fact that U is linearly independent. Therefore, the only possibility left is $n = m$.

To establish the same result for vector spaces with infinite bases, we need to recall a few facts about mappings. A mapping $\varphi : S \to T$ from one set S to another T associates with any $s \in S$ a unique element $\varphi(s) \in T$. The mapping φ is called **one-to-one** if $\varphi(s) = \varphi(s') \;\Rightarrow\; s = s'$ (for all $s, s' \in S$) and **onto** if for any $t \in T$ there is an $s \in S$ such that $\varphi(s) = t$. If φ is one-to-one and onto, it is called an **isomorphism of sets**. If $\varphi : S \to T$ is one-to-one, then it has an **inverse mapping** $\varphi^{-1} : T \to S$ which maps $t \in T$ to the unique $s \in S$ such that $\varphi(s) = t$. If $\varphi : S \to T$ and $\vartheta : T \to Z$ are two mappings, then their **composition**, $\vartheta \circ \varphi : S \to Z$, is the mapping from S to Z given by $(\vartheta \circ \varphi)(s) = \vartheta(\varphi(s))$ (for all $s \in S$). For $\varphi : S \to T$ an isomorphism, evidently $\varphi \circ \varphi^{-1}$ is the identity mapping on T and $\varphi^{-1} \circ \varphi$ the identity on S.

Two sets S and T are said to have the same **cardinality**, written $\kappa_S = \kappa_T$, if they are **isomorphic**, i.e., if there is an isomorphism between them. If S is merely isomorphic to a subset of T, then the cardinality of S is defined to be less than or equal to the cardinality of T and we write $\kappa_S \leq \kappa_T$. It is immediate that $\leq$ is reflexive and transitive. That $\leq$ is antisymmetric is the content of the famous Schroeder-Bernstein theorem (Devlin 1979, Theorem 6.3). Hence, cardinalities form a poset.

Now return to the problem of establishing that, for any two bases U and W of a vector space, $\kappa_U = \kappa_W$. We can suppose that both κ_U and κ_W are infinite; for if, say, κ_U were finite, then since κ_W is infinite we could again run through our earlier 'replacement argument' (cf. (1.24) and (1.25)) and, after a finite number of iterations, 'use up' the basis vectors in U, getting a contradiction with the linear independence of W. Supposing, then, that both κ_U and κ_W are infinite, it is enough to show $\kappa_U \leq \kappa_W$; for by switching the roles of U and W throughout the argument, we get $\kappa_W \leq \kappa_U$, and hence $\kappa_U = \kappa_W$.

Consider an arbitrary vector $|w\rangle \in W$. Because U is a basis, $|w\rangle$ is in the span of some finite subset of U, call it $F_{|w\rangle}$. Furthermore, every $|u\rangle \in U$ is contained in at least one of the sets $F_{|w\rangle}$ for some $|w\rangle \in W$. For if some $|u\rangle \in U$ were not, then because W is a basis and $|u\rangle$ is in its span, $|u\rangle$ would have to be in the span of a subset $F_{|w\rangle}$ of U that (by hypothesis) fails to include $|u\rangle$ itself, contradicting the linear independence of U. So we have that every $|w\rangle \in W$ is associated with a finite subset $F_{|w\rangle}$ of U in such a way that

$$U = \bigcup_{|w\rangle \in W} F_{|w\rangle}. \tag{1.26}$$

But the infinite cardinality of this union, and thus U, is always less than or equal to κ_W. This fact is easily seen if W is countably infinite, because a countable union of finite sets is again countable. If W has higher cardinality, appeal has to be made to the fact that, for any infinite cardinal κ, a union of at most κ sets of cardinality at most κ has cardinality at most κ (Devlin 1979, Corollary 7.8).

The common cardinality of all the bases of a vector space V is called the **dimension** of V (taking the zero vector space to have dimension zero). Since the blip functions in the free vector space over a set S are a basis, and there are as many blip functions as there are elements of S, vector spaces of *any* dimension exist. Moreover, in the next section we shall see that once the dimension is fixed, the vector space itself is fixed 'up to isomorphism'. So there is a sense in which once one has understood free vector spaces, one has understood them all.

The superscripts that occur in R^n, C^n, and ℓ^n are there to indicate that these spaces have finite dimension n, and we shall indicate a generic finite-dimensional space by V^n. The superscript in ℓ^N means it has *countably* infinite dimension (recall $N =$ the natural numbers). By contrast, we have omitted any superscript in 'ℓ' because its dimension is *un*countable, a fact that may be seen by the following argument.

Recall that ℓ is the vector space of all number-valued functions on N. Also recall that the real numbers in the interval $[0,1]$ are uncountable, and that each $r \in [0,1]$ has a unique decimal expansion of the form $.i_1 i_2 i_3 \cdots$ where each digit i_n is a whole number. With reference to r's decimal expansion, define the following infinite sequence of natural numbers,

$$S_r = \{2^{i_1} 3^{i_2} \cdots p_n^{i_n}\}_{n=1}^{\infty}, \tag{1.27}$$

where p_n is the nth prime number. Observe that

$$r \neq r' \Rightarrow \kappa_{S_r \cap S_{r'}} < \infty. \tag{1.28}$$

(Why?) Next, for each $r \in [0,1]$ define an element of ℓ by

$$g_r(n) = \begin{cases} 1 \text{ if } n \in S_r, \\ 0 \text{ if } n \notin S_r. \end{cases} \tag{1.29}$$

Then the uncountable set $\{g_r\}_{r \in [0,1]}$ is linearly independent, and hence ℓ cannot have countable dimension. For suppose that some finite number of these functions satisfies $k_1 g_{r_1} + \cdots + k_m g_{r_m} = 0$. For an arbitrary index j, we need to show $k_j = 0$. It must be the case that

$$S_{r_j} \not\subseteq S_{r_1} \cup \cdots \cup S_{r_{j-1}} \cup S_{r_{j+1}} \cup \cdots \cup S_{r_m}. \tag{1.30}$$

For if not, then evidently

$$S_{r_j} = (S_{r_1} \cap S_{r_j}) \cup \cdots \cup (S_{r_{j-1}} \cap S_{r_j}) \cup (S_{r_{j+1}} \cap S_{r_j}) \cup \cdots \cup (S_{r_m} \cap S_{r_j}); \tag{1.31}$$

and, since S_{r_j} is infinite, at least one of the $S_{r_l} \cap S_{r_j}$ $(l \neq j)$ would also have to be infinite, violating (1.28). Because of (1.30), we may choose an $n \in S_{r_j}$ such that $n \notin S_{r_1} \cup \cdots \cup S_{r_{j-1}} \cup S_{r_{j+1}} \cup \cdots \cup S_{r_m}$, and thus for that choice we obtain

$$0 = k_1 g_{r_1}(n) + \cdots + k_j g_{r_j}(n) + \cdots + k_m g_{r_m}(n) = k_j \cdot 1 = k_j \tag{1.32}$$

as required.

1.7 Linear Mappings and Isomorphism

Let V and W both be vector spaces over the same set of numbers K. To avoid excess brackets, we shall write the result of applying the mapping $\varphi : V \to W$ to a vector $|v\rangle \in V$ as $\varphi|v\rangle$ (rather than '$\varphi(|v\rangle)$'). The mapping φ is called a **linear mapping** if (for any $|v\rangle, |v'\rangle \in V$ and $k, k' \in K$)

$$\varphi(k|v\rangle + k'|v'\rangle) = k\varphi|v\rangle + k'\varphi|v'\rangle, \tag{1.33}$$

where the vector space operations on the left are performed in V and on the right in W. Since linear mappings preserve sums of vectors and products of vectors by numbers, they preserve all the relevant structure of a vector space. (We shall encounter similar 'structure-preserving' mappings when we introduce other mathematical structures later.) Note that if φ is linear, $\varphi|0\rangle = |0\rangle$, because $\varphi|0\rangle = \varphi(|v\rangle - |v\rangle) = \varphi|v\rangle - \varphi|v\rangle = |0\rangle$. We leave the reader to check that the composition of two linear mappings is again linear, and that if φ is linear, so is φ^{-1} (when it exists).

Let $\phi : S \to T$ be a mapping between sets. Recall that for any $A \subseteq S$, the set

$$\phi(A) \overset{\text{def}}{=} \{\phi(s) \in T : s \in A\} \tag{1.34}$$

is called the **image** of A by ϕ, and for any $B \subseteq T$, the set

$$\phi^{-1}(B) \overset{\text{def}}{=} \{s \in S : \phi(s) \in B\} \tag{1.35}$$

is called the **inverse image** of B by ϕ. (Note that using the notation 'ϕ^{-1}' in this way by no means implies that ϕ is invertible. As an exercise in the use of this notation, the reader might like to show that ϕ is one-to-one if and only if $\phi^{-1}(\phi(A)) = A$ for all $A \subseteq S$, and onto if and only if $\phi(\phi^{-1}(B)) = B$ for all $B \subseteq T$.) The image of a subspace $U \subseteq V$ by a linear mapping $\varphi : V \to W$, i.e., $\varphi(U)$, is again a subspace. Clearly, $|0\rangle \in \varphi(U)$ (in virtue of $\varphi|0\rangle = |0\rangle$). And if $|w\rangle, |w'\rangle \in W$, then $\varphi|v\rangle = |w\rangle$ and $\varphi|v'\rangle = |w'\rangle$ (for some $|v\rangle, |v'\rangle \in U$). But we know that $k|v\rangle + k'|v'\rangle \in U$, thus

$$\varphi(k|v\rangle + k'|v'\rangle) = k|w\rangle + k'|w'\rangle \in \varphi(U). \tag{1.36}$$

Similarly, $\varphi^{-1}(U)$ is a subspace whenever $U \subseteq W$ is a subspace.

The action of a linear mapping $\varphi : V \to W$ is completely specified by its action on a basis B for V, because the action of φ on an *arbitrary* $|v\rangle \in V$, where $|v\rangle = k_1|b_1\rangle + \cdots + k_n|b_n\rangle$ and $\{b_i\}_{i=1}^n \subseteq B$, must then be

$$\varphi|v\rangle = \sum_{i=1}^{n} k_i \varphi|b_i\rangle \tag{1.37}$$

by linearity. Note, also, that the set of all linear mappings from one fixed vector space V to another W forms a vector space in its own right, when we define the linear combination $k\varphi + k'\varphi'$ of the linear mappings $\varphi : V \to W$ and $\varphi' : V \to W$ to be the (linear) mapping with action (for all $k, k' \in K$ and $|v\rangle \in V$)

$$(k\varphi + k'\varphi')|v\rangle = k\varphi|v\rangle + k'\varphi'|v\rangle. \tag{1.38}$$

The mapping $\varphi : V \to W$ is an **isomorphism of vector spaces** if it is an isomorphism of sets and is linear. For example, ℓ^n, while not a subspace of ℓ^N, is isomorphic to one (under the obvious isomorphism). Not only do isomorphic vector spaces have the same cardinality—and thus are structurally identical as sets—but operations involving vectors in one of the spaces can be mimicked by operations involving the corresponding vectors (under the isomorphism) in the other space, making them structurally equivalent as vector spaces as well.

The chief result that tidies up the subject is that two vector spaces V and W are isomorphic, written $V \cong W$, if and only if they have the same dimension.

To see why, suppose first that V and W have the same dimension. Let B be a basis in V, L a basis in W, and pick an isomorphism of sets $\psi : B \to L$. Define the linear mapping $\varphi : V \to W$ by $\varphi|b\rangle = \psi|b\rangle$ for all $|b\rangle \in B$ (which, recall, defines it completely). It is then a simple matter to show that this φ is one-to-one and onto. For 'one-to-one' use the fact that the expansion coefficients of any vector in W in terms of the basis L are unique, and for 'onto' use the fact that L spans W.

Conversely, suppose $V \cong W$ and pick an isomorphism of vector spaces $\varphi : V \to W$. It follows that φ's action on elements in any basis B for V will produce a basis $L = \{\varphi|b\rangle : b \in B\}$ for W. To prove that L is linearly independent, note that for any finite subset of L we have

$$\sum_{j=1}^{n} k_j \varphi|b_j\rangle = |0\rangle \Rightarrow \ \varphi \sum_{j=1}^{n} k_j |b_j\rangle = |0\rangle \tag{1.39}$$

$$\Rightarrow \sum_{j=1}^{n} k_j|b_j\rangle = |0\rangle \tag{1.40}$$

$$\Rightarrow k_j = 0 \text{ for } j = 1 \text{ to } n, \tag{1.41}$$

using the linearity of φ, the fact that φ is one-to-one (remembering that $\varphi|0\rangle = |0\rangle$), and the linear independence of B. An equally elementary argument establishes that L spans W. Therefore, since φ is an isomorphism mapping basis B in V to basis L in W, V and W have the same dimension.

A consequence of this result is that if B is a basis for V, then since the free vector space over the *set* B has the same dimension as V, they are isomorphic. So we see that nothing more subtle goes on in an arbitrary vector space V than what goes on in a free vector space. In particular, having mastered ℓ^n—and, in the real case, the geometric interpretation of R^n—one has understood *all* n-dimensional vector spaces.

It is often useful for performing operations on vectors to use a particular isomorphism to translate into the language of matrices in ℓ^n from an arbitrary n-dimensional vector space V^n. The standard way of doing so is to pick a basis B for V^n and consider the mapping φ_B that sends $|v\rangle \in V^n$ to the column matrix in ℓ^n consisting of $|v\rangle$'s (unique) expansion coefficients in the basis B. This φ_B is indeed an isomorphism and, for any vector in V, delivers its **matrix representation** relative to the basis B. So if we have any vector equation in V^n, all its vectors can be replaced by their matrix representations relative to a common basis, leaving any expansion coefficients in the equation unchanged. The the equation can then be manipulated according to the matrix rules that define the vector operations in ℓ^n. The same points hold for a vector space with countable dimension and the matrix representations of its vectors in ℓ^N.

1.8 Operators and Algebras

A **linear operator**—or just **operator**—on a vector space V is a linear mapping $\boldsymbol{F} : V \to V$ from V to itself. The image of V by $\boldsymbol{F}$, i.e., the subspace $\boldsymbol{F}(V)$, is usually called the **range** of $\boldsymbol{F}$. The vector space of all operators on a fixed vector space V has a structure not shared by the vector space of linear mappings between different vector spaces. Because the composition, or product, of two operators $\boldsymbol{F}$ and $\boldsymbol{G}$ is automatically another, $\boldsymbol{FG}$, with action $\boldsymbol{FG}|v\rangle = \boldsymbol{F}(\boldsymbol{G}|v\rangle)$ (for all $|v\rangle \in V$), the operators on V form an 'algebra'.

An **(associative) algebra** $\mathcal{A}$ over K consists of a vector space over K with an additional product mapping that assigns to any two vectors $X, Y \in \mathcal{A}$ a third vector $XY \in \mathcal{A}$, satisfying (for all $k \in K$ and $X, Y, Z \in \mathcal{A}$):

$$\textbf{bilinearity}\text{: } X(Y + kZ) = XY + kXZ, (Y + kZ)X = YX + kZX, \tag{1.42}$$

$$\textbf{associativity}\text{: } X(YZ) = (XY)Z. \tag{1.43}$$

One says 'real algebra $\mathcal{A}$' or 'complex algebra $\mathcal{A}$' according to whether $K = R$ or $K = C$. For the algebra of operators on a vector space V, as opposed to an abstract algebra, we write $\mathcal{A}(V)$. Of course $\mathcal{A}(V)$ will be real or complex according to whether V is real or complex.

Elements of $\mathcal{A}(\ell^n)$ are given by $n \times n$ matrices of numbers, which map column matrices in ℓ^n to column matrices in ℓ^n according to the following transparently linear multiplication rule:

$$\begin{pmatrix} k_{11} & \ldots & k_{1n} \\ . & . \; . \; . & . \\ . & . \; . \; . & . \\ . & . \; . \; . & . \\ k_{n1} & \ldots & k_{nn} \end{pmatrix} \begin{pmatrix} l_1 \\ . \\ . \\ . \\ l_n \end{pmatrix} = \begin{pmatrix} m_1 \\ . \\ . \\ . \\ m_n \end{pmatrix} \tag{1.44}$$

with

$$m_i = \sum_{p=1}^{n} k_{ip} l_p . \tag{1.45}$$

The sum of two operators in $\mathcal{A}(\ell^n)$ is then just the sum of their $n \times n$ matrices, entry for entry, and the product of an operator by a number is the matrix obtained by multiplying each entry of the $n \times n$ matrix by that number. Of course, these operations are analogous to the operations that can be performed on the column matrices in ℓ^n itself.

From the product rule in (1.44) and (1.45), it is straightforwardly deduced that the product of two operators in $\mathcal{A}(\ell^n)$ is determined by the following more general matrix multiplication rule:

$$\begin{pmatrix} k_{11} & \ldots & k_{1n} \\ . & . \; . \; . & . \\ \boldsymbol{k_{i1}} & \boldsymbol{\ldots} & \boldsymbol{k_{in}} \\ . & . \; . \; . & . \\ k_{n1} & \ldots & k_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & . & \boldsymbol{l_{1j}} & . & l_{1n} \\ . & . & \boldsymbol{.} & . & . \\ . & . & \boldsymbol{.} & . & . \\ . & . & \boldsymbol{.} & . & . \\ l_{n1} & . & \boldsymbol{l_{nj}} & . & l_{nn} \end{pmatrix} = \begin{pmatrix} m_{11} & . & . & . & m_{1n} \\ . & . & . & . & . \\ . & . & \boldsymbol{m_{ij}} & . & . \\ . & . & . & . & . \\ m_{n1} & . & . & . & m_{nn} \end{pmatrix} \tag{1.46}$$

with

$$m_{ij} = \sum_{p=1}^{n} k_{ip} l_{pj}. \tag{1.47}$$

Notice that, regarded just as a vector space, the algebra $\mathcal{A}(\ell^n)$ has dimension n^2, since a basis is given by the n^2 matrices with a 1 in a single slot of the matrix and 0 elsewhere. In fact, whatever the dimension κ of V, the dimension of $\mathcal{A}(V)$ is always κ^2. (Why?) Since $\kappa^2 = \kappa$ whenever κ is an infinite cardinal (Devlin 1979, Theorem 7.5), the vector spaces $\mathcal{A}(V)$ and V are isomorphic whenever the latter has infinite dimension.

It is easy to conjure up matrices whose product depends on the order in which they are multiplied. Take for example

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.48}$$

in $\mathcal{A}(R^2)$ (or $\mathcal{A}(C^2)$). For two operators, or more generally two elements X and Y in an abstract algebra, it is often convenient to invoke their **commutator** and **anti-commutator**, defined by:

$$[X, Y] \overset{\text{def}}{=} XY - YX, \tag{1.49}$$

$$[X, Y]_+ \overset{\text{def}}{=} XY + YX. \tag{1.50}$$

If the commutator of any two elements in an algebra is 0, the zero vector of the algebra, it is called **commutative** or **abelian**. Generalizing from the example in (1.48), it is easy to convince oneself that $\mathcal{A}(V)$ will be commutative if and only if the dimension of V is trivial (i.e., 1 or 0).

There are two further important algebraic structures that the operators on a vector space instantiate in virtue of forming an associative algebra. A **Lie algebra** $\mathcal{A}$ over K is a vector space over K with a product, denoted by $\bullet$, that is not necessarily associative, but is still bilinear and satisfies (for all $X, Y, Z \in \mathcal{A}$):

$$\textbf{anti-symmetry}\colon X \bullet Y = -Y \bullet X, \tag{1.51}$$

$$\textbf{Jacobi identity}\colon X \bullet (Y \bullet Z) + Z \bullet (X \bullet Y) + Y \bullet (Z \bullet X) = 0. \tag{1.52}$$

The elements of an associative algebra are readily verified to form a Lie algebra if we define $X \bullet Y$ to be $[X, Y]$. Note that anti-symmetry and the Jacobi identity are trivial when the associative algebra is commutative, and in that case the Lie algebra *is* associative. A **Jordan algebra** $\mathcal{A}$ over K is a vector space over K with a product, denoted by $\circ$, that is (again) not necessarily associative, but is still bilinear and satisfies (for all $X, Y \in \mathcal{A}$):

$$\textbf{symmetry}\colon X \circ Y = Y \circ X, \tag{1.53}$$

$$\textbf{Jordan identity}\colon (X \circ X) \circ (Y \circ X) = ((X \circ X) \circ Y) \circ X. \tag{1.54}$$

Defining $X \circ Y$ to be $[X, Y]_+$, the elements of an associative algebra form a Jordon algebra (and, again, symmetry and the Jordan identity are trivial when the associative algebra is commutative, in which case the Jordan algebra is itself associative). When we refer simply to an algebra '$\mathcal{A}$', or (concretely) to '$\mathcal{A}(V)$', it should always be understood that *associative* algebra is meant; otherwise, we shall say explicitly 'Lie algebra $\mathcal{A}$', 'Jordan algebra $\mathcal{A}(V)$', etc.

A **subalgebra** of an algebra $\mathcal{A}$ is a subset of $\mathcal{A}$ forming an algebra in its own right under the operations it inherits from $\mathcal{A}$. Equivalently, a subset of $\mathcal{A}$ is a subalgebra when it is a subspace closed under products. A trivial example is the commutative subalgebra of $\mathcal{A}(V)$ consisting of all multiples of the identity operator $\boldsymbol{I}$. (It is not part of the definition of an algebra that it has an identity, but the operator algebras we shall consider typically do.) Less trivially, the set of all operators that commute with a given operator forms a subalgebra. (But must it be abelian?) On the other hand, the set of all operators on R^3 that act by rotating vectors about a common fixed axis does *not* form an algebra. (Why?)

The intersection of subalgebras is again a subalgebra, so we can define the **subalgebra** generated by a set of elements in an algebra in the usual way, viz., as the intersection of all subalgebras containing the set. A simple example is the commutative subalgebra generated by a single operator $\boldsymbol{F}$ and the identity operator $\boldsymbol{I}$, which is the set of all polynomials in the operator $\boldsymbol{F}$:

$$k_m \boldsymbol{F}^m + k_{m-1} \boldsymbol{F}^{m-1} + \cdots + k_1 \boldsymbol{F}^1 + k_0 \boldsymbol{I} \tag{1.55}$$

with coefficients drawn from K ($\boldsymbol{F}^m$ denoting $\boldsymbol{F}$'s composition with itself m times).

A mapping $\psi : \mathcal{A} \to \mathcal{B}$ from one algebra to another is called a **homomorphism** if it is a linear mapping of vector spaces and, in addition, preserves products, that is (for all $X, Y \in \mathcal{A}$)

$$\psi(XY) = \psi(X)\psi(Y). \tag{1.56}$$

It is easy to see that the image (and inverse image) of any subalgebra by ψ is again a subalgebra. ψ is an **isomorphism of algebras**, or **algebraic isomorphism**, if it is a one-to-one, onto homomorphism. Whenever algebras $\mathcal{A}$ and $\mathcal{B}$ are isomorphic, we shall again write $\mathcal{A} \cong \mathcal{B}$ and qualify this statement with 'as vector spaces' if we only wish to assert that $\mathcal{A}$ and $\mathcal{B}$ are vector space isomorphic. It turns out that non-isomorphic algebras *can* have the same dimension, but isomorphic algebras obviously cannot have different dimension. Apart from examples, every concept in this and the previous two paragraphs applies equally well to Lie and Jordan algebras.

As one might expect from the fact that any n-dimensional vector space V^n is isomorphic to ℓ^n (both real or both complex), $\mathcal{A}(V^n) \cong \mathcal{A}(\ell^n)$. To see the isomorphism explicitly, fix a basis $B = \{v_i\}_{i=1}^n$ for V^n and consider the mapping ψ_B that sends $\boldsymbol{F} \in \mathcal{A}(V^n)$ to the $n \times n$ matrix with *columns* given, respectively, by the (unique) expansion coefficients in the basis B of the vectors $\boldsymbol{F}|v_1\rangle, \ldots, \boldsymbol{F}|v_n\rangle$. This ψ_B is an isomorphism of algebras that delivers the **matrix representation of the operator $\boldsymbol{F}$** relative to the basis B. Thus any operator equation in $\mathcal{A}(V^n)$ can be replaced by one involving matrices in $\mathcal{A}(\ell^n)$ and manipulated according to the product rule for $n \times n$ matrices given in (1.46) and (1.47). Furthermore, matrix representations of vectors *in* V^n and operators *on* V^n can readily be shown to 'mesh' relative to a fixed basis B for V^n, in the sense that (for any $|v\rangle \in V^n$, $\boldsymbol{F} \in \mathcal{A}(V^n)$)

$$\psi_B(\boldsymbol{F})\varphi_B|v\rangle = \varphi_B(\boldsymbol{F}|v\rangle), \tag{1.57}$$

where φ_B maps a vector to its matrix representation in ℓ^n relative to B. This guarantees that any equations involving operators acting on vectors in V^n can always be replaced by their matrix counterparts in some convenient basis and manipulated according to the matrix rule given in (1.44) and (1.45). We shall exploit this possibility in the next section.

1.9 Eigenvectors and Eigenvalues

Let $\boldsymbol{F}$ be an operator on V. A *nonzero* vector $|v\rangle \in V$ for which $\boldsymbol{F}|v\rangle = k|v\rangle$ for some $k \in K$ is called an **eigenvector** of $\boldsymbol{F}$, with k the corresponding **eigenvalue**. For example, in the free vector space over R, the operator that maps a function $f(x)$ on the real line to $xf(x)$ has as eigenvectors any blip function f_r $(r \in R)$, with r the corresponding eigenvalue.

The eigenvectors of an operator $\boldsymbol{F}$ that correspond to distinct eigenvalues must be linearly independent. For suppose, for reductio ad absurdem, that the eigenvectors $|v_1\rangle, \ldots, |v_m\rangle$, with corresponding distinct eigenvalues $l_1, \ldots, l_m$, are a linearly dependent set. Then there must be a *smallest* index value $j \leq m$ such that $|v_j\rangle$ is a linear combination of the eigenvectors with index less than j, i.e.,

$$|v_j\rangle = \sum_{i<j\leq m} k_i|v_i\rangle. \tag{1.58}$$

Acting with $\boldsymbol{F}$ on both sides of (1.58), we obtain

$$l_j|v_j\rangle = \sum_{i<j\leq m} k_i l_i|v_i\rangle. \tag{1.59}$$

Multiplying (1.58) by l_j and equating the result with (1.59), the linear independence of the vectors prior to $|v_j\rangle$ in the sequence entails $l_j k_i = k_i l_i$ for all $i < j$. But the eigenvalue l_j differs from all the eigenvalues $\{l_i\}_{i<j}$, so $k_i = 0$ for all $i < j$. Inserting this result back into (1.58), we see that the eigenvector $|v_j\rangle$ must be zero—which is absurd.

The eigenvectors of an operator that correspond to a fixed eigenvalue form a subspace, as a consequence of the linearity of the operator, called an **eigenspace**. The **multiplicity** of an eigenvalue is defined to be the dimension of its corresponding eigenspace. Operators that have eigenvalues with a multiplicity greater than 1 (which must have linearly independent eigenvectors for the same eigenvalue) are called **degenerate**. Finally, if there is a basis B of V^n consisting of eigenvectors of $\boldsymbol{F}$ possibly with the same eigenvalues, then the matrix representation of $\boldsymbol{F}$ relative to B will have $\boldsymbol{F}$'s eigenvalues lying along its main diagonal (i.e., the one from top-left to bottom-right) and 0 off-diagonal terms. It often simplifies matrix computations to work in a basis that 'diagonalizes' a given operator like this.

Nothing said so far guarantees that any particular operator even *has* eigenvectors and eigenvalues. Certainly not every operator on an infinite-dimensional space has an eigenvector. Just consider the transparently linear 'shift' operator $\boldsymbol{F}$ on ℓ^N with action (for all $k_1, k_2, \ldots \in K$):

$$\boldsymbol{F} : \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \cdot \\ \cdot \end{pmatrix} \to \begin{pmatrix} 0 \\ k_1 \\ k_2 \\ \cdot \\ \cdot \end{pmatrix}. \tag{1.60}$$

Clearly *no* vector can be an eigenvector for this $\boldsymbol{F}$, since for any k,

$$\boldsymbol{F}\begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} 0 \\ k_1 \\ k_2 \\ \cdot \\ \cdot \end{pmatrix} = k\begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ \cdot \\ \cdot \end{pmatrix}, \tag{1.61}$$

and the second equality can hold only if all of the k_i's are zero, leaving the zero vector as the only possible candidate for an eigenvector.

For the case of a finite-dimensional vector space V^n, the situation is quite different. Every operator on V^n has an eigenvector if and only if either V^n is complex or n is odd. The last stipulation is more intuitive than it looks. For example, the operator on R^2 that rotates every vector about the origin by some fixed angle has no eigenvectors, because the only vector that is even a candidate for an eigenvector (viz., a vector whose direction fails to change under the rotation) is the zero vector. By contrast, any rotation in R^3 will necessarily leave invariant the vectors that lie along the axis of rotation (though it is less obvious that composing two or more rotations about differing axes—remembering that the composition of two linear operators is again a linear operator—must always leave fixed at least one ray after all rotations are complete).

To establish the 'if and only if' claimed above, consider an arbitrary operator $\boldsymbol{F}$ on V^n. We are asking: Under what conditions does there exist a nonzero vector $|v\rangle$ and number k such that $\boldsymbol{F}|v\rangle = k|v\rangle$? That is, by exploiting the isomorphisms $\mathcal{A}(V^n) \cong \mathcal{A}(\ell^n)$ and $V^n \cong \ell^n$, we want to know when every matrix equation of the form (cf. (1.57))

$$\begin{pmatrix} F_{11} & \ldots & F_{1n} \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ F_{n1} & \ldots & F_{nn} \end{pmatrix}\begin{pmatrix} k_1 \\ \cdot \\ \cdot \\ \cdot \\ k_n \end{pmatrix} = k\begin{pmatrix} k_1 \\ \cdot \\ \cdot \\ \cdot \\ k_n \end{pmatrix} \tag{1.62}$$

has a nonzero column matrix in ℓ^n that solves the equation for some number k.

Recalling the rules of matrix multiplication, (1.62) can be rewritten in the equivalent form (cf. the operator equation $(\boldsymbol{F} - k\boldsymbol{I})|v\rangle = |0\rangle$)

$$\begin{pmatrix} F_{11}-k & . & . & . & F_{1n} \\ . & . & . & . & . \\ . & . & F_{ii}-k & . & . \\ . & . & . & . & . \\ F_{n1} & . & . & . & F_{nn}-k \end{pmatrix} \begin{pmatrix} k_1 \\ . \\ . \\ . \\ k_n \end{pmatrix} = \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \qquad (1.63)$$

which is just a fancy way of writing down a system of n linear homogeneous equations in the n unknowns $\{k_i\}_{i=1}^{n}$. It turns out that a nonzero solution to such a system of equations exists if and only if the 'determinant' of the $n \times n$ 'coefficient matrix' for the system in (1.63) is zero (Lipschutz 1968, Theorem 8.4). The details of how the determinant of a general $n \times n$ matrix is defined need not detain us. All that is important for our purposes is that the determinant of a matrix is a linear combination of products of the entries of the matrix, where each product is formed from n entries of the matrix, exactly one of these n coming from each row of the matrix, and exactly one coming from each column. In particular, the first term in the determinant of a matrix is always the product of the elements along its main diagonal, while the last term is the product of the elements along its other, minor, diagonal multiplied by the coefficient $(-1)^{n(n-1)/2}$. (For example, in the simplest 2×2 case, the determinant of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad (1.64)$$

is $ad - bc$, and simple algebra reveals that the pair of equations

$$\begin{aligned} ax + by &= 0 \\ cx + dy &= 0 \end{aligned} \qquad (1.65)$$

has a nonzero solution only if $ad = bc$, i.e., only if its determinant is zero.) Therefore, the $n \times n$ matrix in (1.63) has determinant zero just in case a certain polynomial of degree n in the unknown variable k (with coefficients that are functions of the F_{ij}'s) has a root, which would then have to be an eigenvalue of $\boldsymbol{F}$.

That there is *always* such a root in the case where $K = C$ follows at once from the fundamental theorem of algebra (Churchill and Brown 1984, Sec. 42) which establishes that, regardless of its coefficients, every polynomial of degree n over the complex numbers has exactly n complex roots (some of which may be the same). So we arrive at the conclusion that every operator on a finite-dimensional complex vector space has at least one eigenvector and eigenvalue. As an added bonus, we have learned that the maximum number of distinct eigenvalues of such an operator is limited by the dimension of the space on which it acts.

In the case where $K = R$, obviously not every polynomial p of degree n over the real numbers has a *real* root, e.g., $x^2 + 1$ has only the roots $\pm\sqrt{-1}$. However, every real polynomial p can be factored into real polynomials of degree 1 or 2 (Anderson and Feil 1995, p. 97). Noting that the degree of the product of any two polynomials is the sum of their degrees, it follows that when n is odd, p will have the form $(ax+b)q$ (where q has even degree) and, therefore, will always have at least one root, viz. $-b/a$. On the other hand, when n is even, it is easy to see that the operator $\boldsymbol{F}$ on R^n given by the matrix with $(-1)^{n(n-1)/2}$ in its upper right-hand corner, 1's along the rest of its minor diagonal, and 0's elsewhere, is such that $\boldsymbol{F} - r\boldsymbol{I}$ has the determinant $r^n + 1$ with no real roots. (When $n = 2$, $\boldsymbol{F}$ is just the operator rotating every vector about the origin by 90°.) Thus, when V^n is real and n even, the existence of eigenvectors and values cannot be taken for granted.

1.10 Linear Functionals

There is another kind of linear mapping that plays just as fundamental a role in the theory of vector spaces as operators do. A **linear functional $\boldsymbol{f}$** on V is a linear mapping $\boldsymbol{f} : V \to K$. The terminology 'functional' derives from the fact that, if we take V to be the free vector space over a set, such $\boldsymbol{f}$'s are, literally, functions of functions. The vector space of all linear functionals on V is called the **dual** of V and is denoted V^*. If $V^* \cong V$, we say V is **self-dual**.

Finite-dimensional vector spaces are self-dual. To see why, let $\{b_i\}_{i=1}^n$ be a basis for V^n. We can then construct a basis $\{\boldsymbol{f}_i\}_{i=1}^n$ for $(V^n)^*$, called the **dual basis** of $\{b_i\}_{i=1}^n$, where

$$\boldsymbol{f}_i|b_j\rangle = \delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j. \end{cases} \tag{1.66}$$

Linear independence of the set $\{\boldsymbol{f}_i\}_{i=1}^n$ is proved by

$$\sum_{i=1}^n k_i \boldsymbol{f}_i = 0 \Rightarrow \sum_{i=1}^n k_i \boldsymbol{f}_i|b_j\rangle = 0 \text{ for all } j = 1 \text{ to } n \tag{1.67}$$

$$\Rightarrow k_j = 0 \text{ for all } j = 1 \text{ to } n \tag{1.68}$$

(abusing notation slightly in the first line by letting '0' denote both the number 0 and the functional mapping all of V to that number). That $\{\boldsymbol{f}_i\}_{i=1}^n$ spans

$(V^n)^*$ is verified by noting that the action of an arbitrary linear functional $\boldsymbol{f}$ is reproduced by the linear combination

$$\sum_{i=1}^{n}(\boldsymbol{f}|b_i\rangle)\boldsymbol{f}_i \tag{1.69}$$

which, by construction, agrees with $\boldsymbol{f}$ in its action on all basis vectors $\{b_i\}_{i=1}^n$.

Infinite-dimensional spaces *fail* to be self-dual. For example, the action of a linear functional on ℓ^N is completely determined by its action on a countable basis. So an element of $(\ell^N)^*$ is defined simply by specifying any countable set of numbers. Mapping each such set of numbers to the column matrix in ℓ that contains them defines an isomorphism between $(\ell^N)^*$ and ℓ. But since ℓ fails to have countable dimension, so does $(\ell^N)^*$, and so it could not be isomorphic to ℓ^N.

1.11 Direct Sums

There are at least two ways to take a pair of sets S_1 and S_2 and build a larger one out of them. We can form their **disjoint union**

$$S_1 \cup_d S_2 \overset{\text{def}}{=} \{(x,n) : n = 1 \text{ or } 2, x \in S_n\} \tag{1.70}$$

or we can form their **Cartesian product**

$$S_1 \times S_2 \overset{\text{def}}{=} \{(s_1,s_2) : s_1 \in S_1 \text{ and } s_2 \in S_2\}. \tag{1.71}$$

Evidently, the cardinality of the disjoint union of two sets is just the sum of their individual cardinalities, whereas the cardinality of their Cartesian product is the product of their cardinalities. The analogues of these two constructions in vector space theory, where the relevant notion of size is *basis* cardinality, i.e., dimension, are the 'direct sum' and 'tensor product' of two vector spaces.

We begin, in the present section, with the simpler construction of the two. The **direct sum** of two vector spaces V_1 and V_2 over K, denoted $V_1 \oplus V_2$, is the vector space (over K) consisting of elements of $V_1 \times V_2$ (whose pairs we write as (v_1, v_2) rather than '$(|v_1\rangle, |v_2\rangle)$') endowed with the following natural vector space operations (for all $k \in K$, $|v_1\rangle, |v_1'\rangle \in V_1$, and $|v_2\rangle, |v_2'\rangle \in V_2$):

$$(v_1, v_2) + (v_1', v_2') = (v_1 + v_1', v_2 + v_2'), \tag{1.72}$$

$$k(v_1, v_2) = (kv_1, kv_2). \tag{1.73}$$

Notice how these operations are defined in terms of operations already available in V_1 and V_2. Despite drawing its elements from the Cartesian product $V_1 \times V_2$,

$V_1 \oplus V_2$ is in fact analogous to disjoint union, because its dimension is the sum of the dimensions of its summands. All that needs to be done is to check that for B_1 a basis in V_1 and B_2 a basis in V_2, the set of all vectors of the form $(b_1, 0)$ with $|b_1\rangle \in B_1$, together with the set of all vectors of the form $(0, b_2)$ with $|b_2\rangle \in B_2$, constitute a basis for $V_1 \oplus V_2$.

The direct sum often makes an appearance when considering two subspaces U and W of a vector space V that satisfy

$$U + W = V,\ U \cap W = \{0\}. \tag{1.74}$$

Such subspaces are said to be **complementary**. Equivalently, U and W are complementary when every $|v\rangle \in V$ is a unique linear combination $|v\rangle = |u\rangle + |w\rangle$ with $|u\rangle \in U$ and $|w\rangle \in W$ (uniqueness being enforced by the second condition in (1.74)). Here, $|u\rangle$ is called the **component** of $|v\rangle$ lying in U; and, similarly, $|w\rangle$ is $|v\rangle$'s component in W. Moreover, to say that $U, W \subseteq V$ are complementary is just to say that $V \cong U \oplus W$. For if U and W are complementary, then we may isomorphically map any $|v\rangle \in V$ to the ordered pair of its components $(u, w) \in U \oplus W$.

1.12 Tensor Products

Consider, again, two vector spaces V_1 and V_2 over K, their Cartesian product $V_1 \times V_2$, and a third vector space W. A mapping $\varphi : V_1 \times V_2 \to W$ is called **bilinear** (cf. (1.42)) if it is linear in both arguments; that is, if (for any $|v_1\rangle, |v_1'\rangle \in V_1$, $|v_2\rangle, |v_2'\rangle \in V_2$, and $k, k' \in K$)

$$\varphi(kv_1 + k'v_1', v_2) = k\varphi(v_1, v_2) + k'\varphi(v_1', v_2), \tag{1.75}$$

$$\varphi(v_1, kv_2 + k'v_2') = k\varphi(v_1, v_2) + k'\varphi(v_1, v_2'). \tag{1.76}$$

(Note that, in general, $k\varphi(v_1, v_2) \neq \varphi(kv_1, kv_2)$.) The pair $\langle W, \varphi\rangle$, where W is a vector space and φ a bilinear mapping $\varphi : V_1 \times V_2 \to W$, is called a **tensor product** of V_1 and V_2 if, whenever $\langle W', \varphi'\rangle$ is any other such pair, there is a *unique* linear mapping $\psi : W \to W'$ such that $\psi \circ \varphi = \varphi'$. This definition is conveniently summarized by the diagram of figure 1.2. One says such a diagram 'commutes' if, by following the arrows along a path through the diagram, and composing the corresponding mappings, the same mapping is obtained between fixed endpoints no matter which path between them is traversed. Thus the vector space W and bilinear mapping φ is a tensor product of V_1 and V_2 if, for any W' and bilinear φ' as shown in figure 1.2, there is a unique linear ψ such that the diagram in that figure commutes.

$$\begin{tikzcd}
V_1 \times V_2 \arrow[r, "\varphi"] \arrow[dr, "\varphi'"'] & W \arrow[d, "\psi \text{ (unique)}"] \\
& W'
\end{tikzcd}$$

FIG. 1.2. Tensor product of two vector spaces

The existence of a tensor product for any two vector spaces V_1 and V_2 is readily established. Fix a basis B_1 for V_1, a basis B_2 for V_2, and let $W = F(B_1 \times B_2)$, the free vector space (over K) on the *set* $B_1 \times B_2$. Recall that a basis for $F(B_1 \times B_2)$ is given by the blip functions $f_{(b_1,b_2)}$ with $(b_1, b_2) \in B_1 \times B_2$. Next, consider any $|v_1\rangle \in V_1$ and $|v_2\rangle \in V_2$. These vectors will each be unique finite linear combinations of basis vectors in their respective spaces

$$|v_1\rangle = \sum_{i=1}^{m} k_i |b_1^i\rangle, \quad |v_2\rangle = \sum_{j=1}^{p} l_j |b_2^j\rangle. \tag{1.77}$$

So we may define a mapping $\varphi : V_1 \times V_2 \to F(B_1 \times B_2)$ by taking $\varphi(v_1, v_2)$ to be the following linear combination of blip functions on $B_1 \times B_2$, built from the coefficients that figure in (1.77):

$$\varphi(v_1, v_2) \overset{\text{def}}{=} \sum_{i,j=1}^{m,p} k_i l_j f_{(b_1^i, b_2^j)}. \tag{1.78}$$

It is easily checked that φ is bilinear. All that remains is to show that $\langle F(B_1 \times B_2), \varphi\rangle$ is a tensor product of V_1 and V_2. Given another W' and (bilinear) φ' as in figure 1.2, clearly a necessary condition for $\psi \circ \varphi = \varphi'$ is that, for any basis pair (b_1, b_2), $\psi(f_{(b_1,b_2)}) = \varphi'(b_1, b_2)$. Since the elements of form $f_{(b_1,b_2)}$ are a basis for $W = F(B_1 \times B_2)$, this fixes the linear mapping ψ uniquely. It is then an easy exercise to show, using bilinearity of φ and φ', that $\psi \circ \varphi = \varphi'$ (i.e., that the action of $\psi \circ \varphi$ and φ' is the same on *all* elements of $V_1 \times V_2$, not just those of form (b_1, b_2)).

Next, we observe that tensor products are unique up to isomorphism. For if $\langle W, \varphi\rangle$ and $\langle W', \varphi'\rangle$ are both tensor products of V_1 and V_2, then there is a unique ψ such that the diagram in figure 1.2 commutes, i.e., $\psi \circ \varphi = \varphi'$. Similarly (interchanging primes and unprimes), there is a unique ψ' such that $\psi' \circ \varphi' = \varphi$. It follows that $(\psi' \circ \psi) \circ \varphi = \varphi$, and therefore setting $\mu = \psi' \circ \psi$ in the diagram of figure 1.3 makes *it* commute. Alternatively, it is clear that diagram 1.3 commutes if μ is set equal to the identity operator on W. Therefore, by uniqueness of μ, $\psi' \circ \psi$ is the identity on W. A similar argument (by symmetry) establishes that $\psi \circ \psi'$ is the identity operator on W', and hence $\psi : W \to W'$ is an isomorphism. It is convenient to call this ψ the **natural isomorphism** between W and W'. For example, if in the previous paragraph's construction, we had chosen a different pair of bases $B_1' \subseteq V_1$ and $B_2' \subseteq V_2$, obtaining another tensor product $\langle F(B_1' \times B_2'), \varphi' \rangle$, then the natural isomorphism between $F(B_1 \times B_2)$ and $F(B_1' \times B_2')$ would be the isomorphism mapping the blip function basis of the former into that of the latter.

$$\begin{tikzcd}
V_1 \times V_2 \arrow[r, "\varphi"] \arrow[dr, "\varphi"'] & W \arrow[d, "\mu \text{ (unique)}"] \\
& W
\end{tikzcd}$$

FIG. 1.3. Step in the argument for the uniqueness of tensor products

The (unique, up to isomorphism) tensor product of V_1 and V_2 is standardly denoted by $V_1 \otimes V_2$, the bilinear mapping from $V_1 \times V_2$ to $V_1 \otimes V_2$ by $\otimes$, and the result of applying the map $\otimes$ to a pair (v_1, v_2) is written as $|v_1\rangle \otimes |v_2\rangle$. (Since $\otimes$ is bilinear, and acts on pairs of vectors, it is similar to the product operation in an algebra, except for one glaring disanalogy. What is it?) Fortunately, it is the general features of the space $V_1 \otimes V_2$ that are important, rather than the specific details of how it might have been constructed (which is why one need not reflect any such details in the notation '$V_1 \otimes V_2$'). We turn next to establishing the most important of these general features.

Consider subsets $S_1 \subseteq V_1$, $S_2 \subseteq V_2$, and the set

$$S = \{|s_1\rangle \otimes |s_2\rangle : |s_1\rangle \in S_1, |s_2\rangle \in S_2\} \subseteq V_1 \otimes V_2. \tag{1.79}$$

We claim that if S_1 and S_2 are each linearly independent, then S will be linearly independent as well. To see this, note that (by a straightforward application of Zorn's lemma) U_1 is contained in some basis $B_1 \subseteq V_1$, and U_2 is contained in some basis $B_2 \subseteq V_2$. Thus, it suffices to establish that the set of elements of form $|b_1\rangle \otimes |b_2\rangle$ is linearly independent (because S is contained therein). From our discussion of the uniqueness of tensor products, we know that the diagram of figure 1.4 below commutes, and that ψ is the natural isomorphism. Since $\psi(f_{(b_1,b_2)}) = |b_1\rangle \otimes |b_2\rangle$, and the blip functions form a basis for $F(B_1 \times B_2)$, the elements of form $|b_1\rangle \otimes |b_2\rangle$ must form a basis in $V_1 \times V_2$, and, hence, be linearly independent. Incidentally, this result makes it clear that the dimension of $V_1 \otimes V_2$ is indeed the product of the dimensions of V_1 and V_2, in analogy with the Cartesian product of two sets.

$$\begin{tikzcd}
V_1 \times V_2 \arrow[r, "\varphi"] \arrow[dr, "\otimes"'] & F(B_1 \times B_2) \arrow[d, "\psi\ \text{(unique)}"] \\
& V_1 \otimes V_2
\end{tikzcd}$$

FIG. 1.4. Argument that $\{|b_1\rangle \otimes |b_2\rangle : |b_1\rangle \in B_1, |b_2\rangle \in B_2\}$ is a basis for $V_1 \otimes V_2$

Next, if we suppose, instead, that the sets S_1 and S_2 span V_1 and V_2, respectively, then it follows that $V_1 \otimes V_2$ is itself spanned by S. By the previous paragraph's argument, the set of all elements of form $|b_1\rangle \otimes |b_2\rangle \in V_1 \otimes V_2$ spans the latter. Thus it suffices to exhibit each such element as a linear combination of the vectors of form $|s_1\rangle \otimes |s_2\rangle$. But since we have by hypothesis

$$|b_1\rangle = \sum_{i=1}^{m} k_i |s_1^i\rangle, \quad |b_2\rangle = \sum_{j=1}^{p} l_j |s_2^j\rangle, \tag{1.80}$$

bilinearity of $\otimes$ immediately yields what we require:

$$|b_1\rangle \otimes |b_2\rangle = \sum_{i,j=1}^{m,p} k_i l_j |s_1^i\rangle \otimes |s_2^j\rangle. \tag{1.81}$$

A vector $|w\rangle \in V_1 \otimes V_2$ is called a **product vector**, or **simple tensor**, if there are vectors $|v_1\rangle \in V_1$ and $|v_2\rangle \in V_1$ such that $|w\rangle$ is the tensor product of $|v_1\rangle$ and $|v_2\rangle$, i.e., $|w\rangle = |v_1\rangle \otimes |v_2\rangle$. By the argument just given, then, the set of all product vectors in $V_1 \otimes V_2$ span it, though they are (of course) not linearly independent. As a result, if one is interested in defining a linear mapping $\psi : V_1 \otimes V_2 \rightarrow W$ by first defining its action on all product vectors and then extending it to the rest of $V_1 \otimes V_2$ by linearity, one must first ensure that the extension will be consistent by checking that ψ preserves linear combinations of product vectors that are again product vectors.

The denotation '$V_1 \otimes V_2$' can be a little confusing, since it might give the wrong impression that every element of $V_1 \otimes V_2$ is a product vector. Not so (and nowhere have we required that the mapping $\otimes$ be onto). A vector in $V_1 \otimes V_2$ that is not a product vector is called **entangled** or **nonseparable**. For example, consider the tensor product $V_1^2 \otimes V_2^2$ of a pair of two-dimensional spaces, with bases $\{b_1, b_1'\}$ and $\{b_2, b_2'\}$, respectively. Since it is a vector space, $V_1^2 \otimes V_2^2$ is closed under vector sums and must contain linear combinations like $|b_1\rangle \otimes |b_2\rangle + |b_1'\rangle \otimes |b_2'\rangle$. To see that this vector is entangled, suppose, to the contrary, that

$$|b_1\rangle \otimes |b_2\rangle + |b_1'\rangle \otimes |b_2'\rangle = |v_1\rangle \otimes |v_2\rangle. \tag{1.82}$$

Inserting the expansions $|v_1\rangle = k_1|b_1\rangle + k_1'|b_1'\rangle$ and $|v_2\rangle = k_2|b_2\rangle + k_2'|b_2'\rangle$ into the right-hand side of (1.82), and simplifying (using the bilinearity of $\otimes$) yields:

$$(k_1k_2-1)|b_1\rangle\otimes|b_2\rangle + k_1k_2'|b_1\rangle\otimes|b_2'\rangle + k_1'k_2|b_1'\rangle\otimes|b_2\rangle + (k_1'k_2'-1)|b_1'\rangle\otimes|b_2'\rangle = |0\rangle\otimes|0\rangle. \tag{1.83}$$

But since the vectors

$$|b_1\rangle \otimes |b_2\rangle, \quad |b_1\rangle \otimes |b_2'\rangle, \quad |b_1'\rangle \otimes |b_2\rangle, \quad |b_1'\rangle \otimes |b_2'\rangle, \tag{1.84}$$

form a basis, there are no values for k_1, k_1', k_2, and k_2' that can solve equation (1.83). For the same reason, all of the following are entangled:

$$|b_1\rangle \otimes |b_2\rangle + |b_1'\rangle \otimes |b_2'\rangle, \tag{1.85}$$

$$|b_1\rangle \otimes |b_2\rangle - |b_1'\rangle \otimes |b_2'\rangle, \tag{1.86}$$

$$|b_1\rangle \otimes |b_2'\rangle + |b_1'\rangle \otimes |b_2\rangle, \tag{1.87}$$

$$|b_1\rangle \otimes |b_2'\rangle - |b_1'\rangle \otimes |b_2\rangle. \tag{1.88}$$

It is easy to see that these vectors form a basis in $V_1^2 \otimes V_2^2$ given that the vectors in (1.84) do. Thus the latter represent only the special case of a **product basis**, consisting entirely of product vectors.

The operation of taking the tensor product of two vector spaces may be iterated to produce tensor product spaces with any number of factors, the simplest case being

$$(V_1 \otimes V_2) \otimes V_3. \tag{1.89}$$

However, it does not actually matter where we place the parentheses. For it is not difficult to see that the space

$$V_1 \otimes (V_2 \otimes V_3) \tag{1.90}$$

also qualifies as a tensor product of $V_1 \otimes V_2$ and V_3, and that the natural isomorphism associates any vector of the form $(|v_1\rangle \otimes |v_2\rangle) \otimes |v_3\rangle$ in the first space above with $|v_1\rangle \otimes (|v_2\rangle \otimes |v_3\rangle)$ in the second. If we always adopt this isomorphism in cases of iterated tensor products, both tensor products of vector spaces and of vectors themselves becomes associative, and we are free to drop all parentheses from tensor product expressions. Alternatively, we could choose to avoid the issue of associativity altogether by defining the n-fold tensor product of the spaces $\{V_j\}_{j=1}^n$ directly, by analogy with the case $n = 2$, rather than considering n-fold products as being produced by iteration. For example, in the case of the three-fold tensor product space $V_1 \otimes V_2 \otimes V_3$, the mappings φ and φ' to W and W' (respectively) in figure 1.2 would now have to be from $V_1 \otimes V_2 \otimes V_3$, and both mappings would be *tri*linear. Moreover, a product basis for $V_1 \otimes V_2 \otimes V_3$ is obtained by taking any three bases for its factor spaces and forming all possible three-fold tensor products out of their elements. Observe, also, that there are the obvious natural isomorphisms between $V_1 \otimes V_2 \otimes V_3$ and the spaces in (1.89) and (1.90).

As one might expect, tensor product distributes over direct sum:

$$U \otimes (V \oplus W) \cong (U \otimes V) \oplus (U \otimes W). \tag{1.91}$$

This is established by associating a vector of form $|u\rangle \otimes (v, w)$ in the space on the left with the unique vector of form $(|u\rangle \otimes |v\rangle, |u\rangle \otimes |w\rangle)$ in the space on the right.

Finally, the **tensor product of two operator algebras** $\mathcal{A}(V_1)$ and $\mathcal{A}(V_2)$ is the algebra $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2)$ whose elements are drawn from the tensor product of the two vector spaces $\mathcal{A}(V_1)$ and $\mathcal{A}(V_2)$ and endowed with the following algebraic product, defined initially between simple tensors by

$$(\boldsymbol{F_1} \otimes \boldsymbol{F_2})(\boldsymbol{G_1} \otimes \boldsymbol{G_2}) = (\boldsymbol{F_1 G_1} \otimes \boldsymbol{F_2 G_2}), \tag{1.92}$$

and then (consistently!) extended to the rest of $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2)$ using bilinearity. Note, in this connection, that the set of all operators of the form $\boldsymbol{F_1} \otimes \boldsymbol{I_2}$, where $\boldsymbol{I_2}$ is the identity operator on V_2, is a subalgebra of $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2)$ (and, similarly, with 1 and 2 interchanged).

If κ_1 is the dimension of V_1 and κ_2 the dimension of V_2, then, as vector spaces, $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2) \cong \mathcal{A}(V_1 \otimes V_2)$, because the space on the right has dimension $(\kappa_1 \kappa_2)^2$ while the space on the left has dimension $\kappa_1^2 \kappa_2^2$. Moreover, these algebras will be *algebraically* isomorphic when both κ_1 and κ_2 are finite. For consider the map $\psi : \mathcal{A}(V_1) \otimes \mathcal{A}(V_2) \to \mathcal{A}(V_1 \otimes V_2)$ whereby $\boldsymbol{F_1} \otimes \boldsymbol{F_2} \in \mathcal{A}(V_1) \otimes \mathcal{A}(V_2)$ is associated with the unique linear operator in $\boldsymbol{F_1} \otimes \boldsymbol{F_2} \in \mathcal{A}(V_1 \otimes V_2)$ with action:

$$(\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)(|v_1\rangle \otimes |v_2\rangle) = \boldsymbol{F}_1|v_1\rangle \otimes \boldsymbol{F}_2|v_2\rangle \tag{1.93}$$

on all product vectors in $V_1 \otimes V_2$. (Note that this definition of $\boldsymbol{F_1} \otimes \boldsymbol{F_2}$ extends consistently and uniquely to all of $V_1 \otimes V_2$ by linearity.) It is not difficult to see that ψ is one-to-one and a homomorphism. To show that ψ is onto (and hence $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2) \cong \mathcal{A}(V_1 \otimes V_2)$), consider the image of $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2)$ by ψ, i.e., $\psi(\mathcal{A}(V_1) \otimes \mathcal{A}(V_2))$. Clearly $\mathcal{A}(V_1) \otimes \mathcal{A}(V_2) \cong \psi(\mathcal{A}(V_1) \otimes \mathcal{A}(V_2))$, and thus the latter constitutes a $\kappa_1^2 \kappa_2^2$-dimensional subspace of the $(\kappa_1 \kappa_2)^2$-dimensional space $\mathcal{A}(V_1 \otimes V_2)$. Since each κ_i is finite, this cannot occur unless the subspaces $\psi(\mathcal{A}(V_1) \otimes \mathcal{A}(V_2))$ and $\mathcal{A}(V_1 \otimes V_2)$ coincide, which means ψ must be onto. (Why does this argument fail in the infinite-dimensional case?)

Everything in the previous two paragraphs generalizes in the obvious way to iterated and n-fold tensor products of operator algebras.

Notes and References

Clear and nearly exhaustive treatments of finite-dimensional vector spaces can be found in Lipschutz (1968) and Halmos (1948). Section 1.6's proof that ℓ has uncountable dimension was communicated to us by John L. Bell. For concise and fairly general discussions of arbitrary vector spaces, associative, and Lie algebras—including exercises and applications to modern physics—see Chs. 9–23 in Geroch (1985). MacLane and Birkhoff (1979) is a classic text on algebras. Jordan algebras were first introduced by one of the co-founders of quantum theory, Pascual Jordan (1932), with the axiomatization of the theory in mind. Shortly thereafter, Jordan's collaboration with von Neumann and Wigner (1934) produced a characterization of a large class of finite-dimensional Jordan algebras, and the literature on Jordan algebras is now voluminous (e.g., see Jacobson (1968)). There are also numerous sources for lattice theory, though Birkhoff's

(1967) is probably the bible. The term ‘entangled’ was originally coined by Schrödinger.

2

INNER PRODUCT SPACES

At the beginning of the previous chapter we motivated abstract vector spaces by recalling some elementary properties of vectors in the plane. However, we deliberately left out reference to the lengths of vectors and the angles that they make with one another. In this chapter we discuss the structure necessary to make generalized 'length' and 'angle' discriminations within a vector space.

Consider two vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ that make an angle ϕ with each other, as in figure 2.1. Define their **dot product** in terms of their coordinates by

$$\overrightarrow{OA} \cdot \overrightarrow{OB} \overset{\text{def}}{=} x_1 x_2 + y_1 y_2. \tag{2.1}$$

Despite appearances, the dot product is actually an intrinsic property of the pair of vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ independent of the coordinates used to represent them. This independence can be seen by rotating the coordinate axes counterclockwise through an angle θ to new primed coordinate axes given by (cf. figure 2.1)

$$x' = x\cos\theta + y\sin\theta, \quad y' = -x\sin\theta + y\cos\theta, \tag{2.2}$$

and then verifying that $\overrightarrow{OA} \cdot \overrightarrow{OB}$ is preserved by the rotation, i.e.,

$$x_1' x_2' + y_1' y_2' = x_1 x_2 + y_1 y_2. \tag{2.3}$$

Clearly $\overrightarrow{OA} \cdot \overrightarrow{OB}$ is also preserved under reflections about the x- and y-axes:

$$x' = x, \; y' = -y \; ; \; x' = -x, \; y' = y. \tag{2.4}$$

Since arbitrary reflections about any axis through the origin are obtained by composing the 'basic' reflections in (2.4) with rotations, $\overrightarrow{OA} \cdot \overrightarrow{OB}$ is invariant under *all* rotations and reflections.

From the dot products of vectors we can recover both their lengths and the angles that they make with one another. By the Pythagorean theorem, the length of $\overrightarrow{OA}$ is

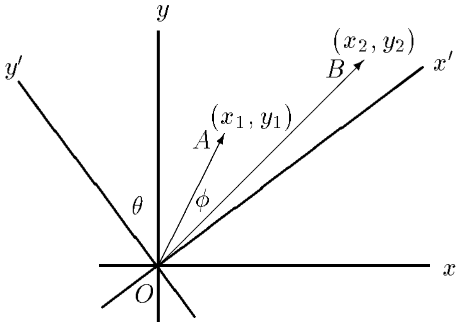

FIG. 2.1. Lengths and angles of vectors

$$||\overrightarrow{OA}|| = \sqrt{\overrightarrow{OA} \cdot \overrightarrow{OA}} = \sqrt{x_1^2 + y_1^2}, \tag{2.5}$$

and similarly for $\overrightarrow{OB}$. Less obvious is the fact that the angle ϕ between $\overrightarrow{OA}$ and $\overrightarrow{OB}$ can be recovered from the formula

$$\cos\phi = \frac{\overrightarrow{OA} \cdot \overrightarrow{OB}}{||\overrightarrow{OA}||\ ||\overrightarrow{OB}||}. \tag{2.6}$$

To see this, note that we may evaluate the expression on the right-hand side of (2.6) with respect to any coordinate axes. Choosing coordinates x'' and y'' so that $\overrightarrow{OB}$ lies along the positive x''-axis, we obtain

$$\overrightarrow{OB} = (||\overrightarrow{OB}||, 0), \quad \overrightarrow{OA} = (||\overrightarrow{OA}|| \cos\phi, ||\overrightarrow{OA}|| \sin\phi), \tag{2.7}$$

and (2.6) is immediate. In particular, ϕ in (2.6) will be $\frac{\pi}{2}$ or $\frac{3\pi}{2}$, and $\overrightarrow{OA}$ and $\overrightarrow{OB}$ perpendicular, exactly when their dot product vanishes.

These observations suggest the idea of an abstract vector space, over the reals *or* complex numbers, on which is defined a 'dot product-like' function that can be used to make intrinsic geometric distinctions within the space. Such spaces are called 'inner product' spaces, and their structure will occupy us for the rest of this chapter.

2.1 Definition

To facilitate algebraic manipulation in inner product spaces, we shall sometimes allow ourselves to denote a vector such as $k_1|v_1\rangle + k_2|v_2\rangle$ by '$|k_1v_1 + k_2v_2\rangle$'. Formally, an **inner product space** V **over** K (sometimes also called a **unitary space** or **pre-Hilbert space**) consists of a vector space V over K with an additional mapping, called the **inner product**, that assigns to any two vectors $|v\rangle, |w\rangle \in V$ a number in K, written $\langle v|w\rangle$. This inner product must possess the following properties (for all $k, k' \in K$ and $|v\rangle, |v'\rangle, |w\rangle, |w'\rangle \in V$):

$$\textbf{conjugate-symmetry}\text{: } \langle v|w\rangle = \langle w|v\rangle^*, \tag{2.8}$$

$$\textbf{positive-definiteness}\text{: } \langle v|v\rangle \geq 0 \text{ with equality if and only if } |v\rangle = |0\rangle, \tag{2.9}$$

$$\textbf{anti-linearity on the left}\text{: } \langle kv + k'v'|w\rangle = k^*\langle v|w\rangle + k'^*\langle v'|w\rangle, \tag{2.10}$$

$$\textbf{linearity on the right}\text{: } \langle v|kw + k'w'\rangle = k\langle v|w\rangle + k'\langle v|w'\rangle. \tag{2.11}$$

The asterisks above, which are redundant if $K = R$, denote complex conjugation, i.e., $c = a + bi \Rightarrow c^* = a - bi$. Also, the requirement of antilinearity is actually redundant since it follows from linearity, conjugate-symmetry, and the fact that for any two complex numbers $(c_1 + c_2)^* = c_1^* + c_2^*$ and $(c_1c_2)^* = c_1^*c_2^*$. And note that, by conjugate-symmetry, $\langle v|v\rangle$ is real, so that the inequality in (2.9) makes sense.

Two immediate consequences of the definition of an inner product should be noted straightaway. First, the zero vector must have zero inner product with all other vectors. (Why?) Second, any two vectors that have the same inner product with all vectors must in fact be the same vector, i.e.,

$$\langle u|v\rangle = \langle u|w\rangle, \text{ for all } |u\rangle \in V \ \Rightarrow\ |v\rangle = |w\rangle. \tag{2.12}$$

For, assuming the antecedent in (2.12) and using linearity, we have $\langle u|v - w\rangle = \langle u|v\rangle - \langle u|w\rangle = 0$ for all $|u\rangle \in V$. In particular, $\langle v - w|v - w\rangle = 0$, which by positive-definiteness requires that $|v\rangle - |w\rangle = |0\rangle$.

The length or **norm** of a vector $|v\rangle$ in an inner product space is defined to be $+\sqrt{\langle v|v\rangle}$, which we shall denote simply by $||v||$. With this 'norm function' on vectors, an inner product space instantiates a more general kind of space. A **normed space** V **over** K is a vector space V over K in which every vector $|v\rangle$

is assigned a number $||v|| \in K$ in such a way that the following properties are satisfied (for all $k \in K$, $|v\rangle, |v'\rangle \in V$):

$$||v|| \geq 0 \text{ with equality if and only if } |v\rangle = |0\rangle, \tag{2.13}$$

$$||kv|| = |k| \; ||v||, \tag{2.14}$$

$$\textbf{triangle inequality}: ||v + w|| \leq ||v|| + ||w||. \tag{2.15}$$

(Here $|k| = +\sqrt{k^*k}$, which reduces to the usual absolute value for k when $K = R$.) The definition of a vector's norm in an inner product space automatically guarantees (2.14), and also (2.13) by positive-definiteness. The triangle inequality can be extracted from the

$$\textbf{Schwarz inequality}: |\langle v|w\rangle| \leq ||v|| \; ||w|| \text{ for all } |v\rangle, |w\rangle \in V. \tag{2.16}$$

If $|w\rangle = |0\rangle$, the Schwarz inequality is trivial. Otherwise, it is proved by invoking positive-definiteness of the inner product to justify writing

$$\left\langle v - \frac{\langle v|w\rangle}{\langle w|w\rangle} w \;\middle|\; v - \frac{\langle v|w\rangle}{\langle w|w\rangle} w \right\rangle \geq 0. \tag{2.17}$$

If we then expand out the left-hand side of (2.17) using linearity and antilinearity, and take the square root of both sides, the Schwarz inequality follows. To obtain from it the triangle inequality, simply substitute $|v\rangle + |w\rangle$ in for *both* $|v\rangle$ and $|w\rangle$ in (2.16), square both sides, and use linearity and antilinearity.

An inner product space *qua* normed space also instantiates a still more general space within which discriminations of distance can be made. A **metric space** is a set, S, on which a distance function, $d(a, b)$, is defined. This function, called the **metric**, maps any two points $a, b \in S$ to a real number and satisfies (for all $a, b, c \in S$):

$$d(a, b) \geq 0, \text{ with equality if and only if } a = b, \tag{2.18}$$

$$d(a, b) = d(b, a), \tag{2.19}$$

$$d(a, c) \leq d(a, b) + d(b, c). \tag{2.20}$$

If we take S to be a normed space V, and define the distance between its 'points' (in this case, vectors) in terms of the norm as $d(|v\rangle, |w\rangle) = ||v - w||$, then (2.18) and (2.19) are automatic, while (2.20) follows from the triangle inequality by making the appropriate substitutions. This definition of distance in V is called the **metric induced by the norm** in V. For vectors in the real plane, it corresponds to taking the distance between the two vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ in figure 2.1 to be the distance between the tips of their arrows, as measured by the length of the vector $\overrightarrow{OA} - \overrightarrow{OB}$ (obtained via the parallelogram law).

2.2 Examples

Consider the free vector space over a set S, $F(S)$, and let $\{f_s\}_{s\in S}$ be the blip function basis (cf. Sections 1.2 and 1.5). Then for any two vectors $|v\rangle, |w\rangle \in F(S)$ (in this case, finite support functions) there are finite subsets $A, B \subseteq S$ and unique expansion coefficients such that

$$|v\rangle = \sum_{s\in A} k_s f_s \text{ and } |w\rangle = \sum_{s\in B} l_s f_s. \tag{2.21}$$

Defining

$$\langle v|w\rangle \stackrel{\text{def}}{=} \sum_{s\in A\cap B} k_s^* l_s \tag{2.22}$$

$F(S)$ is easily seen to be an inner product space. As we have seen, in the case where S is $\{1, \ldots, n\}$, $F(S) = \ell^n$. The inner product of two column matrices in ℓ^n, with respective entries $\{k_i\}_{i=1}^n$ and $\{l_i\}_{i=1}^n$, is usually computed by converting the first matrix into a *row* matrix with the same entries *conjugated*, and then calculating its matrix product with the second matrix according to the rule

$$(k_1^* \cdots k_n^*) \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix} = \sum_{i=1}^n k_i^* l_i \tag{2.23}$$

which is a special case of equation (2.22). In the case of R^2, (2.23) is of course just the dot product of two vectors.

The infinite analogue of the rule in (2.23) is also an inner product for ℓ^N (the case where $S = N$, the natural numbers). However, (2.23) does *not* supply a well-defined inner product on ℓ because the summation on the right in (2.23) will not always be finite for matrices without finite support. (For example, we might have $k_i = l_i = 1$ for all i, in which case we would be trying to 'sum' an infinite number of 1's in (2.23).) To overcome this problem, one considers the largest subspace of ℓ within which the rule in (2.23) *is* defined. To see how this subspace is identified, let us first return to the topic of metric spaces.

An element s in a metric space S is a **limit of an infinite sequence** of points $s_1, s_2, \ldots, s_n, \ldots$ in S if for any $\epsilon > 0$ there exists a positive integer M such that

$$n \geq M \Rightarrow d(s_n, s) < \epsilon. \tag{2.24}$$

(There is no loss in generality in considering only infinite sequences, because finite ones may be regarded as infinite sequences that, at some finite stage in the sequence, become constant. Clearly, then, the limit of a finite sequence, thus understood, is just its last member. For countably infinite sets or sequences, we shall always write $\{s_n\}$ with the understanding that this means '$\{s_n\}_{n=1}^{\infty}$'.) Intuitively, s is a limit of the sequence $\{s_n\}$ if you can get closer and closer to point $s \in S$ by taking points further and further along in the sequence, where the standard of 'closeness' is set by the metric defined on S. (See figure 2.2.) It follows from the properties of a metric (i.e., (2.18)–(2.20)) that a limit of a sequence, if it exists, is in fact unique. For supposing that s' is another limit of $\{s_n\}$, we have

$$d(s', s) = d(s', s_n) + d(s_n, s) \text{ for all } n. \tag{2.25}$$

Since both terms on the right-hand side of (2.25) can be made arbitrarily small for sufficiently large n, it must be the case that $d(s', s) = 0$ and thus $s' = s$. When the limit of a sequence $\{s_n\}$ does exist, the uniqueness of the limit justifies us writing $s_n \to s$, meaning that $\{s_n\}$ approaches s as n goes to infinity. It is easy to see that $s_n \to s$ is equivalent to asserting that $d(s_n, s) \to 0$ in the metric space of the reals R equipped with the usual distance function $d(r_1, r_2) = |r_1 - r_2|$.

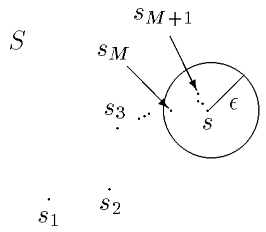

FIG. 2.2. Limit of an infinite sequence in a metric space

In a general metric space, not every sequence need have a limit, since the sequence may 'jump around all over the space without ever settling down'. A necessary condition for a limit to exist is that, beyond a certain point in the sequence, the remaining points in the sequence must get closer and closer to one another (otherwise, how could they be expected to 'home in on' any limit point?). Such sequences are called 'Cauchy', and the precise definition is: a sequence $\{s_n\}$ in a metric space is a **Cauchy sequence** if for any $\epsilon > 0$ there is a positive integer M such that

$$n, m \geq M \Rightarrow d(s_n, s_m) < \epsilon. \tag{2.26}$$

Equivalently, we can express the idea that $\{s_n\}$ is Cauchy by writing $d(s_n, s_m) \to 0$ (in R), regarding $\{d(s_n, s_m)\}$ as a doubly indexed infinite sequence of real numbers.

While being Cauchy is necessary for a sequence to converge, it is far from sufficient. Consider the metric space R (the real line) with the point 0 removed. (Clearly one can always take any metric space and remove points to get a new metric space.) There are plenty of Cauchy sequences in this space which 'want to converge' to the 0, but cannot because 0 is not in the space! By contrast, a metric space is said to be **complete** if every Cauchy sequence converges to a limit within the space. Intuitively, the metric space R, *without* any points deleted, is complete, though we shall not review the formal proof (see Sutherland 1975, Theorem 1.2.9). Moreover, it is a corollary of R's completeness that the metric space of complex numbers C (with distances again measured by absolute values of differences) is complete as well. For if $\{c_n\} \subseteq C$ is Cauchy, then denoting the real and imaginery parts of each c_n by a_n and b_n, we have

$$\sqrt{(a_n - a_m)^2 + (b_n - b_m)^2} = |c_n - c_m| \to 0, \tag{2.27}$$

which entails that both $\{a_n\}$ and $\{b_n\}$ must be Cauchy in R. Thus there exist (unique) $a, b \in R$ such that $a_n \to a$ and $b_n \to b$. Given this, we leave the reader to verify that $c_n = a_n + ib_n \to a + ib$.

Now our problem with defining an inner product on ℓ arose from the fact that, for two infinite column matrices with entries $\{k_i\}$ and $\{l_i\}$, and only finitely many of these entries nonzero, the sum

$$\sum_{i=1}^{\infty} k_i^* l_i \tag{2.28}$$

may not be defined. So when *is* this sum defined? When there is a number $k \in K$ to which the sum converges, in the sense that the following sequence of its 'partial sums' converges to k:

$$\sum_{i=1}^{n} k_i^* l_i \to k. \tag{2.29}$$

So if there is any hope of obtaining an inner product space from ℓ, we must at least restrict ourselves to the subset, which we denote ℓ_2, of column matrices in ℓ that are **square-summable**, i.e., column matrices

$$\begin{pmatrix} k_1 \\ k_2 \\ \vdots \end{pmatrix} \text{ for which } \sum_{i=1}^{\infty} |k_i|^2 \text{ converges in } R. \tag{2.30}$$

In fact, this restriction suffices: ℓ_2 *is* an inner product space with its inner product given by (2.28).

To see why, consider any $k \in K$, and any two square-summable matrices in ℓ_2 with entries $\{k_i\}$ and $\{l_i\}$. Then we have

$$\left|\sum_{i=1}^{n} |kk_i|^2 - \sum_{i=1}^{m} |kk_i|^2\right| = \left|\sum_{i=m+1}^{n} |kk_i|^2\right| \tag{2.31}$$

$$= \sum_{i=m+1}^{n} |kk_i|^2 \tag{2.32}$$

$$= |k|^2 \sum_{i=m+1}^{n} |k_i|^2 \to 0 \tag{2.33}$$

(using the square-summability of $\{k_i\}$ in (2.33)). This shows that the sequence of partial sums

$$\left\{\sum_{i=1}^{n} |kk_i|^2\right\} \tag{2.34}$$

is Cauchy in R, and therefore (since R is complete) that ℓ_2 is closed under multiplication by any number k. Using similar reasoning, we can show that the product in (2.28) is well-defined throughout ℓ_2 by observing that

$$\left|\sum_{i=m+1}^{n} k_i^* l_i\right| \le \sum_{i=m+1}^{n} |k_i^* l_i| = \sum_{i=m+1}^{n} |k_i^*||l_i| \le \sum_{i=m+1}^{n} \left(|k_i|^2 + |l_i|^2\right) \to 0, \tag{2.35}$$

using the triangle inequality in the first step of (2.35) and, in the penultimate step, using $ab \le \max(a^2, b^2) \le a^2 + b^2$ (which holds for any two real numbers). Finally, to see that ℓ_2 is closed under vector sums, note that

$$\left|\sum_{i=m+1}^{n} |k_i + l_i|^2\right| = \left|\sum_{i=m+1}^{n} \left(|k_i|^2 + |l_i|^2 + k_i^* l_i + l_i^* k_i\right)\right| \tag{2.36}$$

$$\leq \sum_{i=m+1}^{n} \left(|k_i|^2 + |l_i|^2 + |k_i^* l_i| + |l_i^* k_i|\right) \to 0, \tag{2.37}$$

where we have invoked the calculation in (2.35) (and the square-summability of $\{k_i\}$ and $\{l_i\}$) when taking the limit in (2.37).

As a last example of inner product spaces, we consider tensor products and direct sums. Any tensor product $V_1 \otimes V_2$ of two vector spaces on which inner products are defined inherits a natural inner product structure from its factor spaces. Writing $|v_1\rangle \otimes |v_2\rangle$ as $|v_1 \otimes v_2\rangle$, define

$$\langle v_1 \otimes v_2 | v_1' \otimes v_2'\rangle \overset{\text{def}}{=} \langle v_1|v_1'\rangle\langle v_2|v_2'\rangle \tag{2.38}$$

and then extend this definition (it extends consistently) to pairs of entangled vectors in $V_1 \otimes V_2$ by assuming $\langle\cdot|\cdot\rangle$ is antilinear in its first argument and linear in its second. The result is an inner product on $V_1 \otimes V_2$. We shall always assume that tensor product spaces have this natural inner product induced on them by the inner products on their factor spaces. Similarly, we shall always assume that $V_1 \oplus V_2$ comes equipped with the inner product defined (completely) by

$$\langle (v_1, v_2)|(v_1', v_2')\rangle \overset{\text{def}}{=} \langle v_1|v_1'\rangle + \langle v_2|v_2'\rangle. \tag{2.39}$$

2.3 Orthogonality and Complete Orthonormal Sets

Two vectors $|v\rangle$ and $|w\rangle$ in an inner product space are said to be **orthogonal** if they have zero inner product. Since $\langle v|w\rangle = \langle w|v\rangle^*$, the relation of orthogonality between vectors is symmetric, in accord with the geometric picture of orthogonality as perpendicularity in R^2 and R^3. Trivially, $|0\rangle$ is orthogonal to every vector, and a quick computation verifies the **generalized Pythagorean theorem**:

$$\{v_i\}_{i=1}^n \text{ are mutually orthogonal } \Rightarrow \left\|\sum_{i=1}^{n} |v_i\rangle\right\|^2 = \sum_{i=1}^{n} \|v_i\|^2. \tag{2.40}$$

We leave the reader the exercise of using (2.40) to show that any set of nonzero mutually orthogonal vectors is linearly independent. (Clearly the converse fails:

vectors in R^2 that make a nonzero acute angle with each other are linearly independent but not orthogonal.)

Now partially order by inclusion the set of all subsets of (nonzero) mutually orthogonal vectors in an inner product space. By a trivial application of Zorn's lemma, this poset has a maximal element, i.e., a mutually orthogonal set not properly contained in any other, called a **complete orthogonal set**. More often than not, one works with complete **orthonormal** sets, which have the additional property that each vector in the set is a **unit vector**, i.e., has norm 1. Dividing any vector $|v\rangle$ by its norm $||v||$, called **normalizing** the vector, produces a unit vector $|v\rangle/||v||$. We shall often denote unit vectors by the letter e, as in '$|e_1\rangle$'.

Evidently, an **orthonormal basis** for an inner product space V is a set of (nonzero) mutually orthonormal vectors that span the vector space V. Every orthonormal basis is also a complete orthonormal set. However, when the dimension of V is uncountably infinite, the converse fails, as illustrated by the following sequence of vectors in ℓ_2:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \ldots \tag{2.41}$$

Clearly no vector in ℓ_2 can be orthogonal to all vectors above unless it is zero, so these vectors define a complete orthonormal set. Yet the vectors in (2.41) do not provide a (vector space) basis, since no matrix in ℓ_2 without finite support is in their span. (They do, however, form an orthonormal basis in ℓ^N.) Thus, the existence of complete orthonormal sets in any inner product space does not guarantee that such spaces always possess orthonormal bases. (However, this situation will improve when we regard ℓ_2, not just as an inner product space, but as a 'Hilbert space'. For orthonormal bases of Hilbert spaces are only required to span a 'dense subset' of the space, and every Hilbert space possesses an orthonormal basis, thus understood. See section !?.)

On the other hand, for any inner product space with countable dimension, there is a well-known procedure, called the **Gram-Schmidt orthonormalization process**, for actually *constructing* an orthonormal basis $\{e_i\}$ out of an arbitrary (not necessarily orthogonal) basis $\{v_i\}$. First, define

$$|e_1\rangle \overset{\text{def}}{=} \frac{|v_1\rangle}{||v_1||}. \tag{2.42}$$

Define the next element of the orthonormal basis by

$$|e_2\rangle \overset{\text{def}}{=} \frac{|w_2\rangle}{\|w_2\|}, \text{ where } |w_2\rangle = |v_2\rangle - \langle v_2|e_1\rangle|e_1\rangle, \tag{2.43}$$

which is well-defined (because $|w_2\rangle \neq |0\rangle$, by the linear independence of $\{v_i\}$), orthogonal to $|e_1\rangle$ (by direct calculation), and such that the spans of the sets $\{e_1, e_2\}$ and $\{v_1, v_2\}$ coincide (because each set lies in the other's span). Continue this process, defining the jth element of the orthonormal basis in terms of the previously defined elements as

$$|e_j\rangle \overset{\text{def}}{=} \frac{|w_j\rangle}{\|w_j\|}, \quad |w_j\rangle = |v_j\rangle - \langle v_j|e_1\rangle|e_1\rangle - \cdots - \langle v_j|e_{j-1}\rangle|e_{j-1}\rangle \tag{2.44}$$

which is again well-defined, orthogonal to all previously defined vectors, and such that the spans of $\{e_1, \ldots, e_j\}$ and $\{v_1, \ldots, v_j\}$ coincide. (The full proof of these facts would proceed by induction on j. Figure 2.3 illustrates the case $j = 3$.) Repeating this process finitely or countably many times, as needed, we obtain a set of vectors $\{e_i\}$ with the same span as $\{v_i\}$, because every vector $|e_j\rangle \in \{e_i\}$ is a (finite) linear combination of vectors $\{v_i\}_{i\leq j} \subseteq \{v_i\}$, and vice-versa. Since we began with a set $\{v_i\}$ that spans the whole space, the constructed orthonormal set $\{e_i\}$ is an orthonormal basis.

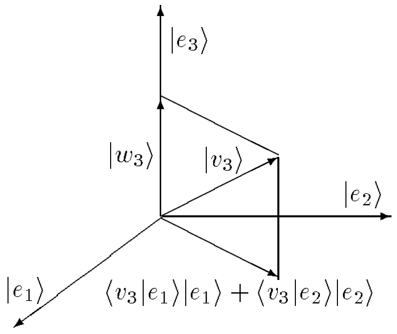
$|e_3\rangle$
$|w_3\rangle$
$|v_3\rangle$
$|e_2\rangle$
$|e_1\rangle$
$\langle v_3|e_1\rangle|e_1\rangle + \langle v_3|e_2\rangle|e_2\rangle$

FIG. 2.3. Gram-Schmidt orthogonalization process for $j = 3$

We end this section with two further remarks about orthonormal bases in inner product spaces with countable dimension. First, inner products determine the coefficients of any vector expanded in terms of an orthonormal basis $\{e_i\}$. Writing

$$|v\rangle = \sum_{j=1}^{n} k_j |e_j\rangle, \tag{2.45}$$

we can, for any index value m, take the inner product of $|e_m\rangle$ with both sides of (2.45), yielding $k_m = \langle e_m|v\rangle$. Second, if we have any two orthonormal bases E_1 and E_2 for V_1 and V_2, then the simple tensors of form

$$\{|e_1 \otimes e_2\rangle : |e_1\rangle \in E_1, |e_2\rangle \in E_2\} \tag{2.46}$$

not only form a basis in $V_1 \otimes V_2$, as usual, but an *orthonormal* basis as a consequence of (2.38). We leave the reader to make out the corresponding claim for the direct sum of two inner product spaces (cf. (2.39)).

2.4 Orthogonal and Orthoclosed Subspaces

Any subspace of an inner product space, i.e., of its underlying vector space, is itself an inner product space, upon restricting the taking of inner products to pairs of vectors in the subspace. U and W are called **orthogonal subspaces**, and we write $U \perp W$, if every vector in the one is orthogonal to every vector in the other. Because $|0\rangle$ is the only vector orthogonal to itself, orthogonal subspaces can only intersect in the zero subspace and are therefore complementary. (Hence, two planes in R^3 that intersect at right angles are *not* orthogonal subspaces.)

For any subspace W, the set of all vectors orthogonal to every vector of W is itself a subspace $W^\perp$ called the **subspace orthogonal to** W. Using this $\perp$ (pronounced: 'perp') operation on subspaces, $U \perp W$ can also be expressed by writing $U \subseteq W^\perp$ or, equivalently, $W \subseteq U^\perp$ (by symmetry). The operation $\perp$ can be applied more than once to a subspace, yielding $(W^\perp)^\perp$ which we write as $W^{\perp\perp}$.

The following properties of $\perp$ follow straight from its definition, and apply for any two subspaces U and W of any inner product space V:

$$0^\perp = V,\ V^\perp = 0,\ W \cap W^\perp = 0, \tag{2.47}$$

$$W \subseteq U \ \Rightarrow\ U^\perp \subseteq W^\perp, \tag{2.48}$$

$$W \subseteq W^{\perp\perp}. \tag{2.49}$$

(Here, 0 denotes the zero subspace.) We shall feel free to invoke any of these properties below without further comment. Note that we cannot strengthen (2.49) to read $W = W^{\perp\perp}$. For a counterexample, take W to be the subspace $\ell^N \subseteq \ell_2$. We have already noted that any vector in ℓ_2 orthogonal to every vector in ℓ^N—in particular, orthogonal to every element of the orthonormal basis for ℓ^N given in (2.41)—must in fact be the zero vector. So $(\ell^N)^\perp = 0$ and $(\ell^N)^{\perp\perp} = \ell_2 \neq \ell^N$.

If a subspace W of an inner product space V *does* satisfy $W = W^{\perp\perp}$ it is called **orthoclosed**. Clearly both 0 and V itself are orthoclosed. And, in view of (2.49), it suffices for W's orthoclosure to establish that $W^{\perp\perp} \subseteq W$. Thus, for any subspace W, $W^\perp$ is orthoclosed because $W \subseteq W^{\perp\perp}$, whence $(W^\perp)^{\perp\perp} \subseteq W^\perp$.

In addition, if a subspace $W \subseteq V$ is finite-dimensional, it is automatically orthoclosed. To see this, it suffices to show that $W^{\perp\perp} \not\subseteq W$ contradicts the finite-dimensionality of W. So suppose that there is a vector $|v\rangle \in W^{\perp\perp}$ that is not in the span of W. Let $\{e_i\}_{i=1}^m$ be an orthonormal basis for (finite-dimensional) W. Then $\{e_i\}_{i=1}^m \cup \{v\}$ is linearly independent and spans a subspace U such that $W \subset U \subseteq W^{\perp\perp}$. Because U itself is an inner product space of finite dimension, we may apply the Gram-Schmidt process to the basis $\{e_i\}_{i=1}^m \cup \{v\}$ in U. The process will leave the first m mutually orthonormal vectors in the basis unchanged and replace $|v\rangle$ with a new normalized vector $|e_{m+1}\rangle$ orthogonal to all the vectors in $\{e_i\}_{i=1}^m$. But there can be no such (nonzero) vector as $|e_{m+1}\rangle$! For it would have to be *both* orthogonal to W, i.e., in $W^\perp$, *and* contained in $U \subseteq W^{\perp\perp}$, yet we know $W^\perp \cap W^{\perp\perp} = 0$.

Finally, the intersection of an arbitrary (not necessarily countable) collection $\{U_\lambda\}_{\lambda\in\Lambda}$ of orthoclosed subspaces is again orthoclosed. For each $U_{\lambda'}^\perp$ is orthogonal to the subspace $\bigcap_{\lambda\in\Lambda} U_\lambda$, which means we have

$$U_{\lambda'}^\perp \subseteq \left(\bigcap_{\lambda\in\Lambda} U_\lambda\right)^\perp \quad\Rightarrow\quad \left(\bigcap_{\lambda\in\Lambda} U_\lambda\right)^{\perp\perp} \subseteq U_{\lambda'}^{\perp\perp} = U_{\lambda'}. \tag{2.50}$$

Since this is true for all λ', $\left(\bigcap_{\lambda\in\Lambda} U_\lambda\right)^{\perp\perp} \subseteq \bigcap_{\lambda\in\Lambda} U_\lambda$ and the intersection is orthoclosed as well.

2.5 The Lattice of Orthoclosed Subspaces

We saw in Section 1.4 that the subspaces of a vector space form a complete atomic lattice. The same is true for the orthoclosed subspaces of an inner product space V. Ordering them by inclusion, we obtain a poset with minimum element 0, maximum element V and with all (necessarily orthoclosed) one-dimensional subspaces as atoms. The meet of any collection of orthoclosed subspaces is clearly just their intersection (which is orthoclosed, by the argument of (2.50)). But what of their join? It must at least contain their span, but that will not necessarily produce an orthoclosed subspace. What we seek, then, is the smallest orthoclosed subspace containing their span. But for any subspace W, the smallest orthoclosed subspace that contains W is just $W^{\perp\perp}$. For if U is another such subspace containing W, we have

$$W \subseteq U \;\Rightarrow\; U^{\perp} \subseteq W^{\perp} \;\Rightarrow\; W^{\perp\perp} \subseteq U^{\perp\perp} = U. \tag{2.51}$$

Thus for the join of an arbitrary collection of orthoclosed subspaces, we must take the orthoclosure of their span. For just two subspaces U and W, we shall denote their orthoclosed span, $(U + W)^{\perp\perp}$, by $U \vee W$ (i.e., we use the same symbol as that used for the join in an abstract lattice).

Due to the presence of the $\perp$ operation on subspaces, the lattice of orthoclosed subspaces of an inner product space has some additional structure. A lattice $\mathcal{L}$ (with minimum and maximum elements 0 and 1) is called **orthocomplemented**, or an **ortholattice**, if it is equipped with an operation ${}^{\perp} : \mathcal{L} \to \mathcal{L}$ (called an orthocomplement on $\mathcal{L}$) that satisfies (for all $a, b \in \mathcal{L}$):

$$a \wedge a^{\perp} = 0, \;\; a \vee a^{\perp} = 1, \tag{2.52}$$

$$a \leq b \;\Rightarrow\; b^{\perp} \leq a^{\perp}, \tag{2.53}$$

$$a^{\perp\perp} = a. \tag{2.54}$$

(Again, we have allowed ourselves to use the symbol ${}^{\perp}$ to denote both the orthocomplement in an abstract ortholattice and the concrete operation of 'take the orthogonal subspace' on subspaces.) To show that the lattice of orthoclosed subspaces does indeed form an ortholattice under ${}^{\perp}$, all that remains to check is the second property in (2.52). Let W be orthoclosed and set $X = W \vee W^{\perp}$ (so X is also orthoclosed). We must show $X = V$. Clearly $W \subseteq X$ and $W^{\perp} \subseteq X$, so we have $X^{\perp} \subseteq W^{\perp}$ and $X^{\perp} \subseteq W^{\perp\perp}$. And since $W^{\perp} \cap W^{\perp\perp} = 0$, it follows that $X^{\perp} = 0$. Therefore, $X^{\perp\perp} = X = V$ as required. For an abstract ortholattice we shall always write $\mathcal{L}_{\perp}$, whereas for a concrete ortholattice of orthoclosed subspaces we shall write $\mathcal{L}_{\perp}(V)$.

It is a simple matter to check that, in any ortholattice $\mathcal{L}_{\perp}$,

$$\textbf{de Morgan's laws}\text{: } (a \vee b)^{\perp} = a^{\perp} \wedge b^{\perp} \text{ and } (a \wedge b)^{\perp} = a^{\perp} \vee b^{\perp} \tag{2.55}$$

hold (and, in particular, these 'laws' hold in $\mathcal{L}_{\perp}(V)$). For example, to obtain the first law, begin with $a \leq a \vee b$ and $b \leq a \vee b$. Then $(a \vee b)^{\perp} \leq a^{\perp}$ and $(a \vee b)^{\perp} \leq b^{\perp}$, whence $(a \vee b)^{\perp} \leq a^{\perp} \wedge b^{\perp}$. To obtain the reverse, i.e., $(a \vee b)^{\perp} \geq a^{\perp} \wedge b^{\perp}$, start with $a^{\perp} \wedge b^{\perp} \leq a^{\perp}$ and $a^{\perp} \wedge b^{\perp} \leq b^{\perp}$. Then we have $a \leq (a^{\perp} \wedge b^{\perp})^{\perp}$ and $b \leq (a^{\perp} \wedge b^{\perp})^{\perp}$, whence $a \vee b \leq (a^{\perp} \wedge b^{\perp})^{\perp}$ which, in turn, entails $a^{\perp} \wedge b^{\perp} \leq (a \vee b)^{\perp}$.

An ortholattice $\mathcal{L}_{\perp}$ is called an **orthomodular lattice** if, in addition, it satisfies (for all $a, b \in \mathcal{L}_{\perp}$):

$$a \leq b \;\Rightarrow\; b = a \vee (b \wedge a^{\perp}). \tag{2.56}$$

Evidently distributive ortholattices are necessarily orthomodular, but not conversely. It turns out that $\mathcal{L}_{\perp}(V)$ will be orthomodular whenever the span of two orthogonal subspaces in $\mathcal{L}_{\perp}(V)$ is itself orthoclosed, i.e., if $\mathcal{L}_{\perp}(V)$ satisfies the condition

$$U, W \in \mathcal{L}_{\perp}(V), \quad U \perp W \quad \Rightarrow \quad U + W \in \mathcal{L}_{\perp}(V). \tag{2.57}$$

To see that (2.57) is sufficient for the orthomodularity of $\mathcal{L}_{\perp}(V)$, consider two subspaces $U \subseteq W$ in $\mathcal{L}_{\perp}(V)$ and set $X = U \vee (W \cap U^{\perp})$. Evidently, $X \subseteq W$, but we need to show that $X = W$. Reworking the left-hand side of

$$(W \cap U^{\perp}) \vee (W \cap U^{\perp})^{\perp} = V = W \vee W^{\perp} \tag{2.58}$$

using de Morgan's laws, we obtain $X \vee W^{\perp} = W \vee W^{\perp}$. Because $X \subseteq W$, both X and W are orthogonal to $W^{\perp}$. Thus (2.57) licenses us to write $X + W^{\perp} = W + W^{\perp}$ from which it follows easily that $X = W$. Note that, when V is finite-dimensional, (2.57) is automatic, and hence $\mathcal{L}_{\perp}(V)$ orthomodular. Moreover, we shall see in section ?? that (2.57) also holds for the inner product spaces employed in quantum theory (viz., 'Hilbert spaces'). When V is such that $\mathcal{L}_{\perp}(V)$ is orthomodular, we shall denote the latter by $\mathcal{L}_{\perp_o}(V)$, and always denote an abstract orthomodular lattice by $\mathcal{L}_{\perp_o}$.

A sublattice of an ortholattice $\mathcal{L}_{\perp}$ is any subset closed under meets, joins and orthocomplements. In fact, taking the complements of both sides of de Morgan's laws, join is expressible in terms of meet and orthocomplement, and meet is expressible in terms of join and orthocomplement. Thus we need only demand that a sublattice of $\mathcal{L}_{\perp}$ be a subset closed under orthocomplements and either joins *or* meets. And, of course, any sublattice of $\mathcal{L}_{\perp_o}$ will itself be orthomodular.

The sublattice generated by a subset S of an ortholattice $\mathcal{L}_{\perp}$ is the smallest sublattice of $\mathcal{L}_{\perp}$ containing S, obtained by closing S under the operations of $\mathcal{L}_{\perp}$. For example, the sublattice generated by two distinct, nonorthogonal rays $A, B \in \mathcal{L}_{\perp_o}(R^3)$ is depicted in Figure 2.4. Note that the sublattice of $\mathcal{L}(R^3)$ generated by A and B would have seven fewer subspaces, since there is no requirement in

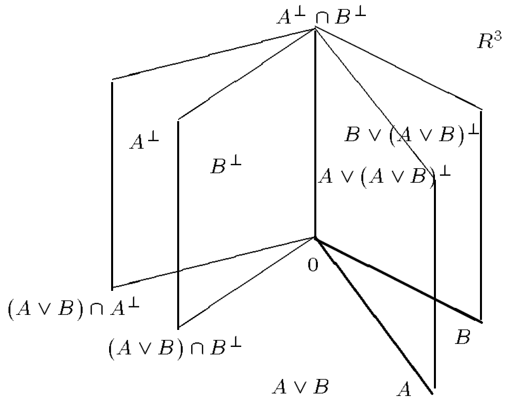
$A^\perp \cap B^\perp$
R^3
$A^\perp$
$B^\perp$
$B \vee (A \vee B)^\perp$
$A \vee (A \vee B)^\perp$
0
$(A \vee B) \cap A^\perp$
$(A \vee B) \cap B^\perp$
B
$A \vee B$
A

FIG. 2.4. Sublattice of $\mathcal{L}_{\perp_o}(R^3)$ generated by A and B

$\mathcal{L}(R^3)$ to close under orthocomplements. In general, it is a complicated affair to determine the structure of a fully generated sublattice of $\mathcal{L}_{\perp_o}(V)$ given generators for the sublattice. But the situation simplifies considerably when the generators of the sublattice are all 'mutually compatible' (our next subject).

2.6 Compatible Subspaces and Boolean Algebras

In addition to orthogonality, there is also a further symmetric relation between elements of an orthomodular lattice $\mathcal{L}_{\perp_o}$. Elements $a, b \in \mathcal{L}_{\perp_o}$ are said to be **compatible**, and one writes $a \leftrightarrow b$, exactly when there exist three mutually orthogonal (not necessarily nonzero) elements $x, y, z \in \mathcal{L}_{\perp_o}$ such that

$$a = x \vee z, \quad b = y \vee z. \tag{2.59}$$

$\mathcal{L}_{\perp_o}$'s orthomodularity automatically ensures that:

$$a \leq b \;\Rightarrow\; a \leftrightarrow b. \tag{2.60}$$

(So, while there was nothing stopping us from defining compatibility between elements in an arbitrary—not necessarily orthomodular—ortholattice, (2.60) would not have been guaranteed.) Similarly, two subspaces $U, W \subseteq \mathcal{L}_{\perp_o}(V)$ that satisfy condition (2.59) are called **compatible subspaces**. The picture is that U and

W are orthogonal, except possibly for an overlap within some third subspace. Thus orthogonal subspaces in $\mathcal{L}_{\perp_o}(V)$ are always compatible, but the converse is false. For example, two planes in $\mathcal{L}_{\perp_o}(R^3)$ at right angles, while not orthogonal, *are* compatible (the overlap subspace being the unique ray in which the planes intersect). As another example, note that $A^\perp \cap B^\perp$ in Figure 2.4 is compatible with every other subspace in that figure, but the two planes on the left are incompatible (as are the two on the right).

For any compatible pair $a, b \in \mathcal{L}_{\perp_o}$, it turns out that the three mutually orthogonal elements in (2.59) are uniquely fixed by a and b via the following equations:

$$x = a \wedge b^\perp, \tag{2.61}$$

$$y = a^\perp \wedge b, \tag{2.62}$$

$$z = a \wedge b. \tag{2.63}$$

To derive (2.61), first observe that

$$x \leq y^\perp, x \leq z^\perp \quad \Rightarrow \quad x \leq y^\perp \wedge z^\perp = (y \vee z)^\perp = b^\perp. \tag{2.64}$$

However, $x \leq a$, and thus, invoking the final inequality in (2.64), $x \leq a \wedge b^\perp$. But we also have the implications

$$\begin{aligned} x \leq a \wedge b^\perp \leq b^\perp \Rightarrow b &\leq x^\perp, & (2.65)\\ \Rightarrow x^\perp &= b \vee (x^\perp \wedge b^\perp), & (2.66)\\ \Rightarrow x &= b^\perp \wedge (x \vee b) & (2.67) \end{aligned}$$

(using orthomodularity and de Morgan's law in (2.66) and (2.67)). Since $x \vee b$ equals $a \vee b$ (after all, they both equal $x \vee y \vee z$), it follows from (2.67) that $x = b^\perp \wedge (a \vee b)$. Invoking $x \leq a \wedge b^\perp$ one last time,

$$x \leq a \wedge b^\perp \leq (a \vee b) \wedge b^\perp = x, \tag{2.68}$$

which entails (2.61). A parallel argument for (2.62) is obtained by interchanging the roles of x and y (and, therefore, a and b). (2.63)'s proof, on the other hand, calls for a few more ortholattice gymnastics:

$$\begin{aligned} x \leq z^\perp, y \leq z^\perp &\Rightarrow x \vee y \leq z^\perp, & (2.69)\\ &\Rightarrow z^\perp = (x \vee y) \vee [z^\perp \wedge (x \vee y)^\perp] \text{ (orthomodularity),} & (2.70) \end{aligned}$$

$$\Rightarrow z = (x \vee y)^{\perp} \wedge (z \vee x \vee y) \text{ (de Morgan)}, \tag{2.71}$$
$$\Rightarrow z = [(a \wedge b^{\perp}) \vee (a^{\perp} \wedge b)]^{\perp} \wedge (a \vee b) \text{ (2.61), (2.62)}, \tag{2.72}$$
$$\Rightarrow z = (a^{\perp} \vee b) \wedge (a \vee b^{\perp}) \wedge (a \vee b) \text{ (de Morgan)}. \tag{2.73}$$

Now $a \wedge b$, since it is $\leq$ both a and b, is $\leq$ all three of $a^{\perp} \vee b$, $a \vee b^{\perp}$, and $a \vee b$. It then follows (using (2.73), and the fact that z is $\leq$ both a and b in the final step) that

$$a \wedge b \leq (a^{\perp} \vee b) \wedge (a \vee b^{\perp}) \wedge (a \vee b) = z \leq a \wedge b, \tag{2.74}$$

completing the proof of (2.63).

We can now establish the following two alternative characterizations of the compatibility of a pair $a, b \in \mathcal{L}_{\perp_o}$:

$$a \leftrightarrow b \quad \Leftrightarrow \quad a = (a \wedge b) \vee (a \wedge b^{\perp}), \tag{2.75}$$
$$\Leftrightarrow \quad b = (b \wedge a) \vee (b \wedge a^{\perp}). \tag{2.76}$$

We need only verify the first equivalence (since the second then follows by the symmetry of compatibility). The implication '$\Rightarrow$' follows immediately from (2.59), (2.61), and (2.63). For '$\Leftarrow$', note that we can use orthodularity to write

$$b = (b \wedge a) \vee \left[b \wedge (b \wedge a)^{\perp}\right]. \tag{2.77}$$

Because $b \wedge (b \wedge a)^{\perp}$, $a \wedge b^{\perp}$, and $a \wedge b$ are mutually orthogonal, (2.77), together with $a = (a \wedge b) \vee (a \wedge b^{\perp})$, jointly entail $a \leftrightarrow b$. The equivalences in (2.75) and (2.76) make it easy to see that

$$a \leftrightarrow b \quad \Rightarrow \quad \text{the elements } \{a, b, a^{\perp}, b^{\perp}\} \text{ are pairwise compatible.} \tag{2.78}$$

As we have seen, lattices need not be distributive. But if three elements $a, b, c \in \mathcal{L}_{\perp_o}$ are mutually compatible, then the distributive laws *will* hold with respect to those elements. Consider distributivity of $\wedge$ over $\vee$ first. Regardless of compatibility, we always have

$$(a \wedge b) \leq a \wedge (b \vee c) \text{ and } (a \wedge c) \leq a \wedge (b \vee c) \tag{2.79}$$

and, so, it must be the case that $(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee c)$. To obtain the reverse inclusion, note that $a \leftrightarrow b$ and $a \leftrightarrow c$ imply

$$a \wedge (b \vee c) = a \wedge \left[(b \wedge a) \vee (b \wedge a^{\perp}) \vee (c \wedge a) \vee (c \wedge a^{\perp})\right]. \tag{2.80}$$

Setting

$$s = (b \wedge a) \vee (c \wedge a) \text{ and } t = (b \wedge a^{\perp}) \vee (c \wedge a^{\perp}), \tag{2.81}$$

we need to show that $a \wedge (s \vee t) \leq s$. But observe that $s \leq a \leq t^{\perp}$ and, in particular, that s and t, being orthogonal, are compatible. It follows from (2.68) (substituting s for a and t for b) that $(s \vee t) \wedge t^{\perp} = s \wedge t^{\perp}$. And so we obtain

$$a \wedge (s \vee t) \leq t^{\perp} \wedge (s \vee t) = s \wedge t^{\perp} \leq s \tag{2.82}$$

and $\wedge$ distributes over $\vee$ as claimed. We leave the reader to verify that $\vee$ also distributes over $\wedge$. (Just use the fact that if $\{a, b, c\}$ are pairwise compatible, so are $\{a^{\perp}, b^{\perp}, c^{\perp}\}$ (cf. (2.78)), together with $\wedge$'s distribution over $\vee$, and de Morgan's laws.)

A **Boolean algebra** is a distributive ortholattice. Thus, by the previous paragraph, any sublattice of $\mathcal{L}_{\perp_o}$ consisting entirely of mutually compatible subspaces is a Boolean subalgebra of $\mathcal{L}_{\perp_o}$. In fact, somewhat more is true: the sublattice generated by any subset of $\mathcal{L}_{\perp_o}$ is Boolean if and only if the generating subset is mutually compatible. For the 'only if' part of this claim, consider any two generators a and b. Invoking distributivity within the generated sublattice, we have

$$a = a \wedge 1 = a \wedge (b \vee b^{\perp}) = (a \wedge b) \vee (a \wedge b^{\perp}) \tag{2.83}$$

thus $a \leftrightarrow b$. (This also establishes that if a sublattice of $\mathcal{L}_{\perp_o}$ is Boolean, all its elements must be mutually compatible; for a sublattice is nothing but the sublattice generated by its members.) For the converse 'if' part of the claim, recall that if the generators of a sublattice are all compatible, then they will be compatible with each other's, and their own, orthocomplements. Moreover, each generator a will be compatible with the join of any two other generators $b_1 \vee b_2$. For, since $a \leftrightarrow b_1$ and $a \leftrightarrow b_2$ by hypothesis,

$$b_1 = (b_1 \wedge a) \vee (b_1 \wedge a^{\perp}) \text{ and } b_2 = (b_2 \wedge a) \vee (b_2 \wedge a^{\perp}). \tag{2.84}$$

Using the pairwise compatibility of $\{b_1, b_2, a, a^{\perp}\}$, and therefore freely employing distributivity, it follows that

$$\begin{aligned} b_1 \vee b_2 &= [(b_1 \wedge a) \vee (b_2 \wedge a)] \vee \left[(b_1 \wedge a^{\perp}) \vee (b_2 \wedge a^{\perp})\right] && (2.85) \\ &= [(b_1 \vee b_2) \wedge a] \vee \left[(b_1 \vee b_2) \cap a^{\perp}\right] && (2.86) \end{aligned}$$

and $a \leftrightarrow (b_1 \vee b_2)$. Thus, because the generators are not just compatible with each other, but with each other's complements and joins, the sublattice they generate must consist entirely of mutually compatible subspaces, and therefore be Boolean.

Since Figure 2.4 contains incompatible subspaces, the sublattice of $\mathcal{L}_{\perp_o}(R^3)$ depicted therein fails to be Boolean. On the other hand, that sublattice consists of seven distinct overlapping Boolean subalgebras. For example, focusing just on the Boolean subalgebras generated by A alone and B alone, they intersect in the (trivial) Boolean subalgebra $\{0, R^3\}$, as shown in figure 2.5. More generally, any sublattice of $\mathcal{L}_{\perp_o}$ is the union of all its Boolean subalgebras, since any element of the sublattice will be contained in the Boolean subalgebra that element generates, which itself must be contained in the given sublattice. In addition, it is clear that any two compatible subspaces in a sublattice will always lie within one of the sublattice's Boolean subalgebras. This fact suggests an alternative way of viewing a sublattice of $\mathcal{L}_{\perp_o}$, viz., as a collection of overlapping Boolean algebras in which any two elements are contained in one of those algebras. Viewed in this way, $\mathcal{L}_{\perp_o}$ instantiates a 'partial Boolean algebra', which is a structure that can be defined without reference to an underlying nondistributive orthomodular lattice.

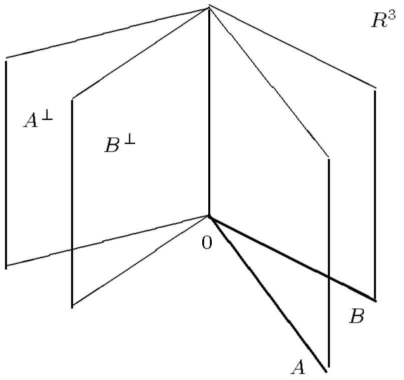

FIG. 2.5. Subalgebra of $\tilde{\mathcal{B}}(R^3)$ generated by A and B

Before proceeding to the formal definition of a partial Boolean algebra, it is important to note that a Boolean algebra may itself be defined without reference to lattices or posets. First note that in any Boolean algebra, i.e., distributive ortholattice, the following identities automatically hold:

$$a \vee b = b \vee a \qquad a \wedge b = b \wedge a, \tag{2.87}$$

$$a \vee (b \vee c) = (a \vee b) \vee c, \qquad a \wedge (b \wedge c) = (a \wedge b) \wedge c, \tag{2.88}$$

$$(a \vee b) \wedge b = b, \qquad (a \wedge b) \vee b = b, \tag{2.89}$$

$$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c), \qquad (a \wedge b) \vee c = (a \vee c) \wedge (b \vee c), \tag{2.90}$$

$$a \vee a^{\perp} = 1, \qquad a \wedge a^{\perp} = 0. \tag{2.91}$$

Now let $\langle \mathcal{B}, \wedge, \vee, \perp, 0, 1 \rangle$ be a structure in which $\mathcal{B}$ is a set with designated elements 0 and 1, $\wedge$ and $\vee$ are binary operations on the set, and $\perp$ is a unary operation. It turns out that if the identities (2.87)–(2.91) hold in this structure, and we define the relation $\leq$ on $\mathcal{B}$ by

$$a \leq b \quad \Leftrightarrow \quad a \wedge b = a, \tag{2.92}$$

then $\mathcal{B}$ must be a distributive ortholattice in which $\wedge$, $\vee$, and $\perp$ are the meet, join, and orthocomplementation operations, and 0 and 1 are the minimum and maximum elements. The complete proof of this claim is left to the reader, but to illustrate the strategy, we shall show that $a \leq b \Rightarrow b^{\perp} \leq a^{\perp}$. First observe that for any $c \in \mathcal{B}$,

$$c \wedge 1 = 1 \wedge c = (c \vee c^{\perp}) \wedge c = (c^{\perp} \vee c) \wedge c = c, \tag{2.93}$$

$$c \vee 0 = 0 \vee c = (c \wedge c^{\perp}) \vee c = (c^{\perp} \wedge c) \vee c = c. \tag{2.94}$$

Using the definition of the partial ordering in (2.92), equations (2.93) and (2.94), and the identities (2.87)–(2.91), we can then construct the following sequence of entailments:

$$\begin{aligned} a \leq b &\Rightarrow a \wedge b = a, & (2.95)\\ &\Rightarrow (a \wedge b) \vee a^{\perp} = a \vee a^{\perp} = 1, & (2.96)\\ &\Rightarrow (a \vee a^{\perp}) \wedge (b \vee a^{\perp}) = 1, & (2.97)\\ &\Rightarrow 1 \wedge (b \vee a^{\perp}) = b \vee a^{\perp} = 1, & (2.98)\\ &\Rightarrow (b \vee a^{\perp}) \wedge b^{\perp} = 1 \wedge b^{\perp} = b^{\perp}, & (2.99)\\ &\Rightarrow (b \wedge b^{\perp}) \vee (a^{\perp} \wedge b^{\perp}) = b^{\perp} & (2.100)\\ &\Rightarrow 0 \vee (a^{\perp} \wedge b^{\perp}) = a^{\perp} \wedge b^{\perp} = b^{\perp}, & (2.101)\\ &\Rightarrow b^{\perp} \wedge a^{\perp} = b^{\perp} \;\; \Rightarrow \;\; b^{\perp} \leq a^{\perp}. & (2.102) \end{aligned}$$

A **partial Boolean algebra** is a structure $\langle \tilde{\mathcal{B}}, \leftrightarrow, \wedge, \vee, \perp, 0, 1 \rangle$ in which $\tilde{\mathcal{B}}$ is a set containing 0 and 1, $\leftrightarrow$ is a reflexive, symmetric relation on $\tilde{\mathcal{B}}$ satisfying

$a \leftrightarrow 1$ for all $a \in \tilde{\mathcal{B}}$, $\perp$ is a unary operation on $\tilde{\mathcal{B}}$, and both $\wedge$ and $\vee$ are binary *partial* operations defined on $\{(a,b) \in \tilde{\mathcal{B}} : a \leftrightarrow b\}$ and satisfying, for any $a \leftrightarrow b$:

$$a \leftrightarrow b^{\perp}, \quad (a \vee b) \leftrightarrow a, \quad (a \wedge b) \leftrightarrow a, \text{ and} \tag{2.103}$$

$$\text{the substructure of } \tilde{\mathcal{B}} \text{ generated by } \{a,b\} \text{ is a Boolean algebra.} \tag{2.104}$$

(There is, of course, redundancy in this definition because of de Morgan's laws.) Taking $\leftrightarrow$, $\wedge$, $\vee$, and $\perp$ to have their usual meaning in $\mathcal{L}_{\perp_o}$, it is clear that every sublattice of $\mathcal{L}_{\perp_o}$ is a partial Boolean algebra, and, in particular, all sublattices of $\mathcal{L}_{\perp_o}(V)$ are. For the partial Boolean algebra of *all* orthoclosed subspaces of an inner product space, we shall write $\tilde{\mathcal{B}}(V)$.

Evidently, a subalgebra of a partial Boolean algebra $\tilde{\mathcal{B}}$ is just a subset of $\tilde{\mathcal{B}}$ closed under $\perp$ and the partial operations $\wedge$ and $\vee$. Moreover, the subalgebra generated by a subset is obtained just by closing under these operations. Thus, the subalgebra of $\tilde{\mathcal{B}}(V)$ generated by a set of subspaces is obtained by closing the set under orthocomplements and the meets (or joins) of *compatible* subspaces. It follows that while every sublattice of $\mathcal{L}_{\perp_o}(V)$ is a (partial Boolean) subalgebra of $\tilde{\mathcal{B}}(V)$, the converse fails because closing under meets and joins of compatible subspaces will not guarantee closure under meets and joins of arbitrary subspaces. For example, *all* the elements of the subalgebra of $\tilde{\mathcal{B}}(R^3)$ generated by two distinct rays A and B are shown in Figure 2.5, whereas we saw in Figure 2.4 that the sublattice of $\mathcal{L}_{\perp_o}(R_3)$ that they generate is somewhat larger.

2.7 Isomorphisms and Unitary Operators

Two inner product spaces over the same set of numbers are isomorphic if there is a vector space isomorphism between them that preserves inner products, i.e., that maps pairs of vectors to pairs of vectors in the image space with the same inner product. (There is slight redundancy in this definition, because any (not necessarily linear) mapping $\varphi : V \to W$ that preserves inner products is necessarily one-to-one—and the latter is, of course, built into the definition of a vector space isomorphism. To show that φ is one-to-one, observe that $\varphi|v\rangle = \varphi|v'\rangle$ implies $\langle u|v\rangle = \langle u|v'\rangle$ for all $|u\rangle \in V$.) For example, the vector space isomorphism discussed in Section 1.7 that maps the vectors in a real (or complex) n-dimensional space to their column matrix representations in R^n (or C^n) preserves inner products, so that this isomorphism is an inner product isomorphism as well. For another example, given two subspaces $U, W \subseteq V$ of an inner product space, $V \approx U \oplus W$ (as inner product spaces) if and only if $U \perp W$ and $U + W = V$.

Obviously, if one knows the inner products between all vectors, the norm of any vector is fixed. Less obvious, but equally true, is the converse: that the norms of all vectors in an inner product space fix the inner product between any two vectors. If the space is complex, a tedious (but elementary) calculation reveals that

$$\langle v|w\rangle = \frac{1}{4}\left(||v+w||^2 - ||v-w||^2 + i||v+iw||^2 - i||v-iw||^2\right), \qquad (2.105)$$

and the appropriate expression for a real inner product space is obtained by simply dropping the last two terms in (2.105). It follows that, given a linear mapping $\varphi : V \to W$, φ will preserve norms if and only if φ preserves inner products. For example, the 'shift' operator invoked in (1.60) is patently norm-preserving, and hence preserves inner products as well.

By the previous paragraph, it is clear that we could have defined an inner product isomorphism as a norm-preserving vector space isomorphism. This definition would then just be a special case of the definition of an isomorphism between two arbitrary normed spaces. Of course, not every normed space need be an inner product space (i.e., need have a norm definable in terms of an inner product on the space). But if the norm on a vector space V happens to satisfy:

$$\textbf{the parallelogram law}: \; ||v+v'||^2 + ||v-v'||^2 = 2||v||^2 + 2||v'||^2, \qquad (2.106)$$

then defining an inner product on V via the formula (2.105) makes V an inner product space. The proof of this is left to the reader, as is the verification that the parallelogram law holds of the norm in any inner product space. Thus, the parallelogram law is both necessary and sufficient for a normed space to define an inner product space.

Before proceeding, we need to make a few remarks about notation. When we wish to take the inner product of a vector $|w\rangle$ with a vector $\boldsymbol{F}|v\rangle$ (obtained by applying the operator $\boldsymbol{F}$ to $|v\rangle$), we shall write the latter as $|\boldsymbol{F}v\rangle$ and the inner product as $\langle w|\boldsymbol{F}v\rangle$. Similarly, the inner product of $\boldsymbol{F}|v\rangle$ with $|w\rangle$ will be written as $\langle\boldsymbol{F}v|w\rangle$. As a simple application of this notation, recall from Section 2.3 that the expansion coefficients of a vector $|v\rangle$ in terms of an orthonormal set $\{e_i\}$ are given by $\{\langle e_i|v\rangle\}$. Therefore the ijth entry of the matrix in ℓ^n that represents a linear operator $\boldsymbol{F}$ relative to an orthonormal basis is $F_{ij} = \langle e_i|\boldsymbol{F}e_j\rangle$ (cf. the discussion prior to (1.57)).

One is often interested in the special case where the inner product isomorphism at issue maps the space back onto itself, in which case it is called a **unitary operator**. For example, in R^3, the unitary operators are rotations about fixed axes, reflections through planes, and compositions thereof. (Is (1.60)'s shift operator unitary?) Because a unitary operator preserves inner products, it will preserve orthogonality between vectors and norms. Therefore unitary operators must map orthonormal bases into orthonormal bases. Conversely, any two orthonormal bases E and E' of an inner product space V (of countable dimension) are related by some unitary operator. For since E and E' have the same cardinality, there is a one-to-one, onto mapping $\boldsymbol{U} : E \to E'$ that can be extended by linearity to an operator on V. With this definition of $\boldsymbol{U}$, it is easy to see that $\langle \boldsymbol{U} e_i | \boldsymbol{U} e_j \rangle = \langle e_i | e_j \rangle$ for any two $|e_i\rangle, |e_j\rangle \in E$. Thus for *any* $|v\rangle, |w\rangle \in V$,

$$\langle \boldsymbol{U} v | \boldsymbol{U} w \rangle = \left\langle \boldsymbol{U} \sum_{j=1}^{m} k_j e_j | \boldsymbol{U} \sum_{p=1}^{n} l_p e_p \right\rangle, \tag{2.107}$$

$$= \sum_{j,p=1}^{m,n} k_j^* l_p \langle \boldsymbol{U} e_j | \boldsymbol{U} e_p \rangle, \tag{2.108}$$

$$= \sum_{j,p=1}^{m,n} k_j^* l_p \langle e_j | e_p \rangle, \tag{2.109}$$

$$= \left\langle \sum_{j=1}^{m} k_j e_j | \sum_{p=1}^{n} l_p e_p \right\rangle = \langle v | w \rangle, \tag{2.110}$$

showing that $\boldsymbol{U}$ is indeed unitary, as claimed. (Henceforth, we shall always reserve the letter $\boldsymbol{U}$ for unitary operators.)

2.8 Bras and Kets

Working in a vector space V with an inner product on it yields a stock of linear functionals on V that can be invoked. Every $|v\rangle \in V$ automatically determines a linear functional $\boldsymbol{f}^v$ with action $\boldsymbol{f}^v |u\rangle = \langle v|u\rangle$ for all $|u\rangle \in V$. It is convenient to denote the dual vector $\boldsymbol{f}^v$ by '$\langle v|$'. Any dual vector of this form, i.e., given by $\langle v|$ for some $|v\rangle \in V$, is called a **bra vector**. The reason is simple: the action of a bra vector on any **ket vector** $|u\rangle \in V$ is given by $\langle v|(|u\rangle) = \langle v|u\rangle$, so the result of that action produces the closed bra(-c-)ket expression '$\langle v|u\rangle$'. This symbolism and terminology is not merely cute, but good book-keeping. For we now know that whenever we see a closed bracket expression, it denotes a number, whereas terms involving an unclosed ket, such as $\langle v|w\rangle|u\rangle$, are vectors, and ones involving an unclosed bra, like $k^*\langle v|$, are dual vectors. Bra-ket notation also helps us to spot (and thus prevent writing down) undefined expressions such as $\langle v|v'\rangle\langle w| + \langle v'|v\rangle|w\rangle$, which is a futile attempt to add a vector and a dual vector (which live in different spaces).

Of course, linear combinations of bra vectors are again bra vectors. One can immediately see from the antilinearity of the inner product that the following identity holds (for all $k, k' \in K$ and $|v\rangle, |v'\rangle \in V$):

$$\langle kv + k'v'| = k^*\langle v| + k'^*\langle v'|. \tag{2.111}$$

Indeed, linearity and antilinearity of inner products can now be restated in terms of the action of bras on kets as

$$\langle v|kw + kw'\rangle = \langle v|(k|w\rangle + k'|w'\rangle),\ \langle kv + k'v'|w\rangle = (k^*\langle v| + k'^*\langle v'|)|w\rangle. \tag{2.112}$$

In writing out inner products of linear combinations of vectors, it will often be convenient to employ bras and kets in this way. We can also employ them to form inner products on tensor product spaces in a new way. Where we formerly wrote $\langle v_1 \otimes v_2|v_1' \otimes v_2'\rangle$ $(\overset{\text{def}}{=} \langle v_1|v_1'\rangle\langle v_2|v_2'\rangle)$, we can now write $((\langle v_1| \otimes \langle v_2|)(|v_1'\rangle \otimes |v_2'\rangle)$ where $\langle v_1| \otimes \langle v_2|$ lives in the space $V_1^* \otimes V_2^*$ and has the same action on elements of $V_1 \otimes V_2$ as $\langle v_1 \otimes v_2| \in (V_1 \otimes V_2)^*$.

The obvious question now arises: Is *every* vector in the dual of an inner product space a *bra* vector? 'Yes' in the finite-dimensional case, but 'No' in general. In the finite case, recall from Section 1.10 that V^n is self-dual, and a basis for $(V^n)^*$ is given by the linear functionals $\{\boldsymbol{f}_i\}_{i=1}^n$ with action $\boldsymbol{f}_i|v_j\rangle = \delta_{ij}$ on some arbitrary basis $\{v_i\}_{i=1}^n$ in V^n. If we take the latter to be an orthonormal basis $\{e_i\}_{i=1}^n$, then evidently $\boldsymbol{f}_i$ has the same action as $\langle e_i|$ for all i, and the linear functionals $\{\boldsymbol{f}_i\}_{i=1}^n$ are all bra vectors. So bra vectors span $(V^n)^*$. And since linear combinations of bras are bras, *all* linear functionals on V^n must be bra vectors. In the infinite-dimensional case, this claim is false. Consider a space V with a countably infinite orthonormal basis $\{e_j\}$, and the element $\boldsymbol{f} \in V^*$ with action $\boldsymbol{f}|e_j\rangle = j$ for all j. Is there a $|v\rangle \in V$ such that $\boldsymbol{f} = \langle v|$? Supposing there is, we would have by the Schwarz inequality

$$j = \boldsymbol{f}|e_j\rangle = |\langle v|e_j\rangle| \leq ||v||\ ||e_j|| = ||v|| \text{ for all } j = 1 \text{ to } \infty, \tag{2.113}$$

which implies the absurdity that $|v\rangle$ has no norm. (Note that the trouble here stems from the action of $\boldsymbol{f}$ being 'unbounded'. For Hilbert spaces, which we discuss later, their duals are defined to be the set of all *bounded* linear functionals, and, as a consequence, Hilbert spaces are self-dual and every dual vector is a bra.)

It is clear that the composition of an operator $\boldsymbol{F}$ on V followed by a linear functional $\boldsymbol{f}$ on V yields the linear functional $\boldsymbol{fF}$ (on V) with action (for all $|v\rangle \in V$):

$$(\boldsymbol{fF})|v\rangle = \boldsymbol{f}|\boldsymbol{F}v\rangle \ (= \boldsymbol{f}(\boldsymbol{F}|v\rangle)). \tag{2.114}$$

If $\boldsymbol{f}$'s action is given by a bra vector $\langle w|$, such a linear functional can be written as $\langle w|\boldsymbol{F}$. So we now have *three* conceptually different ways to read the string of symbols '$\langle w|\boldsymbol{F}v\rangle$': as the inner product of the ket vectors $|w\rangle$ and $|\boldsymbol{F}v\rangle$ $(= \boldsymbol{F}|v\rangle)$, as the action of the bra $\langle w|$ on the ket $|\boldsymbol{F}v\rangle$, and as the action of the linear functional $\langle w|\boldsymbol{F}$ on the ket $|v\rangle$. Of course, '$\langle w|\boldsymbol{F}v\rangle$' denotes the very same number on all three readings. It is simply a matter of convenience which of these readings is adopted at any particular stage in an argument.

Just as for linear functionals, operators on an inner product space can be defined using bras and kets. Any pair of vectors $|u\rangle, |u'\rangle \in V$ automatically determine an operator on V with action $(|u\rangle\langle u'|)|y\rangle = |u\rangle(\langle u|y\rangle)$ $(= (\langle v'|u\rangle)|v\rangle)$ for all $|y\rangle \in V$. Supposing that the $\boldsymbol{F}$ of the previous paragraph is of the ket-bra form $|u\rangle\langle u'|$ (note: not all operators need be of this form), $\langle w|\boldsymbol{F}v\rangle$ can be rewritten as '$\langle w|u\rangle\langle u'|v\rangle$', and this string of symbols now gains a *fourth* interpretation: as simply the product of two numbers.

Notes and References

Again, nice treatments of inner product spaces can be found in the final chapters of Lipschutz (1968) and Halmos (1948). Sutherland (1975) contains an elementary introduction to metric spaces. Both Cohen (1989) and Beltrametti and Cassinelli (1981) contain detailed intermediate level treatments of orthomodular lattices and compatibility, discussed in the context of quantum theory. For more advanced discussions, see Varadarajan (1968) and Kalmbach (1983). Chapter 4 of Bell and Machover (1977) gives a succint review of Boolean algebras. The concept of a partial Boolean algebra was first introduced by Kochen and Specker (1965,1967) in order to found a rival version of quantum logic to that championed by Birkhoff and Von Neumann (1936), which was based on orthomodular lattices. Our definition of a partial Boolean algebra follows Bell (1996). The elegant bra-ket formalism was first introduced by Dirac (1939), and made popular (mainly among physicists) by Dirac's famous book on quantum theory (1958).

3

OPERATORS ON FINITE-DIMENSIONAL COMPLEX INNER PRODUCT SPACES

In the previous two chapters we put no restriction on the dimension of the spaces under discussion, nor did we require that they be defined over the complex numbers as opposed to the reals. In this chapter, we shall focus entirely—apart from certain general definitions— on inner product spaces V^n that are both finite-dimensional and complex. These restrictions make it possible for $\mathcal{A}(V^n)$ to possess 'norm' and 'adjoint' structures that need not be present in the algebra of (all) linear operators on an arbitrary inner product space.

On the other hand, to formulate quantum theory we only require a special kind of complex inner product space, called a 'Hilbert space', that possesses extra 'topological' structure. In the next few chapters, we shall see what that structure amounts to and how it permits a certain subalgebra of the linear operators on a Hilbert space, the 'continuous' linear operators, to possess norm and adjoint structures that reduce to the corresponding structures possessed by $\mathcal{A}(V^n)$ when the Hilbert space is finite-dimensional. Holding topological considerations at bay has the advantage of allowing one to gain enough facility with the theory's mathematical apparatus to begin delving into literature, a substantial portion of which presupposes familiarity only with finite-dimensional Hilbert spaces.

3.1 Operator Norms and Normed Algebras

Let V be an arbitrary (not necessarily finite-dimensional) inner product space. When it exists, the **norm of an operator $\boldsymbol{F}$** on V, denoted $|\boldsymbol{F}|$, is defined to be the smallest real number r such that

$$||\boldsymbol{F}e|| \leq r \text{ for all unit vectors } |e\rangle \in V. \qquad (3.1)$$

Equivalently, we may write $|\boldsymbol{F}| = \bigvee_{||e||=1}\{||\boldsymbol{F}e||\}$, where the join here refers to the lattice of real numbers and is taken over all real numbers $||\boldsymbol{F}e||$ such that $||e|| = 1$.

Intuitively, $|\boldsymbol{F}|$ represents the maximum expansion factor that can be achieved by applying $\boldsymbol{F}$ to any unit vector. But note that, even if such a maximum exists, it need not be attained, i.e., it need not be the case that $||\boldsymbol{F}e|| = |\boldsymbol{F}|$ for some unit $|e\rangle$. Moreover, as hinted above, $\boldsymbol{F}$ need not even possess a norm when V is infinite-dimensional. For let V have countably infinite dimension, and consider an orthonormal basis $\{e_i\} \subseteq V$. Then for the operator defined by $\boldsymbol{F}|e_i\rangle = i|e_i\rangle$, i$|\boldsymbol{F}|$ does not exist, otherwise it would have to exceed every natural number.

On the other hand, every operator on a finite-dimensional space V^n possesses a norm. Let $\{e_i\}_{i=1}^n$ be an orthonormal basis for V^n, and set $l = \bigvee_{||e||=1}\{||\boldsymbol{F}e_i||\}_{i=1}^n$, a fixed number. Expand an arbitrary unit vector $|e\rangle$ in terms of the given orthonormal basis as

$$|e\rangle = \sum_{i=1}^{n} c_i|e_i\rangle, \text{ where } \sum_{i=1}^{n} |c_i|^2 = 1. \tag{3.2}$$

The second equation in (3.2) (which follows from our assumption that $||e|| = 1$) requires that for all i, $|c_i| \leq 1$. Using this inequality, the definition of l, and the triangle inequality, we get

$$||\boldsymbol{F}e|| = ||\sum_{i=1}^{n} c_i\boldsymbol{F}e_i|| \leq \sum_{i=1}^{n} |c_i|\ ||\boldsymbol{F}e_i|| \leq nl, \tag{3.3}$$

so that there is indeed a fixed upper limit on how much any given operator can expand the length of a unit vector.

With every operator on V^n assigned its norm, you are invited to check that $\mathcal{A}(V^n)$ is a normed space (and therefore also a metric space with the metric induced by its norm; see section 2.1). A **normed algebra** $\mathcal{A}$ over K is a normed space over K where the algebraic product and norm together satisfy (for all $X, Y \in \mathcal{A}$):

$$\textbf{product inequality}\text{: } |XY| \leq |X||Y|. \tag{3.4}$$

To see that $\mathcal{A}(V^n)$ also qualifies as a normed *algebra*, note that the product inequality is trivial if one of the operators at issue is zero. Otherwise, observe that for unit vectors such that $\boldsymbol{G}|e\rangle \neq |0\rangle$,

$$||\boldsymbol{F}\boldsymbol{G}e|| = \left\| \ ||\boldsymbol{G}e||\ \boldsymbol{F}\frac{\boldsymbol{G}e}{||\boldsymbol{G}e||} \right\| \leq ||\boldsymbol{G}e||\ |\boldsymbol{F}| \leq |\boldsymbol{G}||\boldsymbol{F}|, \tag{3.5}$$

whence

$$|\boldsymbol{FG}| = \bigvee_{\|e\|=1} \{\|\boldsymbol{FG}e\|\} \leq \bigvee_{\|e\|=1} \{\|\boldsymbol{FG}e\| : \boldsymbol{G}|e\rangle \neq |0\rangle\} \leq |\boldsymbol{G}||\boldsymbol{F}|. \tag{3.6}$$

3.2 Operator Adjoints and *-Algebras

Let $\boldsymbol{F}$ be any operator on V^n. Then there exists a unique linear operator $\boldsymbol{F}^*$, called the **adjoint** of $\boldsymbol{F}$, such that

$$\langle \boldsymbol{F}v|w\rangle = \langle v|\boldsymbol{F}^*w\rangle \text{ for all } |v\rangle, |w\rangle \in V^n \tag{3.7}$$

or, equivalently, such that the linear functionals $\langle \boldsymbol{F}v|$ and $\langle v|\boldsymbol{F}^*$ are the same (i.e., have the same action on all $|v\rangle \in V^n$). Of course, these claims of existence and uniqueness require argument.

We shall dispense with existence first. For any vector $|w\rangle$ consider the linear functional $\langle w|\boldsymbol{F}$. Then since every linear functional on a finite-dimensional inner product space is a bra vector (section 2.8), there is a vector $|w'\rangle \in V^n$ such that $\langle w|\boldsymbol{F} = \langle w'|$. Furthermore, $|w'\rangle$ is obviously the *only* vector in V^n for which $\langle w|\boldsymbol{F} = \langle w'|$, since no two vectors can have the same inner product with all vectors unless they are the same. Next, define a mapping $\boldsymbol{F}^* : V^n \to V^n$ which, for any $|w\rangle$, maps $|w\rangle$ to the unique vector $|w'\rangle$ such that $\langle w|\boldsymbol{F} = \langle w'|$. Then for all $|v\rangle$ and $|w\rangle$,

$$\langle w|\boldsymbol{F}v\rangle = \langle w'|v\rangle \;\Rightarrow\; \langle w|\boldsymbol{F}v\rangle = \langle \boldsymbol{F}^*w|v\rangle \;\Rightarrow\; \langle \boldsymbol{F}v|w\rangle = \langle v|\boldsymbol{F}^*w\rangle. \tag{3.8}$$

(The first entailment follows from the definition of $\boldsymbol{F}^*$, and the second from the conjugate-symmetry of the inner product.) Moreover, $\boldsymbol{F}^*$ is linear, because for any vector $|v\rangle$ we have

$$\begin{aligned}
\langle v|\boldsymbol{F}^*(k_1w_1 + k_2w_2)\rangle &= \langle \boldsymbol{F}v|k_1w_1 + k_2w_2\rangle && (3.9)\\
&= k_1\langle \boldsymbol{F}v|w_1\rangle + k_2\langle \boldsymbol{F}v|w_2\rangle && (3.10)\\
&= k_1\langle v|\boldsymbol{F}^*w_1\rangle + k_2\langle v|\boldsymbol{F}^*w_2\rangle && (3.11)\\
&= \langle v|k_1\boldsymbol{F}^*w_1 + k_2\boldsymbol{F}^*w_2\rangle. && (3.12)
\end{aligned}$$

So the linear operator $\boldsymbol{F}^*$ that we have defined qualifies as an adjoint of $\boldsymbol{F}$.

To see that $\boldsymbol{F}$ can have at most one adjoint, we first record a general fact about operator identities that we shall frequently invoke without comment. If for *all* $|v\rangle$ and $|w\rangle$ the identity $\langle v|\boldsymbol{F}_1w\rangle = \langle v|\boldsymbol{F}_2w\rangle$ holds (or, equivalently, $\langle \boldsymbol{F}_1v|w\rangle = \langle \boldsymbol{F}_2v|w\rangle$ holds, by conjugate-symmetry of the inner product), then the *operator* identity $\boldsymbol{F}_1 = \boldsymbol{F}_2$ must also hold (because no two vectors can have the same inner products with all other vectors unless they are one and the same vector). Therefore, $\boldsymbol{F}$ can have at most one adjoint; for if it had two, $\boldsymbol{F}_1^*$ and $\boldsymbol{F}_2^*$, then by (3.7) they would have to satisfy:

$$\langle \boldsymbol{F}v|w\rangle = \langle v|\boldsymbol{F}_1^* w\rangle = \langle v|\boldsymbol{F}_2^* w\rangle \text{ for all } |v\rangle \text{ and } |w\rangle, \tag{3.13}$$

whence $\boldsymbol{F}_1^* = \boldsymbol{F}_2^*$.

The matrix representation in C^n of the adjoint of an operator is easily determined. Recall once more that the matrix representing an operator $\boldsymbol{F}$ relative to an orthonormal basis has entries $F_{ij} = \langle e_i|\boldsymbol{F}e_j\rangle$. It follows that the matrix representing $\boldsymbol{F}^*$ has entries F_{ji}^*, because

$$\langle e_i|\boldsymbol{F}^* e_j\rangle = \langle \boldsymbol{F}e_i|e_j\rangle = \langle e_j|\boldsymbol{F}e_i\rangle^*. \tag{3.14}$$

Thus, to obtain the matrix for $\boldsymbol{F}^*$, one simply 'takes the transpose' of the matrix for $\boldsymbol{F}$, i.e., reflects the matrix about its main diagonal, and then conjugates all its entries. We could equally well have proved that every operator on V^n has an adjoint by noting that in the space of complex column matrices C^n the conjugate-transpose of an $n \times n$ matrix operator qualifies as its adjoint (via explicitly calculating the matrix products involved in the definition of an adjoint), and then invoking the fact that the inner product spaces C^n and V^n, and the algebras of operators over them, are isomorphic.

Not only does every operator on V^n have a unique adjoint, but also $\mathcal{A}(V^n)$ has the structure of an 'involutive' algebra, or '*-algebra'. A ***-algebra** $\boldsymbol{\mathcal{A}}$ over K consists of an algebra $\mathcal{A}$ over K with an additional mapping $*$ that assigns to any element $X \in \mathcal{A}$ another $X^* \in \mathcal{A}$, called (of course) the adjoint of X, where the operation $*$ is anti-linear (i.e., preserves linear combinations except for a conjugation of their coefficients) and satisfies (for all $X, Y \in \mathcal{A}$):

$$X^{**} = X, \tag{3.15}$$

$$(XY)^* = Y^* X^*. \tag{3.16}$$

Note the analogy here between taking the adjoint of an operator and the conjugate of a complex number. However, because the complex numbers form a commutative *-algebra (over themselves), the reversal of order in (3.16) is unimportant, whereas in the *non*commutative algebra $\mathcal{A}(V^n)$ it cannot be ignored. To establish that both (3.15) and (3.16) hold in $\mathcal{A}(V^n)$, it suffices to note that for all $|v\rangle$ and $|w\rangle$:

$$\langle \boldsymbol{F}^{**}v|w\rangle = \langle v|\boldsymbol{F}^* w\rangle = \langle \boldsymbol{F}v|w\rangle, \tag{3.17}$$

$$\langle v|(\boldsymbol{F}\boldsymbol{G})^* w\rangle = \langle \boldsymbol{F}\boldsymbol{G}v|w\rangle = \langle \boldsymbol{G}v|\boldsymbol{F}^* w\rangle = \langle v|\boldsymbol{G}^*\boldsymbol{F}^* w\rangle. \tag{3.18}$$

The adjoint operation also behaves reasonably towards operator norms; in fact, it preserves them. Consider that

$$\begin{aligned} |\boldsymbol{F}|^2 &= \bigvee_{\|e\|=1} \{\|\boldsymbol{F}e\|^2\}, && (3.19)\\ &= \bigvee_{\|e\|=1} \{\langle e|\boldsymbol{F}^*\boldsymbol{F}e\rangle\}, && (3.20)\\ &\leq \bigvee_{\|e\|=1} \{\|\boldsymbol{F}^*\boldsymbol{F}e\|\}, && (3.21)\\ &= |\boldsymbol{F}^*\boldsymbol{F}|, && (3.22)\\ &\leq |\boldsymbol{F}^*||\boldsymbol{F}|, && (3.23) \end{aligned}$$

exploiting the Schwartz inequality in the third step, and the product inequality in the last. It follows that $|\boldsymbol{F}| \leq |\boldsymbol{F}^*|$ and—interchanging the roles of $\boldsymbol{F}$ and $\boldsymbol{F}^*$ throughout the argument—that $|\boldsymbol{F}| = |\boldsymbol{F}^*|$. With this conclusion in hand, the above inequalities further entail that $|\boldsymbol{F}^*\boldsymbol{F}| = |\boldsymbol{F}|^2$. In fact the latter identity is the stronger of the two: it is an easy exercise to show (using the product inequality) that in any normed *-algebra where the identity $|X^*X| = |X|^2$ holds, the adjoint operation will preserve norms.

Finally, adjoints preserve tensor products. For any pair of product vectors in $V_1 \otimes V_2$, we have

$$\begin{aligned} \langle v_1 \otimes v_2|(\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)^* v_1' \otimes v_2'\rangle &= \langle (\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)v_1 \otimes v_2|v_1' \otimes v_2'\rangle, && (3.24)\\ &= (\langle \boldsymbol{F}_1 v_1| \otimes \langle \boldsymbol{F}_2 v_2|)\,(|v_1'\rangle \otimes |v_2'\rangle), && (3.25)\\ &= \langle \boldsymbol{F}_1 v_1|v_1'\rangle\langle \boldsymbol{F}_2 v_2|v_2'\rangle, && (3.26)\\ &= \langle v_1|\boldsymbol{F}_1^* v_1'\rangle\langle v_2|\boldsymbol{F}_2^* v_2'\rangle, && (3.27)\\ &= (\langle v_1| \otimes \langle v_2|)\,(|\boldsymbol{F}_1^* v_1'\rangle \otimes |\boldsymbol{F}_2^* v_2'\rangle), && (3.28)\\ &= \langle v_1 \otimes v_2|\,(\boldsymbol{F}_1^* \otimes \boldsymbol{F}_2^*)\,(v_1' \otimes v_2')\rangle, && (3.29) \end{aligned}$$

and this calculation suffices to establish $(\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)^* = \boldsymbol{F}_1^* \otimes \boldsymbol{F}_2^*$ as an *operator* identity. For the operators at issue are linear, product vectors span $V_1 \otimes V_2$, and the inner product is linear and antilinear, all of which license the move from equation (3.29) to the assertion that $\langle v|\,(\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)^*\, w\rangle = \langle v|\boldsymbol{F}_1^* \otimes \boldsymbol{F}_2^* w\rangle$ for all $|v\rangle, |w\rangle \in V_1 \otimes V_2$.

3.3 Unitary Operators and Groups

Recall that a unitary operator $\boldsymbol{U}$ on V^n is an inner product isomorphism from V^n to itself (section 2.7). Consider, then, what the adjoint of $\boldsymbol{U}$ must be. Since $\boldsymbol{U}$ preserves inner products,

$$\langle v|w\rangle = \langle \boldsymbol{U}v|\boldsymbol{U}w\rangle = \langle v|\boldsymbol{U}^*\boldsymbol{U}w\rangle \text{ for all } |v\rangle, |w\rangle, \tag{3.30}$$

whence $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}$. But since unitary operators are one-to-one and onto, they are invertible, and we may multiply both sides of $\boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I}$ from the right by $\boldsymbol{U}^{-1}$ yielding $\boldsymbol{U}^* = \boldsymbol{U}^{-1}$. Thus we see that the adjoints of unitary operators are their inverses. Conversely, it is clear that every operator whose adjoint is its inverse preserves inner products, and is therefore unitary. So, in fact, the assertion that $\boldsymbol{U}^* = \boldsymbol{U}^{-1}$ is *equivalent* to the statement that $\boldsymbol{U}$ is unitary.

This characterization of unitarity makes it a matter of simple algebra to show that the set of all unitary operators in $\mathcal{A}(V^n)$ forms a 'group'. A **group** is a set $\mathcal{G}$ in which any two elements $g, g' \in \mathcal{G}$ have a product $gg' \in \mathcal{G}$, where this product operation has three properties. First, it must be associative. Second, there must be a (necessarily unique) identity element $e \in \mathcal{G}$ satisfying

$$eg = ge = e \text{ for all } g \in \mathcal{G}. \tag{3.31}$$

And, third, every element $g \in \mathcal{G}$ must possess a (necessarily unique) inverse, $g^{-1} \in \mathcal{G}$, satisfying

$$gg^{-1} = g^{-1}g = e. \tag{3.32}$$

For unitary operators, the group product is the algebraic product in $\mathcal{A}(V^n)$ (i.e., composition of linear mappings from V^n to itself). To see that the product of two unitary operators is another, observe that

$$(\boldsymbol{U}_1\boldsymbol{U}_2)^*(\boldsymbol{U}_1\boldsymbol{U}_2) = \boldsymbol{U}_2^*\boldsymbol{U}_1^*\boldsymbol{U}_1\boldsymbol{U}_2 = \boldsymbol{U}_2^*\boldsymbol{U}_1^{-1}\boldsymbol{U}_1\boldsymbol{U}_2 = \boldsymbol{U}_2^{-1}\boldsymbol{U}_2 = \boldsymbol{I}, \tag{3.33}$$

and that the inverse of $\boldsymbol{U}_1\boldsymbol{U}_2$, i.e., $(\boldsymbol{U}_1\boldsymbol{U}_2)^{-1}$, is just $\boldsymbol{U}_2^{-1}\boldsymbol{U}_1^{-1}$. We can then multiply (3.33) on the right by $(\boldsymbol{U}_1\boldsymbol{U}_2)^{-1}$ and obtain $(\boldsymbol{U}_1\boldsymbol{U}_2)^* = (\boldsymbol{U}_1\boldsymbol{U}_2)^{-1}$, which is none other than the assertion that $\boldsymbol{U}_1\boldsymbol{U}_2$ is itself unitary. (Exercise: show that the *tensor* product of unitary operators is also unitary.) Clearly the identity operator $\boldsymbol{I}$ must be the identity in the group of unitarity operators, and $\boldsymbol{I}$ is indeed unitary because $\boldsymbol{I}^* = \boldsymbol{I} = \boldsymbol{I}^{-1}$. Finally, the inverse of a unitary operator $\boldsymbol{U}$ is again unitary because

$$(\boldsymbol{U}^{-1})^* = \boldsymbol{U}^{**} = \boldsymbol{U} = (\boldsymbol{U}^{-1})^{-1}. \tag{3.34}$$

3.4 Self-Adjoint Operators and Jordan-Lie Algebras

An operator $\boldsymbol{F}$ is called **self-adjoint** (sometimes also called 'Hermitian') if $\boldsymbol{F} = \boldsymbol{F}^*$. Because the adjoint of a matrix operator on C^n is obtained by reflecting the matrix about its main diagonal and conjugating all entries (see section 3.2), self-adjoint matrices must have real numbers along their main diagonal. And the eigenvalues of a self-adjoint operator are always real. For let $|v\rangle$ be an eigenvector of (self-adjoint) $\boldsymbol{F}$ corresponding to eigenvalue k. Then we have

$$\langle \boldsymbol{F}v|v\rangle = \langle v|\boldsymbol{F}v\rangle \;\Rightarrow\; \langle kv|v\rangle = \langle v|kv\rangle \;\Rightarrow\; k^*\langle v|v\rangle = k\langle v|v\rangle \;\Rightarrow\; k^* = k. \quad (3.35)$$

Evidently, any real linear combination of self-adjoint operators is self-adjoint, but the product of two self-adjoint operators is self-adjoint if and only if they commute. So the self-adjoint operators in $\mathcal{A}(V^n)$ form neither a real associative algebra nor a group. However, they do form a 'Jordan-Lie' algebra, as do the self-adjoint elements of *any* complex *-algebra. (Here, the reader may wish to recall the definitions of a Jordan and a Lie algebra from section 1.8.) A **Jordan-Lie** algebra $\mathcal{S}$ possesses *two* products $\circ$ and $\bullet$—the first making $\mathcal{S}$ a real Jordan algebra and the second a real Lie algebra—which together satisfy, for some fixed real number $r \geq 0$ (and all $X, Y, Z \in \mathcal{S}$):

$$\textbf{Leibniz rule}: X \bullet (Y \circ Z) = (X \bullet Y) \circ Z + Y \circ (X \bullet Z), \quad (3.36)$$

$$\textbf{associator identity}: (X \circ Y) \circ Z - X \circ (Y \circ Z) = r(X \bullet Z) \bullet Y. \quad (3.37)$$

What should the products $\circ$ and $\bullet$ be for the set of all self-adjoint elements $\mathcal{S}(\mathcal{A})$ of a complex *-algebra $\mathcal{A}$? Observe that every element $X \in \mathcal{A}$ has unique real and imaginary parts in $\mathcal{S}(\mathcal{A})$ given by

$$\Re(X) = 1/2(X^* + X), \quad \Im(X) = i/2(X^* - X). \quad (3.38)$$

Evidently, $X = \Re(X) + i\Im(X)$ and both $\Re(X)$ and $\Im(X)$ are self-adjoint. (Moreover, the reader may easily verify that there are no *other* self-adjoint operators A, B in terms of which X may be expressed as $X = A + iB$.) It is natural, then, to define the Jordan and Lie products of two elements in $\mathcal{S}(\mathcal{A})$ by

$$X \circ Y \overset{\text{def}}{=} \Re(XY) = 1/2[X, Y]_+, \quad X \bullet Y \overset{\text{def}}{=} -\Im(XY) = -i/2[X, Y] \quad (3.39)$$

(which are similar to the definitions we adopted in section 1.8). To obtain the Leibniz rule, note that if we can establish the following special case

$$X \bullet (Y \circ Y) = 2(X \bullet Y) \circ Y, \tag{3.40}$$

then the full Leibniz rule can be derived from (3.40) by substituting $Y + Z$ for Y and using bilinearity. Equation (3.40) is verified as follows:

$$\begin{aligned}
2(X \bullet Y) \circ Y &= -i[X, Y] \circ Y, & (3.41)\\
&= -i\left((XY) \circ Y - (YX) \circ Y\right), & (3.42)\\
&= -i/2\left(XY^2 + YXY - (YXY + Y^2X)\right), & (3.43)\\
&= -i/2[X, Y^2] = X \bullet (Y \circ Y). & (3.44)
\end{aligned}$$

Finally, with the definitions in (3.39), the associator identity ends up reducing to

$$(1 - r)(ZXY + YXZ - XZY - YZX) = 0 \tag{3.45}$$

which, when $\mathcal{A}$ is noncommutative, is satisfied if and only if $r = 1$. In the commutative case, (3.45) is of course satisfied regardless of the value of r. But if we *require* the associator identity to hold with $r = 0$, it is not difficult to see that $\mathcal{A}$ must be commutative, so in fact the value $r = 0$ characterizes the commutative case.

We end this section by observing that the (associative) algebra $\mathcal{A}(V^n)$ has bases that consist entirely of self-adjoint operators. Start with any orthonormal basis $\{e_i\}_{i=1}^n$ for V^n, and consider all n^2 operators of the form $|e_i\rangle\langle e_j|$. They are linearly independent, because for any complex coefficients c_{ij} and indices k, k':

$$\sum_{i,j=1}^{n} c_{ij}|e_i\rangle\langle e_j| = \mathbf{0} \Rightarrow \sum_{i,j=1}^{n} c_{ij}\langle e_k|e_i\rangle\langle e_j|e_{k'}\rangle = 0 \Rightarrow c_{kk'} = 0. \tag{3.46}$$

And since the dimension of $\mathcal{A}(V^n)$ is itself n^2, the operators of form $|e_i\rangle\langle e_j|$ must also span $\mathcal{A}(V^n)$. But then so must their *self-adjoint* real and imaginary parts; i.e., the operators of form

$$\boldsymbol{F}_{ij} = \Re(|e_i\rangle\langle e_j|), \quad \boldsymbol{G}_{ij} = \Im(|e_i\rangle\langle e_j|), \tag{3.47}$$

span $\mathcal{A}(V^n)$ as well. Because $\boldsymbol{F}_{ij} = \boldsymbol{F}_{ji}$ and $\boldsymbol{G}_{ij} = -\boldsymbol{G}_{ji}$, these $\boldsymbol{F}$'s and $\boldsymbol{G}$'s are not all linearly independent. So we can drop some of them without changing the fact that they span $\mathcal{A}(V^n)$. In particular, the following subset of the $\boldsymbol{F}$'s and $\boldsymbol{G}$'s obviously continues to span $\mathcal{A}(V^n)$:

$$\{\boldsymbol{F}_{ij} : i \leq j\} \cup \{\boldsymbol{G}_{ij} : i > j\}. \tag{3.48}$$

But now there are only n^2 $\boldsymbol{F}$'s and $\boldsymbol{G}$'s left, so they must be linearly independent as well, and hence form a basis in $\mathcal{A}(V^n)$.

3.5 Projection Operators and Subspaces

An operator $\boldsymbol{P}$ is called a **projection operator**, or just a **projection**, if it is both self-adjoint and **idempotent**, i.e., if $\boldsymbol{P}$ satisfies $\boldsymbol{P}^2 = \boldsymbol{P}$. For example, the following manifestly self-adjoint matrices (in $\mathcal{A}(C^2)$)

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}, \text{ and } \begin{pmatrix} 2/3 & (-1+i)/3 \\ (-1-i)/3 & 1/3 \end{pmatrix}, \tag{3.49}$$

are all projections because they equal their own squares. It is easy to see that an operator $\boldsymbol{P}$ is both self-adjoint and idempotent exactly when $\boldsymbol{P}\boldsymbol{P}^* = \boldsymbol{P}$. Thus, the tensor product of two projections is a again a projection because

$$(\boldsymbol{P}_1 \otimes \boldsymbol{P}_2)(\boldsymbol{P}_1 \otimes \boldsymbol{P}_2)^* = (\boldsymbol{P}_1 \otimes \boldsymbol{P}_2)(\boldsymbol{P}_1^* \otimes \boldsymbol{P}_2^*) \tag{3.50}$$

$$= (\boldsymbol{P}_1\boldsymbol{P}_1^*) \otimes (\boldsymbol{P}_2\boldsymbol{P}_2^*) = \boldsymbol{P}_1 \otimes \boldsymbol{P}_2. \tag{3.51}$$

The reason for their name is that projections have an important geometric interpretation. Recall (from section 2.5) that for any subspace $U \subseteq V$, $V = U \vee U^\perp$. We have also seen that when V is finite-dimensional, all its subspaces will be orthoclosed, so that the operations $\vee$ and $+$ on subspaces will coincide. It follows that, for every subspace $U \subseteq V^n$, $V^n = U + U^\perp$ and, therefore, that every vector $|v\rangle \in V^n$ can be written uniquely as the sum of a component vector in U and one in $U^\perp$. Let the mapping $\boldsymbol{P}_U$ send an arbitrary vector to its component in U. This mapping is clearly a linear operator—with eigenspaces U and $U^\perp$ and corresponding eigenvalues 1 and 0—and it is also idempotent. Moreover, $\boldsymbol{P}_U$ is self-adjoint, since for any two vectors $|v\rangle$ and $|w\rangle$ with respective components $|v_U\rangle, |v_{U^\perp}\rangle$ and $|w_U\rangle, |w_{U^\perp}\rangle$ in U and $U^\perp$, we have

$$\langle \boldsymbol{P}_U v|w\rangle = \langle v_U|\big(|w_U\rangle + |w_{U^\perp}\rangle\big), \tag{3.52}$$

$$= \langle v_U|w_U\rangle + 0, \tag{3.53}$$

$$= \langle v_U|w_U\rangle + \langle v_{U^\perp}|w_U\rangle, \tag{3.54}$$

$$= \big(\langle v_U| + \langle v_{U^\perp}|\big)|w_U\rangle = \langle v|\boldsymbol{P}_U w\rangle. \tag{3.55}$$

Therefore $\boldsymbol{P}_U$ qualifies as a projection operator.

Not only does every subspace U determine an operator $\boldsymbol{P}_U$ projecting vectors onto their components in that subspace, but also every projection operator determines a subspace onto which it projects. That is, if $\boldsymbol{P}$ is a projection, then there is always some subspace U such that $\boldsymbol{P} = \boldsymbol{P}_U$, viz., the subspace given by $U = \{\boldsymbol{P}|v\rangle : |v\rangle \in V\}$. To verify this claim, it suffices to show that $\boldsymbol{P}$ and $\boldsymbol{P}_U$

agree in their action on the components $|v_U\rangle$ and $|v_{U^\perp}\rangle$ of an arbitrary vector $|v\rangle$. Since $|v_U\rangle \in U$, there is a vector $|w\rangle$ such that $\boldsymbol{P}|w\rangle = |v_U\rangle$. Thus, using the idempotency of $\boldsymbol{P}$,

$$\boldsymbol{P}|v_U\rangle = \boldsymbol{P}^2|w\rangle = \boldsymbol{P}|w\rangle = |v_U\rangle = \boldsymbol{P}_U|v_U\rangle. \tag{3.56}$$

And since for any $|v\rangle$, $\boldsymbol{P}|v\rangle \in U$, it is clear that $\boldsymbol{P}_U\boldsymbol{P}|v\rangle = \boldsymbol{P}|v\rangle$ for all $|v\rangle$, whence $\boldsymbol{P}_U\boldsymbol{P} = \boldsymbol{P}$. Taking the adjoint of both sides of this latter equation, and using the self-adjointness of both $\boldsymbol{P}$ and $\boldsymbol{P}_U$, we obtain $\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}_U$. Therefore,

$$\boldsymbol{P}|v_{U^\perp}\rangle = \boldsymbol{P}\boldsymbol{P}_U|v_{U^\perp}\rangle = |0\rangle = \boldsymbol{P}_U|v_{U^\perp}\rangle. \tag{3.57}$$

Rather than saying 'the subspace onto which $\boldsymbol{P}$ projects', it is convenient to refer to that subspace simply as the **range** of $\boldsymbol{P}$. Due to the correspondence between projections and subspaces, we can speak of one-dimensional projections, two-dimensional projections, etc., referring to the dimension of their ranges. For *one*-dimensional projections, we can adopt the special notation $\boldsymbol{P}_{|v\rangle}$ for the projection onto the one-dimensional subspace generated by the vector $|v\rangle$. We can also write a one-dimensional projection using ket-bra notation. Evidently, $|v\rangle\langle w|$ maps all vectors to the one-dimensional subspace generated by $|v\rangle$. If we demand that $|v\rangle\langle w|$ be self-adjoint, then we are requiring $|v\rangle\langle w| = |w\rangle\langle v|$ which holds only if $|w\rangle = r|v\rangle$ for some nonzero real number r. Further demanding that $r|v\rangle\langle v|$ be idempotent requires that $r = 1$ and $|v\rangle$ be a unit vector. Therefore, the projection $\boldsymbol{P}_{|v\rangle}$ may be written in an equivalent manner as $|e\rangle\langle e|$, where $|e\rangle$ is any unit vector lying in the subspace generated by $|v\rangle$.

Projection operators have useful algebraic relations that reflect the relations between their corresponding subspaces. For any subspaces U and W,

$$\boldsymbol{P}_U\boldsymbol{P}_W = \boldsymbol{P}_W\boldsymbol{P}_U = \boldsymbol{0} \;\Leftrightarrow\; \boldsymbol{P}_U + \boldsymbol{P}_W = \boldsymbol{P}_{U+W} \;\Leftrightarrow\; U \perp W, \tag{3.58}$$

$$\boldsymbol{P}_U\boldsymbol{P}_W = \boldsymbol{P}_{U\cap W} \;\Leftrightarrow\; [\boldsymbol{P}_U, \boldsymbol{P}_W] = \boldsymbol{0} \;\Leftrightarrow\; U \leftrightarrow W. \tag{3.59}$$

Again borrowing the language of subspaces to describe projections, (3.58) licenses us to say that projections whose product is zero are orthogonal (meaning: their ranges are orthogonal), and (3.59) licenses us to say that commuting projections are compatible.

The equivalences in (3.58) are no more than a special case of those in (3.59) with U and W taken to intersect only in the zero subspace. We therefore focus our efforts on proving (3.59), for which it suffices to establish the chain of implications:

$$\boldsymbol{P}_U\boldsymbol{P}_W = \boldsymbol{P}_{U\cap W} \;\Rightarrow\; [\boldsymbol{P}_U, \boldsymbol{P}_W] = \boldsymbol{0} \;\Rightarrow\; U \leftrightarrow W \;\Rightarrow\; \boldsymbol{P}_U\boldsymbol{P}_W = \boldsymbol{P}_{U\cap W}. \quad (3.60)$$

So here goes. If $\boldsymbol{P}_U\boldsymbol{P}_W = \boldsymbol{P}_{U\cap W}$ then since $\boldsymbol{P}_{U\cap W}$ is self-adjoint, so is $\boldsymbol{P}_U\boldsymbol{P}_W$, which means $\boldsymbol{P}_U$ and $\boldsymbol{P}_W$ have to commute. Next, assume they commute. Simple algebra then reveals that all three of the operators $\boldsymbol{P}_X$, $\boldsymbol{P}_Y$, and $\boldsymbol{P}_Z$ given by

$$\boldsymbol{P}_X = \boldsymbol{P}_U - \boldsymbol{P}_U\boldsymbol{P}_W, \quad \boldsymbol{P}_Y = \boldsymbol{P}_W - \boldsymbol{P}_U\boldsymbol{P}_W, \quad \boldsymbol{P}_Z = \boldsymbol{P}_U\boldsymbol{P}_W, \qquad (3.61)$$

are projections, and that the product of any two of them is the zero projection. It follows that the subspaces X, Y, and Z are mutually orthogonal. For let $|x\rangle \in X$ and $|z\rangle \in Z$. Then $\boldsymbol{P}_X|x\rangle = |x\rangle$ and $\boldsymbol{P}_Z|z\rangle = |z\rangle$, so that

$$\langle x|z\rangle = \langle \boldsymbol{P}_X x|\boldsymbol{P}_Z z\rangle = \langle x|\boldsymbol{P}_X\boldsymbol{P}_Z z\rangle = 0, \qquad (3.62)$$

and similar statements obviously hold for X and Y, and for Y and Z. Moreover, $\boldsymbol{P}_U = \boldsymbol{P}_X + \boldsymbol{P}_Z$ and $\boldsymbol{P}_W = \boldsymbol{P}_Y + \boldsymbol{P}_Z$ (from (3.61)), so that $U = X + Z$ and $W = Y + Z$, because for any $|u\rangle \in U$,

$$|u\rangle = \boldsymbol{P}_U|u\rangle = \boldsymbol{P}_X|u\rangle + \boldsymbol{P}_Z|u\rangle \in X + Z, \qquad (3.63)$$

and for any $|x\rangle \in X$ and $|z\rangle \in Z$,

$$|x\rangle + |z\rangle = \boldsymbol{P}_X|x\rangle + \boldsymbol{P}_Z|z\rangle = \boldsymbol{P}_U(|x\rangle + |z\rangle) \in U \qquad (3.64)$$

(and, similarly, for the claim that $W = Y + Z$). But then $U \leftrightarrow W$. Finally, supposing that U and W are compatible, there are mutually orthogonal subspaces X, Y, and Z such that $U = X + Z$ and $W = Y + Z$. Since orthogonal subspaces are compatible, we can use distributivity to obtain $U \cap W = Z$, and all we need to show now is that $\boldsymbol{P}_Z = \boldsymbol{P}_U\boldsymbol{P}_W$. This may be shown simply by expanding an arbitrary vector $|v\rangle$ as $|v_X\rangle + |v_Y\rangle + |v_Z\rangle + |v'\rangle$, where $|v'\rangle$ is the component of $|v\rangle$ in the subspace orthogonal to all three of X, Y, and Z, and then observing that $\boldsymbol{P}_U\boldsymbol{P}_W$ maps $|v\rangle$ to $|v_Z\rangle$, just as $\boldsymbol{P}_Z$ does.

Since projections are another way of talking about subspaces, one sometimes refers to $\mathcal{L}_{\perp_o}(V^n)$ as the orthomodular lattice of *projections* on V^n. We leave the reader to verify that the partial ordering and lattice operations in $\mathcal{L}_{\perp_o}(V^n)$, thus understand, can be expressed as follows:

$$\boldsymbol{P}_U \le \boldsymbol{P}_W \;\Leftrightarrow\; \boldsymbol{P}_W\boldsymbol{P}_U = \boldsymbol{P}_U, \qquad (3.65)$$

$$\boldsymbol{P}_U \wedge \boldsymbol{P}_W \;=\; \boldsymbol{P}_{U\cap W}, \qquad (3.66)$$

$$\boldsymbol{P}_U \vee \boldsymbol{P}_W = \boldsymbol{P}_{U+W}, \tag{3.67}$$

$$\boldsymbol{P}_U^{\perp} = I - \boldsymbol{P}_U. \tag{3.68}$$

It follows from (3.58) and (3.67) that the join of orthogonal projections is simply their sum, and from (3.59) and (3.66) that the meet of compatible projections is their product. Observe, also, that the greatest element in $\mathcal{L}_{\perp}(V^n)$ is $\boldsymbol{P}_{V^n} = \boldsymbol{I}$, the least is $\boldsymbol{P}_0 = \boldsymbol{0}$, and that orthomodularity is now just the trivial assertion that

$$\boldsymbol{P}_W \boldsymbol{P}_U = \boldsymbol{P}_U \Rightarrow \boldsymbol{P}_W = \boldsymbol{P}_U + \boldsymbol{P}_W(\boldsymbol{I} - \boldsymbol{P}_U). \tag{3.69}$$

3.6 Normal Operators and The Spectral Theorem

In light of (3.58), any m-dimensional projection $\boldsymbol{P}_U$ on V^n $(m > 1)$ can be written as a sum of (necessarily, lower-dimensional) mutually orthogonal projections with ranges that together span U. Any set of mutually orthogonal (nonzero) projections that sum to the identity projection I (projecting onto the whole of V^n) is called a **resolution of the identity**. For example, if $\{e_i\}_{i=1}^n$ is any orthonormal basis, then

$$\boldsymbol{I} = \sum_{i=1}^{n} |e_i\rangle\langle e_i| \tag{3.70}$$

is a resolution of the identity into one-dimensional projections.

Given any resolution of the identity $\{\boldsymbol{P}_i\}_{i=1}^m$ $(m \leq n)$ on V^n, we can build a self-adjoint operator, $\boldsymbol{F}$, by choosing a set $\{r_i\}_{i=1}^m$ of distinct real numbers and defining

$$\boldsymbol{F} = \sum_{i=1}^{m} r_i \boldsymbol{P}_i. \tag{3.71}$$

$\boldsymbol{F}$ is manifestly self-adjoint (because its coefficients are real and projections are themselves self-adjoint), and the numbers $\{r_i\}_{i=1}^m$ are its eigenvalues. That *every* self-adjoint operator on V^n can be expanded *uniquely* in the above way—i.e., as a linear combination, with distinct real coefficients, of projections in some resolution of the identity—is a consequence of the **spectral theorem**.

In fact, the most general class of operators to which the spectral theorem applies are the 'normal' operators. An operator $\boldsymbol{N}$ on V^n is called **normal** if it commutes with its adjoint, i.e., $\boldsymbol{N}\boldsymbol{N}^* = \boldsymbol{N}^*\boldsymbol{N}$. (The reader might find it amusing to show that this condition is equivalent to $[\Re(\boldsymbol{N}), \Im(\boldsymbol{N})] = 0$.) Evidently self-adjoint operators (including projections) are normal, as are unitary operators.

The only fact we need about normal operators to prove the spectral theorem is the claim that: $|v\rangle$ is an eigenvector of $\boldsymbol{N}$ corresponding to eigenvalue c if and only if $|v\rangle$ is an eigenvector of $\boldsymbol{N}^*$ corresponding to eigenvalue c^*. The argument for this claim is straightforward. Observe that if $\boldsymbol{N}$ is normal, $\langle \boldsymbol{N}v|\boldsymbol{N}v\rangle = \langle \boldsymbol{N}^*v|\boldsymbol{N}^*v\rangle$, and therefore, taking square roots, $||\boldsymbol{N}v|| = ||\boldsymbol{N}^*v||$. Now given that $\boldsymbol{N}$ is normal, the operator $\boldsymbol{N} - c\boldsymbol{I}$, with adjoint $\boldsymbol{N}^* - c^*\boldsymbol{I}$, is also normal. The claim then follows immediately from

$$||(\boldsymbol{N} - c\boldsymbol{I})v|| = ||(\boldsymbol{N}^* - c^*\boldsymbol{I})v|| \tag{3.72}$$

using the positive-definiteness of norms.

We are now ready to prove the spectral theorem. Fix an arbitrary normal operator $\boldsymbol{N}$ on V^n. Since the assumption throughout this chapter has been that V^n is complex, every operator on V^n has at least one eigenvector (by the argument in section 1.9). So $\boldsymbol{N}$ must have an eigenvector $|v\rangle \in V^n$ with corresponding eigenvalue c_1. Let $\boldsymbol{P}_1 = \boldsymbol{P}_{U_1}$ be the projection onto the c_1-eigenspace of $\boldsymbol{N}$, and consider the subspace $U_1^\perp$. This subspace is **invariant** under $\boldsymbol{N}$, meaning that the action of $\boldsymbol{N}$ on any $|u\rangle \in U_1^\perp$ produces another vector lying within the subspace $U_1^\perp$. The invariance of $U_1^\perp$ under $\boldsymbol{N}$ follows from the fact that for any $|v\rangle \in U_1$,

$$\langle v|\boldsymbol{N}u\rangle = \langle \boldsymbol{N}^*v|u\rangle = \langle c_1^*v|u\rangle = c_1\langle v|u\rangle = 0, \tag{3.73}$$

using the claim of the previous paragraph in the second step. Since $\boldsymbol{N}$ leaves $U_1^\perp$ invariant, the restriction of the mapping $\boldsymbol{N}$ to $U_1^\perp$ is a normal operator on the (finite-dimensional) complex inner product space $U_1^\perp$. We may therefore repeat exactly the same argument within that space starting again with the observation that $\boldsymbol{N}$ must have an eigenvector within $U_1^\perp$ and corresponding eigenvalue c_2 ($\neq c_1$). Again, let $\boldsymbol{P}_2 = \boldsymbol{P}_{U_2}$ be the projection onto the c_2-eigenspace of $\boldsymbol{N}$ in $U_1^\perp$ and look at the subspace $(U_1 + U_2)^\perp$. This subspace is again left invariant under $\boldsymbol{N}$, and therefore must contain another of its eigenspaces, etc. Since V^n is n-dimensional, this argument must terminate after $m \leq n$ iterations (at the point where $(U_1 + U_2 + \cdots + U_m)^\perp$ is the zero subspace) leaving us with a resolution of the identity $\{\boldsymbol{P}_i\}_{i=1}^m$, consisting of projections onto all the different eigenspaces of $\boldsymbol{N}$, and a set of distinct complex numbers $\{c_i\}_{i=1}^m$, the corresponding eigenvalues. We now claim, first, that $\boldsymbol{N}$ may be expanded as

$$\boldsymbol{N} = \sum_{i=1}^{m} c_i\boldsymbol{P}_i, \tag{3.74}$$

and, second, that there is no other resolution of the identity and no other set of distinct complex numbers in terms of which $\boldsymbol{N}$ may be so expanded.

For the first claim, note that for any $|v\rangle \in V^n$, $\boldsymbol{P}_i|v\rangle$ lies in the c_i-eigenspace of $\boldsymbol{N}$, so that $\boldsymbol{N}(\boldsymbol{P}_i|v\rangle) = c_i(\boldsymbol{P}_i|v\rangle)$. Therefore, for any $|v\rangle \in V^n$,

$$\begin{aligned} \boldsymbol{N}|v\rangle = \boldsymbol{N}\boldsymbol{I}|v\rangle = \boldsymbol{N}\left(\sum_{i=1}^{m} \boldsymbol{P}_i\right)|v\rangle \qquad (3.75)\\ = \sum_{i=1}^{m} \boldsymbol{N}\boldsymbol{P}_i|v\rangle \qquad (3.76)\\ = \sum_{i=1}^{m} c_i\boldsymbol{P}_i|v\rangle = \left(\sum_{i=1}^{m} c_i\boldsymbol{P}_i\right)|v\rangle. \qquad (3.77) \end{aligned}$$

For the second claim (of uniqueness), suppose there were another way to expand $\boldsymbol{N}$ as

$$\boldsymbol{N} = \sum_{j=1}^{m'} c'_j\boldsymbol{P}'_j \qquad (3.78)$$

with $\{\boldsymbol{P}'_j\}_{j=1}^{m'}$ a resolution of the identity and $\{c'_j\}_{j=1}^{m'}$ a set of distinct complex numbers. For any index k and any $|v_k\rangle$ in the range of $\boldsymbol{P}'_k$, we see that

$$\boldsymbol{N}|v_k\rangle = \sum_{j=1}^{m'} c_j\boldsymbol{P}'_j|v_k\rangle = \sum_{j=1}^{m'} c_j\delta_{jk}|v_k\rangle = c_k|v_k\rangle, \qquad (3.79)$$

so that the numbers $\{c'_j\}_{j=1}^{m'}$ must be a subset of the eigenvalues of $\boldsymbol{N}$, i.e., a subset of $\{c_i\}_{i=1}^{m}$, and, moreover, each $\boldsymbol{P}'_k$ must have a range lying within some eigenspace of $\boldsymbol{N}$. But the range of $\boldsymbol{P}'_k$ cannot be a *proper* subspace of the range of some $\boldsymbol{P}_i$—for then there would be some nonzero vector in that range that $\boldsymbol{P}'_k$ mapped to zero, contradicting the fact that $\{\boldsymbol{P}'_j\}_{j=1}^{m'}$ provides a resolution of the identity (and that $\boldsymbol{I}$ *never* maps a nonzero vector to zero). It follows that each $\boldsymbol{P}'_k$ actually *equals* some $\boldsymbol{P}_i$, and, furthermore, that the resolution of the identity $\{\boldsymbol{P}'_j\}_{j=1}^{m'}$ must coincide with the resolution of the identity $\{\boldsymbol{P}_i\}_{i=1}^{m}$. In particular, we must have $m = m'$. But we have already seen that $\{c'_j\}_{j=1}^{m'} \subseteq \{c_i\}_{i=1}^{m}$. So, in fact, it must be the case that $\{c'_j\}_{j=1}^{m'} = \{c_i\}_{i=1}^{m}$, and the proof of uniqueness is complete.

The expansion (3.74) of a normal operator $\boldsymbol{N}$ in terms of a unique linear combination of projection operators in a resolution of the identity, with the different eigenvalues of the normal operator as coefficients, is called the **spectral decomposition** of $\boldsymbol{N}$. The projection operators in the resolution of identity are called the **spectral projections**, or **eigenprojections**, of $\boldsymbol{N}$.

We can learn alot about normal operators through their spectral decompositions. For example, we have already learned that the eigenspaces of a normal operator $\boldsymbol{N}$ are mutually orthogonal and span V^n. So, in particular, choosing an orthonormal basis within each eigenspace of $\boldsymbol{N}$ and taking the union of all the choices, we obtain an orthonormal basis for V^n consisting entirely of eigenvectors of $\boldsymbol{N}$, called an **eigenbasis** for $\boldsymbol{N}$. Thus every vector in V^n can be expanded in terms of eigenvectors of any normal operator.

In the case of a self-adjoint operator, we can immediately infer from the uniqueness of its spectral decomposition that all its eigenvalues must be real (which, of course, we already knew). And we can infer from the following simple computation (and the uniqueness of the spectral resolution of the identity operator itself) that any eigenvalue c_i of a unitary operator $\boldsymbol{U}$ must satisfy $|c_i|^2 = 1$:

$$\boldsymbol{I} = \boldsymbol{U}\boldsymbol{U}^* = \left(\sum_{i=1}^{m} c_i \boldsymbol{P}_i\right)\left(\sum_{j=1}^{m} c_j^* \boldsymbol{P}_j^*\right), \tag{3.80}$$

$$= \sum_{i,j=1}^{m} c_i c_j^* \boldsymbol{P}_i \boldsymbol{P}_j, \tag{3.81}$$

$$= \sum_{i,j=1}^{m} c_i c_j^* \delta_{ij} \boldsymbol{P}_i^2 = \sum_{i=1}^{m} |c_i|^2 \boldsymbol{P}_i. \tag{3.82}$$

Finally, if we take the tensor product of two normal operators, insert their spectral decompositions, and expand the result using bilinearity of $\otimes$, i.e.,

$$\boldsymbol{N}_1 \otimes \boldsymbol{N}_2 = \sum_{i=1}^{m} c_i \boldsymbol{P}_i \otimes \sum_{j=1}^{m'} c_j' \boldsymbol{P}_j' = \sum_{i,j=1}^{m,m'} c_i c_j' \boldsymbol{P}_i \otimes \boldsymbol{P}_j', \tag{3.83}$$

then the resulting product decomposition will be the *spectral* decomposition for $\boldsymbol{N}_1 \otimes \boldsymbol{N}_2$ (itself normal) if and only if the products $\{c_i c_j'\}_{i,j=1}^{m,m'}$ are all distinct. So one could start with factor operators with no degeneracy, and yet their tensor product might turn out to be highly degenerate depending upon what the products of their eigenvalues are. The moral is that the eigenspaces of the product of two normal operators are *not* generally given by tensor products of the eigenspaces of its factor operators.

3.7 Commuting and Compatible Self-Adjoint Operators

Two self-adjoint operators commute if and only if all their spectral projections commute with each other. For let the two operators be $\boldsymbol{F}$ and $\boldsymbol{G}$, with spectral decompositions

$$\boldsymbol{F} = \sum_{i=1}^{m} r_i \boldsymbol{P}_i \text{ and } \boldsymbol{G} = \sum_{j=1}^{m'} r'_j \boldsymbol{P}'_j. \tag{3.84}$$

Clearly, then, assuming that all the eigenprojections above commute with each other, the fact that $\boldsymbol{F}$ and $\boldsymbol{G}$ are simply linear combinations thereof entails $[\boldsymbol{F}, \boldsymbol{G}] = \boldsymbol{0}$. For the converse, fix an arbitrary i and consider any vector $|v_i\rangle$ in the range of $\boldsymbol{P}_i$, i.e., any eigenvector in the r_i-eigenspace of $\boldsymbol{F}$. Since $[\boldsymbol{F}, \boldsymbol{G}] = \boldsymbol{0}$, we have

$$\boldsymbol{F}|v_i\rangle = r_i|v_i\rangle \;\Rightarrow\; \boldsymbol{GF}|v_i\rangle = r_i\boldsymbol{G}|v_i\rangle \;\Rightarrow\; \boldsymbol{F}(\boldsymbol{G}|v_i\rangle) = r_i(\boldsymbol{G}|v_i\rangle) \tag{3.85}$$

which shows that the r_i-eigenspace of $\boldsymbol{F}$ is invariant under $\boldsymbol{G}$. It follows that for any $|v\rangle \in V^n$, $\boldsymbol{P}_i\boldsymbol{G}\boldsymbol{P}_i|v\rangle = \boldsymbol{G}\boldsymbol{P}_i|v\rangle$ since $\boldsymbol{P}_i|v\rangle$, and therefore $\boldsymbol{G}\boldsymbol{P}_i|v\rangle$, lies in the range of $\boldsymbol{P}_i$. So we have $\boldsymbol{P}_i\boldsymbol{G}\boldsymbol{P}_i = \boldsymbol{G}\boldsymbol{P}_i$, and taking adjoints of both sides, we also have $\boldsymbol{P}_i\boldsymbol{G}\boldsymbol{P}_i = \boldsymbol{P}_i\boldsymbol{G}$. Together these equalities establish that $[\boldsymbol{P}_i, \boldsymbol{G}] = \boldsymbol{0}$. Now just rerun this entire argument with $\boldsymbol{G}$ playing the role of $\boldsymbol{F}$ and $\boldsymbol{P}_i$ the role of $\boldsymbol{G}$. The conclusion is that $[\boldsymbol{P}_i, \boldsymbol{P}'_j] = \boldsymbol{0}$ for arbitrary j, so all the spectral projections of $\boldsymbol{F}$ and $\boldsymbol{G}$ indeed commute, as promised.

Next, suppose we are given any finite collection of self-adjoint operators. Then they will mutually commute if and only if they share an eigenbasis. Suppose, first, that self-adjoint operators $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ etc. have a common eigenbasis. For any two of these operators, say $\boldsymbol{B}$ and $\boldsymbol{C}$, obviously the action of the commutator $[\boldsymbol{B}, \boldsymbol{C}]$ on any vector in the shared eigenbasis will yield $|0\rangle$. So, since every vector in the space is a linear combination of basis vectors, $[\boldsymbol{B}, \boldsymbol{C}] = \boldsymbol{0}$. Conversely, suppose that self-adjoint $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, etc. mutually commute. Let their eigenprojections be given by $\{\boldsymbol{P}_i\}_{i=1}^{m_\alpha}$, $\{\boldsymbol{P}'_j\}_{j=1}^{m_\beta}$, $\{\boldsymbol{P}''_k\}_{k=1}^{m_\gamma}$, etc. and consider the set of products (focussing only on the ones that are nonzero):

$$\left\{\boldsymbol{P}_i\boldsymbol{P}'_j\boldsymbol{P}''_k\cdots\right\}_{i,j,k,\ldots=1}^{m_\alpha,m_\beta,m_\gamma,\ldots}. \tag{3.86}$$

Since $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, etc. commute, so do all their eigenprojections (by the previous paragraph's argument), so that each product of projections in the set above is itself a projection. Moreover, since each operator's eigenprojections resolve the identity, the set of product projections above does so as well (verify!). But the range of $\boldsymbol{P}_i\boldsymbol{P}'_j\boldsymbol{P}''_k\cdots$ is contained within $\boldsymbol{A}$'s ith eigenspace, $\boldsymbol{B}$'s jth eigenspace, etc. Therefore, there is a set of mutually orthogonal subspaces spanning V^n each of which contains simultaneous eigenvectors for all of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, etc. To obtain a common eigen*basis*, simply select an orthonormal basis within each of these subspaces and take the union of all the selections.

You should convince yourself of the following corollary to the result we have just established: commuting self-adjoint operators share a common eigenbasis if and only if they can be **simultaneously diagonalized**, i.e., when there is an orthonormal basis relative to which the matrix representations for all the operators assume a diagonal form (with zero entries off the main diagonal). If there is a *unique* basis that simultaneously diagonalizes a set of commuting self-adjoint operators, or equivalently, if the set shares but a single eigenbasis, then it is called **complete**. For example, any resolution of the identity into one-dimensional projections is a complete commuting set of self-adjoint operators.

To introduce the notion of compatible self-adjoint operators, we first need to discuss functions of a self-adjoint operator. Consider any real-valued polynomial function on the real line:

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x^1 + a_0. \tag{3.87}$$

Since the operators on V^n are an algebra, we may form the corresponding *operator-valued* polynomial function of any self-adjoint operator $\boldsymbol{F} \in \mathcal{A}(V^n)$:

$$p(\boldsymbol{F}) = a_n\boldsymbol{F}^n + a_{n-1}\boldsymbol{F}^{n-1} + \cdots + a_1\boldsymbol{F}^1 + a_0\boldsymbol{I}. \tag{3.88}$$

Inserting $\boldsymbol{F}$'s spectral decomposition $\sum_{i=1}^m r_i\boldsymbol{P}_i$ in place of $\boldsymbol{F}$ in (3.88) and simplifying yields

$$p(\boldsymbol{F}) = \sum_{i=1}^m p(r_i)\boldsymbol{P}_i. \tag{3.89}$$

Thus a polynomial function of a self-adjoint operator $\boldsymbol{F}$ has identically the same eigenprojections as $\boldsymbol{F}$, and its eigenvalues are just the corresponding polynomial functions of the eigenvalues of $\boldsymbol{F}$.

(3.89) motivates defining $f(\boldsymbol{F})$, where f is *any* real-valued function on the real line, to be

$$f(\boldsymbol{F}) \overset{\text{def}}{=} \sum_{i=1}^{m} f(r_i)\boldsymbol{P}_i. \tag{3.90}$$

However, since the operators we are considering in this chapter possess only finitely many eigenvalues, any operator that is a function of $\boldsymbol{F}$ in the more general sense of (3.90) is always some *polynomial* function of $\boldsymbol{F}$. For, given any self-adjoint $\boldsymbol{F}$ and function f, we can construct out of the (distinct) eigenvalues of $\boldsymbol{F}$ the (well-defined) polynomial

$$p(x) = \sum_{i=1}^{m} \left(\frac{(x-r_1)\cdots(x-r_{i-1})(x-r_{i+1})\cdots(x-r_m)}{(r_i-r_1)\cdots(r_i-r_{i-1})(r_i-r_{i+1})\cdots(r_i-r_m)} f(r_i) \right). \tag{3.91}$$

Because the ith term in this summation takes the value $\delta_{ij} f(r_i)$ when $x = r_j$, p agrees with f in its action on all the eigenvalues of $\boldsymbol{F}$, which agreement suffices for $p(\boldsymbol{F}) = f(\boldsymbol{F})$.

As an example of a frequently invoked function of a self-adjoint operator $\boldsymbol{F}$, consider the sum of the spectral projections of $\boldsymbol{F}$ that correspond to some subset Δ of its eigenvalues:

$$\boldsymbol{P}_\Delta = \sum_{\{i:r_i\in\Delta\}} \boldsymbol{P}_i. \tag{3.92}$$

This sum is itself a projection, whose range coincides with the span of the eigenspaces corresponding to the values in Δ. It is easy to see that $\boldsymbol{P}_\Delta$ is a characteristic function χ_Δ of $\boldsymbol{F}$, where χ_Δ is defined by

$$\chi_\Delta(x) = \begin{cases} 1 \text{ if } x \in \Delta, \\ 0 \text{ if } x \notin \Delta. \end{cases} \tag{3.93}$$

In particular, the spectral projections of $\boldsymbol{F}$ themselves are all (particular) characteristic functions of $\boldsymbol{F}$.

A set of self-adjoint operators is called **jointly compatible** if there is some self-adjoint operator that they are all a function of. When the self-adjoint operators are projections, and there are just two of them, this coincides with our previous definition of compatibility between projections (section 2.6). For if two projections are compatible, they have compatible ranges U and W that satisfy $U = A+B$ and $W = A+C$ for some mutually orthogonal subspaces A, B, and C. We can then choose a self-adjoint operator $\boldsymbol{F}$ that has A, B, and C amongst its eigenspaces and write the projections onto U and W as characteristic functions of $\boldsymbol{F}$. Conversely, if two projections are both functions of a single self-adjoint operator, then obviously they must commute, and *ipso facto* be compatible (cf. (3.59)).

Just as compatibility and commutativity of projections come down to the same thing, so also for self-adjoint operators generally. Obviously any set of jointly compatible self-adjoint operators must commute in virtue of [their] all being functions of a single self-adjoint operator. Conversely, suppose self-adjoint $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, etc. mutually commute. Then there is an orthonormal basis $\{e_i\}_{i=1}^n$ with respect to which they can be simultaneously diagonalized. If a_i is the eigenvalue of $\boldsymbol{A}$ corresponding to $|e_i\rangle$, and similarly for b_i and $\boldsymbol{B}$, c_i and $\boldsymbol{C}$, etc., then it is clear that these operators can be written as

$$\boldsymbol{A} = \sum_{i=1}^{n} a_i |e_i\rangle\langle e_i|, \ \boldsymbol{B} = \sum_{i=1}^{n} b_i |e_i\rangle\langle e_i|, \ \boldsymbol{C} = \sum_{i=1}^{n} c_i |e_i\rangle\langle e_i|, \text{ etc.} \tag{3.94}$$

(which are not necessarily their *spectral* decompositions, since they may be degenerate). To obtain a self-adjoint operator of which $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, etc. are all a function, just pick an operator $\boldsymbol{F}$ with spectral decomposition:

$$\boldsymbol{F} = \sum_{i=1}^{n} r_i |e_i\rangle\langle e_i|. \tag{3.95}$$

Because the r_i's are all distinct, there exist well-defined functions f, g, h, etc. satisfying, for all i:

$$f(r_i) = a_i, \ g(r_i) = b_i, \ h(r_i) = c_i, \ \text{etc.}, \tag{3.96}$$

from which it is evident that $\boldsymbol{A} = \boldsymbol{f}(\boldsymbol{F})$, $\boldsymbol{B} = \boldsymbol{g}(\boldsymbol{F})$, $\boldsymbol{C} = \boldsymbol{h}(\boldsymbol{F})$, etc. as desired.

Notes and References

Clear and nearly exhaustive treatments of finite-dimensional vector spaces can be found in Lipschutz (1968) and Halmos (1948). Section 1.6's proof that ℓ has uncountable dimension was communicated to us by John L. Bell. For concise and fairly general discussions of arbitrary vector spaces, associative, and Lie algebras—including exercises and applications to modern physics—see Chs. 9–23 in Geroch (1985). MacLane and Birkhoff (1979) is a classic text on algebras. Jordan algebras were first introduced by one of the co-founders of quantum theory, Pascual Jordan (1932), with the axiomatization of the theory in mind. Shortly thereafter, Jordan's collaboration with von Neumann and Wigner (1934) produced a characterization of a large class of finite-dimensional Jordan algebras, and the literature on Jordan algebras is now voluminous (e.g., see Jacobson (1968)). There are also numerous sources for lattice theory, though Birkhoff's (1967) is probably the bible. The term 'entangled' was originally coined by Schrödinger.