Chapter 8

Vector Spaces in Quantum Mechanics

We have seen in the previous Chapter that there is a sense in which the state of a quantum system can be thought of as being made up of other possible states. The aim here is to use the example of the Stern-Gerlach experiment to develop this idea further, and to show that the states of a quantum system can be represented by vectors in a complex vector space. To begin, we review some of the basic ideas of vectors, using the example of vectors in real two dimensional space.

8.1 Vectors in Two Dimensional Space

The primitive prototype of what we mean by a vector is provided by the position vector in ordinary space. Below is a summary of the important properties of such vectors in physical space based on their interpretation as mathematical objects that have both magnitude and direction. As well as position, velocity and force and so on are also examples of such vectors. The intention is not to give a complete discussion, but to highlight a number of important properties of such vectors that have analogues in the case of quantum states, including the property that two vectors can be combined to produce another vector, and that 'how much' of one vector is contained in another can be measured via the inner product of two vectors.

8.1.1 Linear Combinations of Vectors – Vector Addition

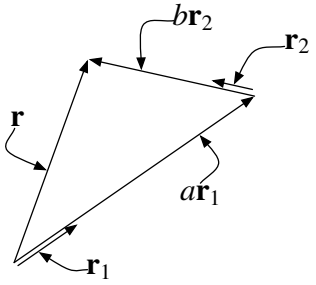

Figure 8.1: An example of vector addition. Two basis vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ combined to give $\mathbf{r}$.

Consider two non-collinear vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, as illustrated in Fig. (8.1). Multiples of these pair of vectors can be added together in accordance with the triangle addition rule illustrated in Fig. (8.1) to form another vector. Conversely, any other vector $\mathbf{r}$ can be expressed in terms of $\mathbf{r}_1$ and $\mathbf{r}_2$ using appropriate values for the (real number) components a and b of $\mathbf{r}$, i.e.

$$\mathbf{r} = a\mathbf{r}_1 + b\mathbf{r}_2. \tag{8.1}$$

The right hand side of this equation is known as a *linear combination* of the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. The particular points to take away from this is that combining vectors produces other vectors, analogous to our observation above that combining states produces other states, and that the components a and b are a measure of how much of $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively go towards making up the vector $\mathbf{r}$.

The two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ introduced above are arbitrary except insofar as they are not collinear. What we mean by collinear is that there is no way that multiples of these vectors can be combined to produce the zero vector, or, in other words,

$$a\mathbf{r}_1 + b\mathbf{r}_2 = 0 \quad \Longrightarrow \quad a, b = 0. \tag{8.2}$$

The usual terminology is to say that these two vectors are *linearly independent*.

Further, as any vector in the plane can be written as a linear combination of $\mathbf{r}_1$ and $\mathbf{r}_2$, they act as the basic building blocks for any vector in the plane, and hence are known as *basis vectors*. There is effectively an infinite number of choices for the basis vectors, and in fact it is possible to choose three or more vectors to be basis vectors. But the minimum number is two, if we wish to be able to describe any vector in the plane as a linear combination of basis vectors. The collection of all the vectors that can be constructed by taking linear combinations of these basis vectors using any real numbers a and b as components is known as a real vector space, and since two basis vectors are needed, the vector space is said to be of dimension two.

This vector space possess more structure than that implied by simply forming various linear combinations. The various vectors that can be drawn in a plane, as in Fig. (8.1), i.e. for which the coefficients a and b take any real value, can be clearly seen to have different lengths and relative orientations. These properties of vectors are encompassed in the definition of the inner, scalar or dot product of pairs of vectors.

8.1.2 Inner or Scalar Products

The inner or scalar product of the two position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined by

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = r_1 r_2 \cos\theta \tag{8.3}$$

where r_1 and r_2 are the lengths of $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively, and θ is the angle between them.

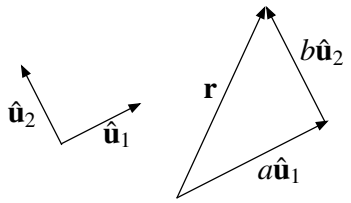

Figure 8.2: An example of a pair of unit vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ combining to give $\mathbf{r}$.

Using the idea of an inner or scalar product $\mathbf{r}_1 \cdot \mathbf{r}_2$ it is possible to introduce a particularly useful pair of basis vectors. To this end, consider two vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ that satisfy

$$\hat{\mathbf{u}}_1 \cdot \hat{\mathbf{u}}_1 = \hat{\mathbf{u}}_2 \cdot \hat{\mathbf{u}}_2 = 1 \tag{8.4}$$

i.e. they have unit length, hence they are known as unit vectors, and

$$\hat{\mathbf{u}}_1 \cdot \hat{\mathbf{u}}_2 = 0 \tag{8.5}$$

i.e. they are orthogonal. This pair of vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ is said to be orthonormal. As before, they form a basis in that multiples of each vector can be added together in the fashion illustrated in Fig. (8.2) to form another vector. Conversely, any vector $\mathbf{r}$ can be expressed using appropriate values for the components of $\mathbf{r}$, i.e.

$$\mathbf{r} = a\hat{\mathbf{u}}_1 + b\hat{\mathbf{u}}_2. \tag{8.6}$$

The components a and b, which represent 'how much' of the vector $\mathbf{r}$ is made up of the vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ are given by

$$a = \hat{\mathbf{u}}_1 \cdot \mathbf{r} \quad \text{and} \quad b = \hat{\mathbf{u}}_2 \cdot \mathbf{r} \tag{8.7}$$

A well known example of a vector expressed as a linear combination of a pair of unit vectors is given by the position vector $\mathbf{r}$ written with respect to the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$:

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}}. \tag{8.8}$$

The position vector is the 'original' or prototypical vector in the sense that the properties of position vectors can be generalized, essentially as a creative exercise in pure mathematics, so as to arrive at the notion of an abstract vector which has nothing to do with position in space, but which nevertheless has mathematical properties firmly founded on the intuitive properties assigned to position vectors. This generalization is central to arriving at the mathematical language of quantum mechanics.

8.2 Generalization to higher dimensions and complex vectors

The above properties of position vectors are built around basic concepts of two, or perhaps, three dimensional physical space. But the mathematics itself, considered as an abstract entity in its own right, can be extended in a number of ways.

Arbitrary dimensions Space has three dimensions: e.g. for the x, y, and z directions in space we require three basis vectors, the familiar unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, which we will here refer to as $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$ and $\hat{\mathbf{u}}_3$ respectively. These basis vectors are sufficient to construct any arbitrary vector by taking appropriate linear combinations. But we can imagine a 'space' with four, five or more dimensions, including infinity. This would require defining more orthonormal unit vectors, $\hat{\mathbf{u}}_4, \hat{\mathbf{u}}_5, \ldots, \hat{\mathbf{u}}_N$ where

$$\begin{aligned} \hat{\mathbf{u}}_n \cdot \hat{\mathbf{u}}_m &= 1 \quad n = m \\ &= 0 \quad n \neq m \end{aligned} \tag{8.9}$$

so that a typical vector would look like

$$\mathbf{v} = a_1\hat{\mathbf{u}}_1 + a_2\hat{\mathbf{u}}_2 + \ldots + a_N\hat{\mathbf{u}}_N \tag{8.10}$$

and by the orthonormality property of the basis vectors, we can specify 'how much' of each of these basis vectors contributes towards the final vector $\mathbf{v}$, i.e.

$$a_n = \hat{\mathbf{u}}_n \cdot \mathbf{v}. \tag{8.11}$$

Clearly we are no longer dealing with ordinary three dimensional space; in fact we are dealing with a space of dimension N.

Complex components The next generalization we can make is to relax the requirement that the coefficients or 'components' a_n be real numbers; we can take them to be complex in general. Thus, for instance, we are going to admit vectors like

$$\mathbf{v} = (1 + 2i)\hat{\mathbf{i}} + (3 - 4i)\hat{\mathbf{j}} \tag{8.12}$$

or more generally, we would allow the coefficients a_n appearing in Eq. (8.10) to be complex.

Inner product Finally, the idea of an inner product has to be looked at again. If we had two complex vectors, such as $\mathbf{v}$ in Eq. (8.12), and some other complex vector $\mathbf{u}$, then it might seem obvious to take the inner product to be simply $\mathbf{u} \cdot \mathbf{v}$. However, there are good reasons to not follow this simple path, the most obvious concerning the issue of defining the 'length' of such a vector. In the case of a real position vector, the length r of vector $\mathbf{r}$ can be written in terms of the scalar product of $\mathbf{r}$ with itself as $r = \sqrt{\mathbf{r} \cdot \mathbf{r}}$. But if we were to adopt this for the complex vector $\mathbf{v}$ given above then we find that

$$\mathbf{v} \cdot \mathbf{v} = (1 + 2i)^2 + (3 - 4i)^2 = -10 - 20i$$

which is a complex number. So, the 'length' $v = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ would also be a complex number, which is not a satisfactory state of affairs if we needed to have the concept of length for complex vectors: there is no way to be able to say that one vector was 'longer' than another. In contrast, however

$$\mathbf{v}^* \cdot \mathbf{v} = |1 + 2i|^2 + |3 - 4i|^2 = 30$$

which is real, and positive, so we can define a length of a complex vector as $v = \sqrt{\mathbf{v}^* \cdot \mathbf{v}}$. This then suggests that we adopt the following definition:

$$\text{The inner product of the two complex vectors } \mathbf{u} \text{ and } \mathbf{v} = \mathbf{u}^* \cdot \mathbf{v}. \tag{8.13}$$

One obvious consequence of this is that the order of the factors matters, i.e.

$$\mathbf{u}^* \cdot \mathbf{v} \neq \mathbf{v}^* \cdot \mathbf{u}. \tag{8.14}$$

In fact,

$$\mathbf{u}^* \cdot \mathbf{v} = (\mathbf{v}^* \cdot \mathbf{u})^*. \tag{8.15}$$

Orthonormal vectors now satisfy

$$\begin{aligned} \hat{\mathbf{u}}_n^* \cdot \hat{\mathbf{u}}_m &= 1 \quad n = m \\ &= 0 \quad n \neq m \end{aligned} \tag{8.16}$$

Complex basis vectors are not exotic mathematical entities. A simple example is

$$\begin{aligned} \hat{\mathbf{u}}_1 &= \frac{3\mathbf{i} + 4i\mathbf{j}}{5} \\ \hat{\mathbf{u}}_2 &= \frac{4\mathbf{i} - 3i\mathbf{j}}{5} \end{aligned} \tag{8.17}$$

which can be easily shown to be orthonormal. For instance

$$\hat{\mathbf{u}}_1^* \cdot \hat{\mathbf{u}}_1 = \frac{3\mathbf{i} - 4i\mathbf{j}}{5} \cdot \frac{3\mathbf{i} + 4i\mathbf{j}}{5} = \frac{9 + 16}{25} = 1$$

and so on.

Thus by fairly straightforward generalizations of the basic concept of a position vector it is possible to construct a more abstract notion of a vector which lives in a vector space of arbitrary dimensions, and which can have complex components. In other words, the study of the properties of ordinary vectors and their generalizations as outlined above, leads to a set of rules that these vectors obey. If it turns out that in some context which *a priori* has nothing to do with vectors in ordinary space, there arises mathematical quantities that obey these same set of rules, then these mathematical quantities will be looked upon as vectors. If we call these quantities $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ ..., then what we must have is

- A rule that enables us to 'add' multiples of these quantities together so as to generate another such object

  $$a\mathbf{u} + b\mathbf{v} = \mathbf{w}$$

  where a and b are complex numbers. For ordinary vectors, the rule was simply the triangle rule for adding position vectors. For more abstract vectors, the rule could be much more complicated.

- A rule that tells us how to define, and hence calculate, the inner product $\mathbf{u}^* \cdot \mathbf{v}$. For ordinary vectors, we made use of the pre-existing notions of the length of a vector and the angle between vectors to give us our definition. In the general case, we provide a definition, and the notion of the length and angle between vectors emerges from the definition!! Often, in pure mathematical contexts, the notation $(\mathbf{u}, \mathbf{v})$ is used to indicate an inner product rather than the dot product notation. It is this bracket notation that segues into the 'bra-ket' notation due to Dirac. Thus we require that the inner product satisfy the requirements:

  $$\begin{aligned} (\mathbf{u}, \mathbf{v}) &= \text{a complex number} \\ (\mathbf{u}, \mathbf{v})^* &= (\mathbf{v}, \mathbf{u}) \\ (\mathbf{v}, \mathbf{v}) &\geq 0 \\ (\mathbf{w}, a\mathbf{u} + b\mathbf{v}) &= a(\mathbf{w}, \mathbf{u}) + b(\mathbf{w}, \mathbf{v}) \end{aligned}$$

What we will see in the case of quantum states is that the mathematical properties of these states *as determined by their physical meaning* provide us with both these set of rules: i.e. we find that we can 'add' states to generate other states, and we can 'multiply' these states in a way that obeys the rules of an inner product. Thus we conclude that quantum states can be treated as abstract vectors. All the mathematical machinery of vectors then follows: vector spaces, unit vectors, linear independence, dimension ....

In the following Section, these parallels between what we have seen here and the behaviour of quantum states is made explicit.

8.3 Spin Half Quantum States as Vectors

We now need to examine the way in which the quantum states for a spin half system can be seen to fit in with the idea of being considered as vectors. To see how this comes about, we will show that there is a perfect analogy between Eqs. (7.18) and (7.19) and corresponding relationships for ordinary (complex) vectors. Returning to the Stern-Gerlach example discussed in the preceding Chapter we obtained there an expression Eq. (7.19)

$$|S\rangle = |+\rangle\langle +|S\rangle + |-\rangle\langle -|S\rangle.$$

for the state of a spin half atom expressed in terms of the states $|\pm\rangle$, which are states for which the atom has a z component of spin equal to $\pm\frac{1}{2}\hbar$, and $\langle\pm|S\rangle$ are probability amplitudes (complex numbers) whose magnitude in some sense tells us 'how much' of the states $|\pm\rangle$ are to be found in the state $|S\rangle$. This result was obtained by 'cancelling' the common factor '$\langle S'|$' from the result

$$\langle S'|S\rangle = \langle S'|+\rangle\langle +|S\rangle + \langle S'|-\rangle\langle -|S\rangle.$$

What we should bear in mind here is that we can recover this relationship between probability amplitudes by reintroducing '$\langle S'|$' into Eq. (7.19) for any chosen final state, yielding an expression for the probability amplitudes as needed. Thus, as has been said before, Eq. (7.19) effectively represents a 'template' into which we insert the appropriate information in order to recover the required probability amplitudes. We can also note that there is nothing sacred about choosing to cancel the the common factor $|S\rangle$ – we could equally as well cancel the factor $|S\rangle$, yielding

$$\langle S'| = \langle S'|+\rangle\langle +| + \langle S'|-\rangle\langle -|. \tag{8.18}$$

Having carried out this cancellation procedure, what has reappeared is the state of a quantum system i.e. $|S\rangle$ which was introduced earlier in a different context, specifically as being nothing more than a way of writing down all that we knew about the state of a quantum system. There, the notation had no mathematical significance, but in the manner in which it appears here, it seems to have acquired a mathematical meaning of some kind. The aim is to see what this meaning might be, and in doing so, we will show that the expression for $|S\rangle$ has many of the properties that we associate with expressing a vector as a sum of its components.

We begin by considering the probability amplitudes $\langle S'|S\rangle$ themselves. These are complex numbers in general for arbitrary spin directions, (but they were real in the particular Stern-Gerlach example used above), such that their modulus squared $|\langle S'|S\rangle|^2$ is the probability $P(S'|S)$ of observing the spin to be in the state $|S'\rangle$ given that it was in the state $|S\rangle$. In particular, $\langle S|S\rangle$ is the probability amplitude of observing the spin to be in the state $|S\rangle$ given that it was in the state $|S\rangle$. This will have to be unity, i.e. $P(S|S) = |\langle S|S\rangle|^2 = 1$. Thus we can conclude that

$$\langle S|S\rangle = e^{i\eta} \tag{8.19}$$

where η is an arbitrary phase. It turns out that this phase always cancels out in any calculation of observable quantities, so it is conventionally set to zero, and hence

$$\langle S|S\rangle = 1. \tag{8.20}$$

The state $|S\rangle$ is said to be normalized to unity. As a particular case, this last result implies that

$$\langle +|+\rangle = 1. \tag{8.21}$$

We can now consider the probability amplitude $\langle +|S\rangle$ obtained by replacing S' by + in the above expression for $\langle S'|S\rangle$:

$$\langle +|S\rangle = \langle +|+\rangle\langle +|S\rangle + \langle +|-\rangle\langle -|S\rangle. \tag{8.22}$$

We have seen that we can put $\langle +|+\rangle = 1$, so we have

$$\langle +|-\rangle\langle -|S\rangle = 0 \tag{8.23}$$

which has to be true no matter what the state $|S\rangle$ happens to be, i.e. no matter what value the probability amplitude $\langle -|S\rangle$ is. Thus we conclude that

$$\langle +|-\rangle = 0. \tag{8.24}$$

What this is telling us is that if an atom has been measured as being in the state $|-\rangle$, for instance, it has emerged from a Stern-Gerlach apparatus in the $S_z = -\frac{1}{2}\hbar$ beam, and if it is then passed through a second such apparatus, then there will be zero probability of it emerging in the $S_z = \frac{1}{2}\hbar$ beam (it has a 100% chance of emerging in the $S_z = -\frac{1}{2}\hbar$ beam). The two states $|+\rangle$ and $|-\rangle$ represent *mutually exclusive* possibilities.

Similarly we can show that

$$\langle -|-\rangle = 1 \quad \text{and} \quad \langle -|+\rangle = 0. \tag{8.25}$$

Thus we can set up a comparison:

$$\begin{aligned}
\langle +|+\rangle = 1 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_1^* \cdot \hat{\mathbf{u}}_1 = 1 \\
\langle +|-\rangle = 0 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_1 = 0 \\
\langle -|-\rangle = 1 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_2 = 1 \\
\langle -|+\rangle = 0 \quad &\longleftrightarrow \quad \hat{\mathbf{u}}_2^* \cdot \hat{\mathbf{u}}_1 = 0
\end{aligned} \tag{8.26}$$

where we have chosen to make the comparison between the probability amplitudes and the inner product of complex unit vectors as we are dealing with probability amplitudes that are, in general, complex numbers. This comparison implies the following correspondences:

$$\begin{aligned}
|+\rangle &\longleftrightarrow \hat{\mathbf{u}}_1 \quad |-\rangle \quad \longleftrightarrow \hat{\mathbf{u}}_2 \\
\langle +| &\longleftrightarrow \hat{\mathbf{u}}_1^* \quad \langle -| \quad \longleftrightarrow \hat{\mathbf{u}}_2^*.
\end{aligned} \tag{8.27}$$

We know that $\langle +|S\rangle$ and $\langle -|S\rangle$ are both just complex numbers, so call them a and b respectively. If we now write

$$|S\rangle = a|+\rangle + b|-\rangle \tag{8.28}$$

we establish a perfect correspondence with the expression

$$\mathbf{v} = a\,\hat{\mathbf{u}}_1 + b\,\hat{\mathbf{u}}_2. \tag{8.29}$$

On the basis of this result, we are then tempted to interpret the ket $|S\rangle$ as a vector expressed as a linear combination of two orthonormal basis vectors $|\pm\rangle$. We can push the analogy further if we once again use the fact that $\langle S|S\rangle = 1$, so that

$$\langle S|S\rangle = 1 = \langle S|-\rangle\langle -|S\rangle + \langle S|+\rangle\langle +|S\rangle \tag{8.30}$$

On the other hand, the total probability of observing the system in either of the states $|\pm\rangle$ must add up to unity, which means that

$$P(+|S) + P(-|S) = |\langle +|S\rangle|^2 + |\langle -|S\rangle|^2 = 1. \tag{8.31}$$

By comparing the last two equations, and noting that

$$|\langle \pm|S\rangle|^2 = \langle \pm|S\rangle\langle \pm|S\rangle^* \tag{8.32}$$

we conclude that

$$\langle \pm|S\rangle = \langle S|\pm\rangle^*. \tag{8.33}$$

If we now consider

$$\langle S'| = \langle S'|+\rangle\langle +| + \langle S'|-\rangle\langle -|$$

and use the result, Eq. (8.33), $\langle \pm|S'\rangle = \langle S'|\pm\rangle^*$, we can write this as

$$\langle S'| = \langle +|S'\rangle^*\langle +| + \langle -|S'\rangle^*\langle -| \tag{8.34}$$

or, expressed in terms of $a' = \langle +|S'\rangle$ and $b' = \langle -|S'\rangle$, we have

$$\langle S'| = a'^*\langle +| + b'^*\langle -| \tag{8.35}$$

which has a perfect correspondence with an ordinary vector $\hat{\mathbf{v}}'$ in the form

$$\mathbf{v}'^* = a'^*\hat{\mathbf{u}}_1^* + b'^*\hat{\mathbf{u}}_2^*. \tag{8.36}$$

So the bra $\langle S'|$ is itself a vector, a bra vector, which can be thought of as being just the complex conjugate of the corresponding ket vector $|S'\rangle$. But while it is occasionally useful to think this way, it is not strictly true mathematically, and this way of viewing a bra vector will not be employed here. Instead, as will be shown shortly, an interpretation of ket vectors as column vectors leads to the interpretation of bra vectors as row vectors. A more mathematical standpoint also leads to interpretation of bra vectors as 'linear functionals', that is, a bra is a mathematical operator that acts on a ket vector to produce a complex number.

Finally, to complete the correspondence, we note that the probability amplitude $\langle S'|S\rangle$ can be written

$$\langle S'|S\rangle = a'^*a + b'^*b \tag{8.37}$$

which can be compared with the inner product $\mathbf{v}'^*\cdot\mathbf{v}$

$$\mathbf{v}'^*\cdot\mathbf{v} = a'^*a + b'^*b \tag{8.38}$$

which tells us that the probability amplitude can be considered as being simply the inner product of the two vectors $|S'\rangle$ and $|S\rangle$, i.e.

$$\langle S'|S\rangle = \langle S'|\cdot|S\rangle. \tag{8.39}$$

In other words, we have a perfect analogy between the two dimensional complex vector space formed by linear combinations of the unit vectors $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ discussed in Section 8.1 and a complex vector space consisting of all the linear combinations of the states $|\pm\rangle$. The ket vectors $|\pm\rangle$ are referred to as *basis states*, analogous to $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ being referred to as basis vectors.

Different spin states can be constructed by forming linear combinations $|S\rangle = a|+\rangle + b|-\rangle$ of these basis states, with a and b being allowed to range over all the complex numbers, though we have only been looking at linear combinations with real coefficients. By limiting a and b to be real numbers, we are constructing states whose measured component of spin all lie in the same plane, which, with the system of axes we have been using here, is the XZ plane, corresponding to polar angle $\phi = 0$. If the coefficients a and b are complex, the state $|S\rangle$ represents a state in which the measured component $S = \mathbf{S}\cdot\mathbf{m}$ is along a direction $\hat{\mathbf{m}}$ that points out of this plane, as discussed in Section 8.3.2. So, any linear combination in which a and b are any complex numbers also defines a possible spin state of the spin half system. Thus all the possible linear combinations of $|\pm\rangle$, i.e.

combinations of the form $a|+\rangle + b|-\rangle$ where a and b are complex numbers form a *complex vector space* known as the *state space* of the system.

The quantum state vectors can also be 'multiplied' together – the inner product of the two vectors $|S'\rangle$ and $|S\rangle$ is just the probability amplitude $\langle S'|S\rangle$. In particular, the basis states are normalized to unity, i.e. they are unit vectors, and they are orthogonal to each other, i.e. they form a pair of *orthonormal basis states*.

The terminology often adopted is to say that the state vector $|S\rangle = a|+\rangle + b|-\rangle$ is a 'linear superposition' of the two states $|\pm\rangle$. The probability amplitudes $a = \langle +|S\rangle$ and $b = \langle -|S\rangle$ represent 'how much' of the states $|\pm\rangle$ are contained within the state $|S\rangle$ to the extent that $|\langle \pm|S\rangle|^2$ is the probability of the z component of spin being found to have the values $\pm\frac{1}{2}\hbar$.

8.3.1 The Normalization Condition

One difference between ordinary vectors and quantum state vectors is the importance of the 'normalization condition', i.e. the requirement that $\langle S|S\rangle = 1$, which must hold true given the interpretation of the inner product as a probability amplitude. But how can this be reconciled with the statement above that *any* linear combination of the basis states is a possible state of the system? How can a state vector such as $|\widetilde{S}\rangle = |+\rangle + |-\rangle$ which has the property

$$\langle \widetilde{S}|\widetilde{S}\rangle = 2 \tag{8.40}$$

be a physically acceptable state vector as it seems to be saying that the probability of finding the system in the state $|\widetilde{S}\rangle$ given that it is in the state $|\widetilde{S}\rangle$ is 4, which does not make sense. But, if we define a new vector $|S\rangle$ by

$$|S\rangle = \frac{|\widetilde{S}\rangle}{\sqrt{\langle \widetilde{S}|\widetilde{S}\rangle}} = \frac{1}{\sqrt{2}}|\widetilde{S}\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle + |-\rangle\right) \tag{8.41}$$

then automatically $|S\rangle$ will have the required normalization property – it is said to be normalized to unity. So, rather than abandoning giving a physical interpretation of state vectors which are not normalized to unity, the approach adopted is that we can multiply any state vector by any factor and say that it still represents the same *physical* state, i.e. $|S\rangle$ and $|\widetilde{S}\rangle = a|S\rangle$, where a is any complex number, represent the same physical state. However, it is only the normalized state $|S\rangle$ that should be used in any calculations in order to be certain that probability is properly accounted for.

8.3.2 The General Spin Half State

We can generalize the result obtained in the preceding Chapter for the state of a spin half system. That result applied only to the case of spin directions oriented in the XY plane. But it ought not matter how the vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ are oriented in space as the probability amplitude in Eq. (7.40) depends only on the angle between these two vectors. For arbitrary orientations in space, we can specify these vectors by their spherical polar coordinates as illustrated in Fig. (8.3).

If we suppose that $\hat{\mathbf{n}}$ has the polar coordinates (θ_i, ϕ_i) then we have:

$$\hat{\mathbf{n}} = \sin\theta_i \cos\phi_i \,\hat{\mathbf{i}} + \sin\theta_i \sin\phi_i \,\hat{\mathbf{j}} + \cos\theta_i \,\hat{\mathbf{k}} \quad (8.42)$$

and similarly for $\hat{\mathbf{m}}$, with polar coordinates (θ_f, ϕ_f). The angle γ between these two vectors can be determined from the inner product $\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}$:

$$\begin{aligned}\hat{\mathbf{n}} \cdot \hat{\mathbf{m}} &= \cos[\gamma(\theta_f, \phi_f, \theta_i, \phi_i)] \\ &= \cos\theta_i \cos\theta_f + \cos(\phi_f - \phi_i) \sin\theta_i \sin\theta_f.\end{aligned} \quad (8.43)$$

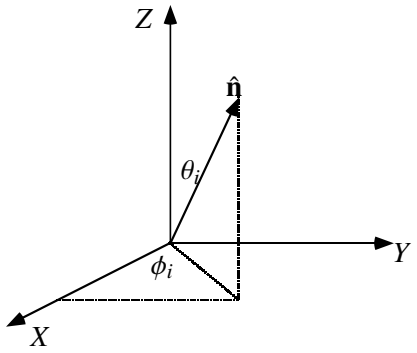

Figure 8.3: Polar angles for defining direction of unit vector $\hat{\mathbf{n}}$

We can now write for the probability amplitude $\langle S_f = \frac{1}{2}\hbar | S_i = \frac{1}{2}\hbar\rangle$ the expression

$$\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle = \cos(\tfrac{1}{2}\gamma) e^{i\xi(\theta_f, \phi_f, \theta_i, \phi_i)} \quad (8.44)$$

where $\exp[i\xi(\theta_f, \phi_f, \theta_i, \phi_i)]$ is a phase factor to be determined. It is necessary to introduce this phase factor as there is no requirement that the phases of the probability amplitudes for arbitrary orientations in space will be necessarily the same as those for the vectors in the XY plane.

We can determine this phase factor by requiring that the probability amplitudes defined in Eq. (8.44) satisfy the sum over amplitudes formula, Eq. (7.42). If we specialize this expression to the particular case of $\hat{\mathbf{l}} = \hat{\mathbf{k}}$, i.e. in the z direction, then we have

$$\begin{aligned}\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle =& \langle S_f = \tfrac{1}{2}\hbar | S_z = \tfrac{1}{2}\hbar\rangle\langle S_z = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle \\ &+ \langle S_f = \tfrac{1}{2}\hbar | S_z = -\tfrac{1}{2}\hbar\rangle\langle S_z = -\tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle \\ =& \cos[\tfrac{1}{2}\gamma(\theta_f, \phi_f, 0, 0)] \cos[\tfrac{1}{2}\gamma(0, 0, \theta_i, \phi_i)] e^{i(\xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i))} \\ &+ \cos[\tfrac{1}{2}\gamma(\theta_f, \phi_f, \pi, 0)] \cos[\tfrac{1}{2}\gamma(\pi, 0, \theta_i, \phi_i)] e^{i(\xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i))}\end{aligned} \quad (8.45)$$

The value of γ for the various values of $(\theta_f, \phi_f, \theta_i, \phi_i)$ can be evaluated by use of Eq. (8.43) to give

$$\begin{aligned}&\gamma(\theta_f, \phi_f, 0, 0) = \theta_f \quad \gamma(0, 0, \theta_i, \phi_i) = \theta_i \\ &\gamma(\theta_f, \phi_f, \pi, 0) = \pi + \theta_f \quad \gamma(\pi, 0, \theta_i, \phi_i) = \pi + \theta_i\end{aligned} \quad (8.46)$$

so that

$$\begin{aligned}\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle =& \cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) e^{i(\xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i))} \\ &+ \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i(\xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i))} \\ =& e^{i\alpha} \left[\cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i\beta}\right].\end{aligned} \quad (8.47)$$

where $\alpha = \xi(\theta_f, \phi_f, 0, 0) + \xi(0, 0, \theta_i, \phi_i)$ and $\beta = \xi(\theta_f, \phi_f, \pi, 0) + \xi(\pi, 0, \theta_i, \phi_i) - \alpha$. We can absorb $\exp(i\alpha)$ into the 'yet-to-be-determined' phase factor $\exp(i\xi)$, which amounts to putting $\alpha = 0$, so the final step is to determine β. This we do by noting that

$$\begin{aligned}\left|\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle\right|^2 =& \cos^2(\tfrac{1}{2}\gamma) \\ =& \tfrac{1}{2}(1 + \cos\gamma) \\ =& \tfrac{1}{2}(1 + \cos\theta_i \cos\theta_f + \cos(\phi_f - \phi_i) \sin\theta_i \sin\theta_f).\end{aligned} \quad (8.48)$$

But this is also given, from Eq. (8.47), by

$$\begin{aligned}\left|\langle S_f = \tfrac{1}{2}\hbar | S_i = \tfrac{1}{2}\hbar\rangle\right|^2 =& \left|\cos(\tfrac{1}{2}\theta_f) \cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f) \sin(\tfrac{1}{2}\theta_i) e^{i\beta}\right|^2 \\ =& \cos^2(\tfrac{1}{2}\theta_f) \cos^2(\tfrac{1}{2}\theta_i) + \sin^2(\tfrac{1}{2}\theta_f) \sin^2(\tfrac{1}{2}\theta_i) + \tfrac{1}{2} \sin\theta_i \sin\theta_f \cos\beta.\end{aligned} \quad (8.49)$$

Using standard trigonometric relations, this last expression can be rearranged to read

$$\left|\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle\right|^2 = \tfrac{1}{2}(1 + \cos\theta_i\cos\theta_f + \cos\beta\sin\theta_i\sin\theta_f) \tag{8.50}$$

which can be compared with Eq. (8.48) to give us $\beta = \pm(\phi_f - \phi_i)$. Choosing the negative sign (in order to have a result consistent with the earlier result Eq. (7.40) for $\theta_i = \theta_f = \pi/2$) gives the required result

$$\langle S_f = \tfrac{1}{2}\hbar|S_i = \tfrac{1}{2}\hbar\rangle = \cos(\tfrac{1}{2}\theta_f)\cos(\tfrac{1}{2}\theta_i) + \sin(\tfrac{1}{2}\theta_f)\sin(\tfrac{1}{2}\theta_i)e^{-i(\phi_f-\phi_i)}. \tag{8.51}$$

which can be readily shown to reduce to the earlier result Eq. (7.40) if we put $\theta_i = \theta_f = \pi/2$, this corresponding to the unit vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ lying in the XY plane.

We can make use of our general formula for the probability amplitudes, Eq. (8.51) to make explicit the probability amplitudes appearing in expression Eq. (7.19). Since here the states $|\pm\rangle$ correspond to the 'intermediate' states $|S_I = \pm\frac{1}{2}\hbar\rangle$ for $\hat{\mathbf{l}} = \hat{\mathbf{k}}$, i.e. in the z direction, so these are the states $|S_z = \pm\frac{1}{2}\hbar\rangle$. The state $|S_z = \frac{1}{2}\hbar\rangle = |+\rangle$, is associated with polar angles $\theta = 0, \phi = 0$ and $\theta = \pi, \phi = 0$ for the state $|S_z = -\frac{1}{2}\hbar\rangle = |-\rangle$. Finally, $|S\rangle \equiv |S_i = \frac{1}{2}\hbar\rangle$ with $\hat{\mathbf{n}}$ in the direction corresponding to the polar angles θ, ϕ. Overall, we have

$$\begin{aligned}
\langle +|S\rangle &= \cos(\tfrac{1}{2}\,0)\cos(\tfrac{1}{2}\theta) + e^{-i(0-\phi)}\sin(\tfrac{1}{2}\,0)\sin(\tfrac{1}{2}\theta)\\
&= \cos(\tfrac{1}{2}\theta)\\
\langle -|S\rangle &= \cos(\tfrac{1}{2}\pi)\cos(\tfrac{1}{2}\theta) + e^{-i(0-\phi)}\sin(\tfrac{1}{2}\pi)\sin(\tfrac{1}{2}\theta)\\
&= e^{i\phi}\sin(\tfrac{1}{2}\theta)
\end{aligned} \tag{8.52}$$

and hence

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle. \tag{8.53}$$

The associated probabilities are then

$$|\langle +|S\rangle|^2 = \cos^2(\tfrac{1}{2}\theta) \qquad |\langle -|S\rangle|^2 = \sin^2(\tfrac{1}{2}\theta). \tag{8.54}$$

Eq. (8.53) is a remarkably important formula used throughout quantum mechanics, finding application not only to explicitly spin half systems, but in fact to *any* system that possesses two basis states, otherwise known as a *qubit*.

8.3.3 Is every linear combination a state of the system?

One of the important properties of vectors is that two or more of them can be combined as a 'linear combination' to produce a third. If we are to consider quantum states as vectors, then this very basic property must also be possessed by quantum states, i.e. we ought to be able to show that by taking a linear combination of two basis states, the result is a vector that represents a possible state of the system. To see how this comes about, at least in the case of spin half, we make use of the general expression for the state $|S\rangle$ Eq. (8.53):

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle. \tag{8.55}$$

The atomic spin has been prepared in a state $|S\rangle$ where $S = \mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar$ by passing the atom through a Stern-Gerlach apparatus with its magnetic field in the $\hat{\mathbf{n}}$ direction, and the atom emerges in the $S = \frac{1}{2}\hbar$ beam. What this equation is then saying is that the combination $\cos(\frac{1}{2}\theta)|+\rangle + \exp(i\phi)\sin(\frac{1}{2}\theta)|-\rangle$, and $|S\rangle$, both represent the same thing – the atomic spin is in a state for which $S = \frac{1}{2}\hbar$. In other words, if we were presented with some arbitrary linear combination of the states $|\pm\rangle$, we ought to be able to work out exactly what state $|S\rangle$ this linear combination represents. Thus, for instance. consider the combination:

$$\frac{1}{\sqrt{2}}|+\rangle + i\frac{1}{\sqrt{2}}|-\rangle \tag{8.56}$$

we immediately see that $\cos(\frac{1}{2}\theta) = 1/\sqrt{2}$, $\exp(i\phi) = i$, and $\sin(\frac{1}{2}\theta) = 1/\sqrt{2}$, and hence $\theta = \phi = \pi/2$. Thus the magnetic field is pointing in the y direction, and hence the spin state of the atom is the state $|S\rangle = |S_y = \frac{1}{2}\hbar\rangle$.

What if we were presented with the combination $2|+\rangle + i2|-\rangle$? Here, we cannot find any angle θ, so it appears that this combination is not a possible state of the atomic spin. But we can write this as

$$2\sqrt{2}\left[\frac{1}{\sqrt{2}}|+\rangle + i\frac{1}{\sqrt{2}}|-\rangle\right] \tag{8.57}$$

which we can now understand as representing $2\sqrt{2}|S_y = \frac{1}{2}\hbar\rangle$. As we have argued above, we consider $2\sqrt{2}|S_y = \frac{1}{2}\hbar\rangle$ as describing the same physical state as $|S_y = \frac{1}{2}\hbar\rangle$. Thus, we can say that $2|+\rangle + i2|-\rangle$ is also a state of the system, namely $2\sqrt{2}|S_x = \frac{1}{2}\hbar\rangle$, which represents the same *physical* information about the state of the system as $|S_x = \frac{1}{2}\hbar\rangle$.

In the same way any combination $C_+|+\rangle + C_-|-\rangle$ where $C_\pm$ are complex numbers in general will always represent some state of the system, in general given by

$$\sqrt{|C_+|^2 + |C_-|^2}\,|S = \tfrac{1}{2}\hbar\rangle \tag{8.58}$$

where

$$S = \mathbf{S}\cdot\hat{\mathbf{m}} \tag{8.59}$$

and where $\hat{\mathbf{m}}$ is a unit vector in the direction defined by the polar angles

$$\theta = 2\tan^{-1}\left|C_-/C_+\right| \qquad \phi = \arg[C_-/C_+]. \tag{8.60}$$

Conversely, given any state of the system, we can work out how to write it in the form $C_+|+\rangle + C_-|-\rangle$. Further, we can repeat the whole of the above discussion for any other choice of the intermediate states $|S_I = \pm\frac{1}{2}\hbar\rangle$.

Ex 8.1 A spin half particle is prepared in the spin state

$$|S\rangle = 2|+\rangle + (1 + \sqrt{3}i)|-\rangle.$$

where $|S\rangle \equiv |\mathbf{S}\cdot\hat{\mathbf{n}} = \frac{1}{2}\hbar\rangle$. What is the component $\hat{\mathbf{n}}$ of spin of the particle which is known to be $\frac{1}{2}\hbar$?

We first need to check that this state is normalized to unity. The bra vector corresponding to this state is

$$\langle S| = 2\langle +| + (1 - \sqrt{3}i)\langle -|$$

so that the inner product $\langle S|S\rangle$ is given by

$$\begin{aligned}\langle S|S\rangle &= \Big(2\langle +| + (1 - \sqrt{3}i)\langle -|\Big)\Big(2|+\rangle + (1 + \sqrt{3}i)|-\rangle\Big)\\ &= 4\langle +|+\rangle + 2(1 + \sqrt{3}i)\langle +|-\rangle + 2(1 - \sqrt{3}i)\langle -|+\rangle\\ &\quad + (1 - \sqrt{3}i)(1 + \sqrt{3}i)\langle -|-\rangle.\end{aligned}$$

Using the fact that $\{|+\rangle, |-\rangle\}$ form an orthonormal basis for the state space for the spin of the particle, we have $\langle +|+\rangle = \langle -|-\rangle = 1$ and $\langle -|+\rangle = \langle +|-\rangle = 0$ so that

$$\langle S|S\rangle = 8.$$

This state is not normalized to unity, so in order to calculate probabilities correctly, we must renormalize $|S\rangle$. We do this by defining

$$|\widetilde{S}\rangle = \frac{|S\rangle}{\sqrt{8}} = \frac{1}{2\sqrt{2}}\Big(2|+\rangle + (1+\sqrt{3}i)|-\rangle\Big)$$

which now satisfies $\langle\widetilde{S}|\widetilde{S}\rangle = 1$. But note that the new state vector $|\widetilde{S}\rangle$ still represents the *same physical state* as the original vector $|S\rangle$.

We can now compare the coefficients of $|\widetilde{S}\rangle$ with the general expression for the state of a spin half particle, that is

$$|S\rangle = \cos(\tfrac{1}{2}\theta)|+\rangle + e^{i\phi}\sin(\tfrac{1}{2}\theta)|-\rangle$$

and immediately read off

$$\cos(\tfrac{1}{2}\theta) = \frac{1}{\sqrt{2}} \qquad \sin(\tfrac{1}{2}\theta)e^{i\phi} = \frac{1+\sqrt{3}i}{2\sqrt{2}}$$

from which we extract $\theta = \pi/2$ and $\phi = \pi/3$, the polar angle components of $\hat{\mathbf{n}}$.

It is this last fact that a state $|S\rangle$ can be written as the linear combination or linear superposition $C_+|+\rangle + C_-|-\rangle$ of two other states, analogous to Eq. (8.1) for the arbitrary vector $\mathbf{v}$, and conversely that any linear superposition of states is itself another state is the essential property that these states need to possess in order for them to be interpreted as vectors belonging to some vector vector space, specifically here, since the coefficients $C_\pm$ are in general complex, a complex vector space. This more general case is what is usually encountered in quantum mechanics, and below we will assume that the probability amplitudes are, in general, complex.

8.4 Constructing State Spaces

The approach adopted here starts with the sort of result that is obtained when the 'cancellation trick' is employed. In the spin half case, what is obtained is the expression

$$|S\rangle = |+\rangle\langle+|S\rangle + |-\rangle\langle-|S\rangle$$

which is interpreted as a vector equation, with the arbitrary state $|S\rangle$ expressed as a linear combination of the pair of orthonormal basis states $|\pm\rangle$. This brings us to the general idea that lies behind setting up a quantum mechanical description of a physical system, and that is to identify a set of orthonormal *basis* states for the system. These basis states have to satisfy a number of properties that we can extract from our study of the two slit and spin half cases.

In the case of two slit interference, the states $|1\rangle$ and $|2\rangle$ are states associated with the two possible positions of the particle: at the position of one or the other of the slits cut into a screen. We have also seen that the electron could be observed to be at one slit *or* the other – the two possibilities are mutually exclusive. Finally, these two possibilities cover *all* the possibilities, at least within the constraints of the two slit model. So we take these states as our basis states, and construct an arbitrary state of our system by taking linear combinations of these two states.

In the case of particle spin, the basis states are, for instance, $|\pm\rangle$, or written out more fully, $|S_z = \pm\frac{1}{2}\hbar\rangle$, corresponding to the two possible values of the z component of the spin of the particle. These two possibilities are mutually exclusive – the atom emerges in either the $S_z = \frac{1}{2}\hbar$ *or* the $S_z = -\frac{1}{2}\hbar$ beam, and complete in the sense that all possibilities are covered – the atom never emerges in any other beam.

In each case, note was taken of the possible values of some measurable property of the system: the position of the particle at any one of the slits, or S_z, the z component of spin. The collection of possibilities was exhaustive — all possibilities were included — and mutually exclusive — the system could be observed to be in one state or the other. In fact, we label these states by the possible values that could be measured, e.g. spin states $|S_z = \pm\frac{1}{2}\hbar\rangle$. In other words, we identify a measurable property of the system, otherwise known as an *observable*, determine all the possible values that this observable can have, either by experiment or by theoretical argument, and for each possible value of this observable, we associate a state labelled by the observable's value. This collection of states so determined we take as our basis states for the system. From these two examples, we can propose a number of properties that our basis states would have to have in the general case:

1. Each basis state represents a mutually exclusive possibility, that is, if the system is observed to be in one of the basis states, it is definitely the case that it will not be observed in any of the others
2. The basis states are associated with the possible values of some measurable property of the system.
3. The basis states must be complete in that they cover all possible such values.

Note that the basis states are not unique. This is most readily seen in the case of spin half. We could have taken as our basis states *any* pair of states $|S = \pm\frac{1}{2}\hbar\rangle$ where $S = \mathbf{S}\cdot\hat{\mathbf{n}}$ with $\hat{\mathbf{n}}$ an arbitrary unit vector.

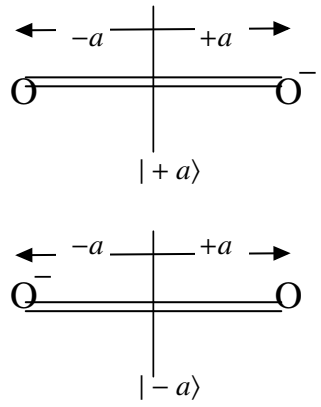

Figure 8.4: O_2^- ion with the two possible positions of the electron, corresponding to the two states $|\pm a\rangle$.

So let us see how these ideas can be applied to a completely different system. This system we will take to be an O_2^- as illustrated in the adjacent figure. Is there any natural choice for what observable we might look upon as determining the basis states of this system? One possibility that suggests itself is the position of the extra electron. For a real O_2^- ion the electron could be found in myriad different possible places, but for our purposes we will assume that there are only two possibilities: at the position of either of the oxygen atoms, that is at $x = \pm a$. The two possibilities correspond to two possible states for the ion, $|\pm a\rangle$. These states we recognize as having all the properties that we have listed above for basis states, i.e. the electron is observed to be in *either* state $|+a\rangle$ *or* state $|-a\rangle$, and that there is nowhere else for the electron to be found (within the limits of our model). By analogy with the spin half example, we then say that

$$\langle a|a\rangle = \langle -a|-a\rangle = 1, \quad \langle -a|a\rangle = \langle a|-a\rangle = 0 \tag{8.61}$$

i.e. the states $|\pm a\rangle$ form a pair of orthonormal basis states for the state space of the ion. This state space has dimension 2. An arbitrary state of the ion will be given by

$$|\psi\rangle = |+a\rangle\langle +a|\psi\rangle + |-a\rangle\langle -a|\psi\rangle \tag{8.62}$$

As another example we can consider a (fictitious) two level atom, an atom which, if its energy is measured, is found to have only two values, E_1 or E_2, with $E_1 > E_2$. Such an atom is, of course, an idealization, but one that has proved to extremely valuable one in understanding the details of the interaction of quasimonochromatic light fields, such as that produced by a laser, with a real atom.

The nature of the interaction is such that the atom behaves as if it has only two energy levels, so that the simplification being considered here is taken as the basis of often used theoretical models.

Given that the energy can only have two values, and moreover that the energy is measured to be *either* E_1 *or* E_2 in a way totally analogous to measuring a component of the spin of a spin half system, we can then assign to the atom two possible states, call them $|E_1\rangle$ and $|E_2\rangle$, or $|e\rangle$ and $|g\rangle$, where $e \equiv$ excited state and $g \equiv$ ground state. We then have

$$\begin{aligned} \langle e|e\rangle &= \langle g|g\rangle = 1 \\ \langle e|g\rangle &= \langle g|e\rangle = 0 \end{aligned}. \tag{8.63}$$

These states then act as the orthonormal basis states of our two level atom, so that any state of the two level atom can be written as a linear combination

$$|\psi\rangle = a|e\rangle + b|g\rangle. \tag{8.64}$$

8.4.1 A General Formulation

We will now see how these ideas can be applied to more general kinds of physical systems. To begin, we have to set up these basis states for the given system. So suppose we perform an *exhaustive* series of measurements of some observable property of the system – call it Q. For example, we could determine through which slits it is possible for an electron to pass in the two slit experiment above in which case $Q \equiv$ 'the positions of the two slits', or the possible values of the Z component of the spin of a particle, in which case $Q \equiv S_z$, or the possible values of the position of an electron on an O_2^- ion, as discussed above, in which case $Q \equiv$ 'the position of the electron on the O_2^- ion'. The generic name given to Q is that it is an 'observable'. We will give many more examples of observables later in this Chapter, and look at the concept again in a later Chapter.

Whatever the system, what we mean by exhaustive is that we determine *all* the possible values that the observed quantity Q might have. For instance, we determine that, in the two slit interference experiment, the electron can pass through, (that is, be observed to be at) the position of one or the other of the two slits, or that a spin half particle would be observed to have either of the values $S_z = \pm\frac{1}{2}\hbar$, and no other value. Or for the O_2^- ion, the electron can be found *either* on the atom at position $x = -a$ *or* on the atom at position $x = +a$. In general, for an arbitrary observable Q, let us represent these observed values by $q_1, q_2, q_3, \ldots, q_N$, i.e. N in total, all of which will be real numbers. Of course, there might be other observable properties of the system that we might be able to measure, but for the present we will suppose that we only need concern ourselves with just one. In keeping with the way that we developed the idea of state earlier, we then let $|q_1\rangle$ represent the state for which Q definitely has the value q_1, and similarly for all the others possibilities.

We now turn to Eq. (7.48), the result obtained earlier when considering generalizations of the two slit experiment to the multiple slit case, or the generalization of the Stern-Gerlach experiment to the arbitrary spin case. There, a sum was made over probability amplitudes for the different 'pathways' from a given initial state $|\psi\rangle$ to some final state $|\phi\rangle$ via a set of intermediate slit or spin states. Here, we generalize that result for an arbitrary set of intermediate states $\{|q_1\rangle, |q_2\rangle, \ldots\}$ as defined above, and make the following claim, the fundamental rule of quantum mechanics, that if the system is initially prepared in the state $|\psi\rangle$, then the probability amplitude of finding it in the state $|\phi\rangle$ is given by

$$\langle\phi|\psi\rangle = \sum_{n=1}^{N} \langle\phi|q_n\rangle\langle q_n|\psi\rangle \tag{8.65}$$

which tells us that the total probability amplitude of finding the system in the final state $|\phi\rangle$ is just the sum of the probability amplitudes of the system 'passing through' any of the states $\{|q_n\rangle; n =$

$1, 2, \ldots, N\}$. The expression Eq. (8.65) is often referred to as a *closure relation* though confusingly, it is also sometimes referred to as a completeness relation, a term which we apply to another expression below.

We further claim, from the basic meaning of probability amplitude, that for any state $|\psi\rangle$ of the system, we must have

$$\langle\psi|\psi\rangle = 1 \tag{8.66}$$

known as the normalization condition. This follows from the requirement that $|\langle\psi|\psi\rangle|^2 = 1$, i.e. that if the system is in the state $|\psi\rangle$, then it is definitely (i.e. with probability one) in the state $|\psi\rangle$.

We then recognize that the probability amplitudes $\langle q_m|q_n\rangle$ must have the following properties:

- $\langle q_m|q_n\rangle = 0$ if $m \neq n$. This amounts to stating that if the system is in the state $|q_n\rangle$, i.e. wherein the observable Q is known to have the value q_n, then there is zero possibility of finding it in the state $|q_m\rangle$. Thus the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are *mutually exclusive*.
- $\langle q_n|q_n\rangle = 1$. This asserts that if the system is in the state for which the quantity Q has the value q_n, then it is certain to be found in the state in which it has the value q_n.
- The states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are also exhaustive in that they cover *all* the possible values that could be observed of the observable Q. These states are said to be *complete* — simply because they cover all possibilities.

These three properties are analogous to the properties that we associate with the inner products of members of an orthonormal set of basis vectors for a complex inner product space, which suggests we interpret the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ as an orthonormal set of basis vectors, or basis states, for our system. We then interpret that fundamental law, Eq. (8.65) as an expression for the inner product of two state vectors, $|\psi\rangle$ and $|\phi\rangle$ in terms of the components of these vectors with respect to the basis states $|q_n\rangle$. Expressions for these state vectors in terms of the basis states can then be obtained by what we have referred to earlier as the 'cancellation trick' to give

$$\begin{aligned} |\psi\rangle &= \sum_{n=1}^{N} |q_n\rangle\langle q_n|\psi\rangle \\ \langle\phi| &= \sum_{n=1}^{N} \langle\phi|q_n\rangle\langle q_n| \end{aligned} \quad . \tag{8.67}$$

We have argued above that the states $\{|q_n\rangle; n = 1, 2, \ldots, N\}$ are, in a sense, complete. We can use this to argue that *any* state of the system can be expressed in the form Eq. (8.67). To see how this follows, suppose $|\psi\rangle$ is some arbitrary state of the system. Then, for at least one of the states $|q_n\rangle$, we must have $\langle q_n|\psi\rangle \neq 0$, i.e. we must have a non-zero probability of observing the system in *one* of the states $|q_n\rangle$. If this were not the case, then for the system in a state such as $|\psi\rangle$ it would be saying that if we measure Q, we don't get an answer! This does not make physical sense. So physical consistency means that it must be the case that any state of the system can be written as in Eq. (8.67) and for that reason, these expressions are referred to as *completeness relations*.

We can also make the inverse claim that *any* such linear combination represents a possible state of the system. The justification for this is not as clear cut as there are physical systems for which there are limitations on allowed linear combinations (so-called super-selection rules), but it appears to be a rule that holds true unless there are good physical reasons, on a case-by-case basis, why it should not. We will assume it to be true here.

If we choose $|\phi\rangle = |\psi\rangle$ in Eq. (8.65) and we use the normalization condition we find that

$$\langle\psi|\psi\rangle = \sum_{n=1}^{N}\langle\psi|q_n\rangle\langle q_n|\psi\rangle = 1. \tag{8.68}$$

But we must also have, since the probability of finding the system in any of the states $|q_n\rangle$ must add up to unity that

$$\sum_n |\langle q_n|\psi\rangle|^2 = 1 \tag{8.69}$$

This can also be understood as being a consequence of our interpretation of the states $\{|q_n\rangle; n = 1, 2, \ldots N\}$ as a complete set of mutually exclusive possibilities, complete in the sense that the total probability of ending up in any of the mutually exclusive possible final states $|q_n\rangle$ adds up to unity — there is nowhere else for the system to be found. By subtracting the last two expressions we arrive at

$$\sum_{n=1}^{N}\left(\langle\psi|q_n\rangle - \langle q_n|\psi\rangle^*\right)\langle q_n|\psi\rangle = 0. \tag{8.70}$$

A sufficient condition for this result to hold true is

$$\langle\psi|q_n\rangle = \langle q_n|\psi\rangle^* \tag{8.71}$$

and hence that, in general,

$$\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*. \tag{8.72}$$

Thus, by a simple extension of the arguments presented in Section 8.3 in the case of spin half quantum states it can be seen that the general results above are also completely analogous to the properties of vectors in a complex vector space. This mathematical formalism will be discussed more fully in the next Chapter, but for the present we can summarize the essential ideas based on what we have already put forward earlier. The important points then are as follows:

1. The collection of all the possible state vectors of a quantum system forms a complex vector space known as the *state space* of the system.
2. The probability amplitudes are identified as the inner product of these state vectors.
3. The intermediate states $\{|q_n\rangle; n = 1, 2, \ldots\}$ form a complete orthonormal set of basis states of this state space, i.e. any state vector $|\psi\rangle$ can be written as a linear combination of these basis states.
4. The number of basis states is known as the dimension of the state space.

8.4.2 Further Examples of State Spaces

The ideas developed above can now be applied to constructing a state space for a physical system. The basic idea is as discussed in Section 7.4 which enables us to define a set of basis states for the state space of the system. By establishing a set of basis states, in a sense, we 'bring the state space into existence', and once this is done, we are free to use all the mathematical machinery available for analysing the properties of the state space so constructed. The question can be asked as to whether or not the ideas presented in Section 7.4, admittedly extracted from only a handful of examples, can be applied with success to any other system. This is a question that can only be answered by applying the rules formulated there and considering the consequences. In Section 8.5 we will discuss where these ideas, if naively applied, fail to work. Otherwise, these ideas, when fully formed, constitute the basis of quantum physics.

In accordance with the ideas developed in Section 7.4, constructing a state space for a physical system can be carried out by recognizing the intermediate states through which a system can pass as it makes its way from some initial state to some observed final state, as was done in the case of the two slit, or spin half systems. Thus, in the two slit example, the two possible intermediate states are those for which the particle is to be found at the position of either of the two slits. In the spin half example, the two intermediate states are those in which the spin is observed to have either of the two values $S_z = \pm\frac{1}{2}\hbar$; these are the states we have been calling $|\pm\rangle$. These intermediate states are states of the system that can be identified through an argument based on the idea that some physical property of the system can be exhaustively measured to yield a set of values that we then use to label a complete set of basis states for the state space of the system.

Negatively Charged Ions Here the system is a molecule which has acquired an extra electron, which can be assumed to found only on any one of the atoms making up the molecule. This is, of course, an approximation. The electron could be found anywhere around the atoms, or in the space between the atoms, in a way that depends on the nature of the chemical bond between the atoms. Here we are making use of a coarse notion of position, i.e. we are assuming that the electron can be observed to reside on one atom or the other, and we do not really care about exactly where on each atom the electron might be found. The idea is best illustrated by the simple example of the O_2^- ion in which the electron can be found on one or the other of the oxygen atoms (see Fig. (8.4)) as discussed on p98.

This kind of model can be generalized to situations involving different geometries, such as atoms arranged in a ring e.g. an ozone ion O_3^-. In this case, the state space will be spanned by three basis states corresponding to the three possible positions at which the electron can be observed. This model (and its generalizations to an arbitrary number of atoms arranged in a ring) is valuable as it gives rise to results that serve as an approximate treatment of angular momentum in quantum mechanics.

Spin Flipping In this case, we have a spin half particle (for instance) in a constant magnetic field, so the two possible states are the familiar spin up or spin down states. If, in addition, we add a *rotating* magnetic field at right angles to the constant field, there arises a time dependent probability of the spin flipping from one orientation to the other. As the spin up and spin down states are of different energies, this represents a change in energy of the particle, a change that can be detected, and is the basis of the electron spin and nuclear magnetic resonance imaging much used in medical work. Obviously this is a state space of dimension two.

Ammonia molecule Here the system is the ammonia molecule NH_3 in which the nitrogen atom is at the apex of a triangular pyramid with the three hydrogen atoms forming an equilateral triangle as the base. The nitrogen atom can be positioned either above or below the plane of the hydrogen atoms, these two possibilities we take as two possible states of the ammonia molecule. (The N atom can move between these two positions by 'quantum tunnelling' through the potential barrier lying in the plane of the hydrogen atoms.) Once again, this is a state space of dimension 2.

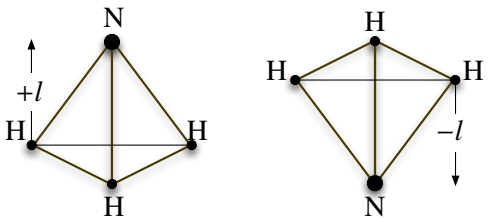

Figure 8.5: Ammonia molecule in two states distinguished by the position of the nitrogen atom, either above or below the plane of the hydrogen atoms, corresponding to the states $|+l\rangle$ and $|-l\rangle$ respectively.

Benzene Molecule An example of quite a different character is that of the benzene molecule, illustrated in Fig. 8.6. The two states of the molecule are distinguished by the positioning of the double bonds between pairs of carbon atoms. The molecule, at least with regard to the arrangements of double bonds can be found in two different states which, for want of a better name, we will call $|\alpha\rangle$ and $|\beta\rangle$. The state space is therefore of dimension 2, and an arbitrary state of the molecule would be given by

$$|\psi\rangle = a|\alpha\rangle + b|\beta\rangle. \tag{8.73}$$

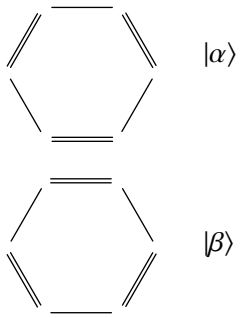

Figure 8.6: Two arrangements of the double bonds in a benzene molecule corresponding to two states $|\alpha\rangle$ and $|\beta\rangle$.

In all the cases considered above, the states were labelled by one piece of data only. It is possible, under certain circumstances, to generalise this to situations where two or more labels are needed to specify the basis states.

8.4.3 States with multiple labels

We have been solely concerned with states with a single label, Recall the general definition of a ket:

$$\Big|\ \text{All the data concerning the system that can be known without mutual interference or contradiction.}\ \Big\rangle$$

As we have seen, for simple systems, the state can be labelled with a single piece of information, e.g. $|S_z = \frac{1}{2}\hbar\rangle$ for a spin half particle for which $S_z = \frac{1}{2}\hbar$, $|x\rangle$ for a particle positioned at the point x, $|n\rangle$ for a single mode cavity with n photons. But, for more complex systems, it is reasonable to expect that more information is needed to fully specify the state of the system — this is certainly the case for classical systems. But, perhaps not unexpectedly, this is not a straightforward procedure for quantum systems. The important difference is embodied in the condition 'without mutual interference': some observables interfere with others in that the measurement of one 'scrambles' the predetermined value of the other, as was seen in the case of the components of spin, or as seen when trying to measure the position and momentum of a particle. If there is no such interference between two observables, then they are said to be *compatible*, and two (or more) labels can be legitimately used to specify the state e.g. $|S_x = \frac{1}{2}\hbar, x\rangle$ for a spin half particle with x component of spin $S_x = \frac{1}{2}\hbar$ AND at the position x. Thus, in this example, we are assuming that the spin of a particle and its position are compatible.

As another example, we can consider a system made up of more than one particle e.g. a system of two spin half particles. The state could then be written as

$$\big|\text{data for first particle,data for second particle}\big\rangle.$$

Its possible states would then be:

$$|S_z = \tfrac{1}{2}\hbar, S_z = \tfrac{1}{2}\hbar\rangle,\ \ |S_z = \tfrac{1}{2}\hbar, S_z = -\tfrac{1}{2}\hbar\rangle,\ \ |S_z = -\tfrac{1}{2}\hbar, S_z = \tfrac{1}{2}\hbar\rangle,\ \ |S_z = -\tfrac{1}{2}\hbar, S_z = -\tfrac{1}{2}\hbar\rangle. \tag{8.74}$$

We can write this more simply as $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ by making the identifications $S_z = \frac{1}{2}\hbar \to 0$ and $S_z = -\frac{1}{2}\hbar \to 1$. The states are orthonormal:

$$\begin{aligned} \langle 00|00\rangle &= 1\langle 01|01\rangle = 1 \ \text{ etc} \\ \langle 00|01\rangle &= 0\langle 01|10\rangle = 0 \ \text{ etc} \end{aligned} \tag{8.75}$$

An arbitrary state is $|\psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$. These states form a set of orthonormal basis states for a state space of dimension 4. The idea can be extended to many particles, or other more complex systems e.g. whole atoms, or molecules or solid state systems and so on.

The notion of 'compatible observables' and determining whether or not two (or more) observables are compatible, is studied later in a more rigorous way once the idea of an observable being represented by Hermitean operators has been formulated.

Qubits Note that a single spin half has the basis states $|0\rangle$ and $|1\rangle$ in our new notation. These two states can be looked on as corresponding to the two binary numbers 0 and 1. A linear combination

$$|\psi\rangle = c_0|0\rangle + c_1|1\rangle.$$

can be formed which represents the possibility of a memory registering a bit of information simultaneously as both a 0 and a 1. This is in contrast with a classical bit, which can be registered as *either* a 0 *or* a 1. The spin half system is an example of a *qubit*. But spin half is not in any way special: *any* two state system: a spin half particle, a two level atom, the two configurations of a benzene molecule, a quantum dot either occupied or not occupied by an electron .... Quantum computation then involves manipulating the whole state $|\psi\rangle$, which, in effect, amounts to performing two calculations at once, differing by the initial setting of the memory bit.

This idea introduced can be readily extended. Thus, if we have two spin half particles, we have the possible states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. The data labelling the state $|00\rangle$ represents the number zero, $|01\rangle$ the number one, $|10\rangle$ the number two, and $|11\rangle$ the number three. We now have two qubits, and a state space of dimension four, and we can set up linear combinations such as

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle \tag{8.76}$$

and we can then perform calculations making use, simultaneously, of four different possible values for whatever quantity the states are intended to represent. With three atoms, or four and so on, the state space becomes much larger: of dimension 2^N in fact where N is the number of qubits, and the basis states represent the numbers ranging from 0 to $2^N - 1$ in binary notation.

The viability of this scheme relies on the linear combination not being destroyed by decoherence. Decoherence is the consequence of noise destroying a linear combination of quantum states, turning a quantum state into a mixture of alternate classical possibilities.

8.5 States of Macroscopic Systems — the role of decoherence

In the examples given above, it was assumed that an exhaustive list of results that could be obtained in the measurement of some observable of a quantum system could be used to set up the basis states for the state space of the system. The value of doing this is, of course, to be determine by the success or otherwise of these ideas. That quantum mechanics is such an overwhelmingly successful theory indicates that there is something correct in this procedure, but the question that arises is this: why does it not appear to work for macroscopic systems, i.e. for systems which we know can be fully adequately explained by standard classical physics? The answer appears to lie in the fact that in all the examples discussed above, whether or not the Hilbert space is of finite or infinite dimension, i.e. whether or not we are talking about spin up or spin down of a spin half particle, or the position of a particle in space, the implicit assumption is that the system we are considering is totally isolated from all other systems, in particular from any influence of the surrounding environment. After all, when we talked about a system, such as an O_2^- ion, we are ignoring all the other physical influences that could act on this system, i.e. we do not need to mention, in our specification of the state of the system, anything other than properties that directly pertain to the system of interest. The assumption is made, as it is in classical physics, that such

influences are sufficiently weak that they can be ignored to a good approximation. In effect, we are supposing that the systems under consideration are isolated systems, that is, systems that are isolated from the effect of any external perturbations.

Classically, at the macroscopic level, we can usually continue to ignore weak perturbing influences when specifying the state of a system. In fact, when defining a 'system' we typically include in what we refer to as the 'system', all the participants in the physical process being described that interact strongly with each other. Anything else that weakly affects these constituents is ignored. For instance, when describing the orbital dynamics of the Earth as it revolves around the Sun, we might need to take into account the gravitational pull of the Moon – the system is the Earth, the Sun and the Moon. But we do not really need to take into account the effect of the background microwave radiation left over after the Big Bang. Or, when describing the collision between two billiard balls, it is probably necessary to include the effect of rolling friction, but it not really necessary to take into account the frictional drag due to air resistance. Of course, sometimes it is necessary to include external influences even when weak: to describe a system coming to thermal equilibrium with its surroundings it is necessary to extend the system by including the environment in the dynamical model. In any of these examples, the same classical physics methods and philosophy applies.

There is a subtle difference when it comes to trying to apply the quantum ideas developed so far to macroscopic systems. The same, weak perturbations that can be put to one side in a classical description of a macroscopic system turn out to have a far-reaching effect if included in a quantum description of the same system. If we were to attempt to describe a macroscopic system according to the laws of quantum mechanics, we would find that any linear superposition of different possible states of the system evolves on a fantastically short time scale to a classical mixture of the different possibilities. For instance, if we were to attempt to describe the state of a set of car keys in terms of two possibilities: in your pocket $|p\rangle$ or in your brief case $|b\rangle$, then a state of the form

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|p\rangle + |b\rangle) \tag{8.77}$$

could be used to represent a possible 'quantum state' of the keys. But this quantum state would be exceedingly short lived (on a time scale $\sim 10^{-40}$ sec), and would evolve into the two alternative possibilities: a 50% chance of the keys being in the state $|p\rangle$, i.e. a 50% chance of finding your keys in your pocket, and a 50% chance of being in the state $|b\rangle$, i.e. a 50% chance of finding them in your brief case. But this is no longer a superposition of these two states. Instead, the keys are *either* in the state $|p\rangle$ *or* the state $|b\rangle$. What this effectively means is that randomness is still there, i.e. repeating an experiment under identical conditions can give randomly varying results. But the state of the keys is no longer represented by a state vector, so there are no longer any quantum interference effects present. The randomness can then be looked upon as being totally classical in nature, i.e. as being due to our ignorance of information that is in principle there, but impossibly difficult to access. In effect, the quantum system behaves like a noisy classical system.

The process that washes out the purely quantum effects is known as decoherence. Since it is effectively impossible to isolate any macroscopic system from the influence of its surrounding environment[1], all macroscopic systems are subject to decoherence. This process is believed to play a crucial role in why, at the macroscopic level, physical systems, which are *all* intrinsically quantum mechanical, behave in accordance with the classical laws of physics. It is also one of the main corrupting influences that prevent a quantum computer from functioning as it should. Quantum computers rely for their functioning on the 'qubits' remaining in linear superpositions of states, but the ever-present decohering effects of the environment will tend to destroy these delicate quantum states before a computation is completed, or else at the very least introduce errors as the

[1] 'No man is an Iland, intire of itselfe' – J. Donne, *Devotions upon Emergent Occasions* Meditation XVII (1693)

computation proceeds. Controlling decoherence is therefore one of the major challenges in the development of viable quantum computers.

So, the bottom line is that it is only for protected isolated systems that quantum effects are most readily observed, and it is for microscopic systems that this state of affairs is to be found. But that is not to say that quantum effects are not present at the macroscopic level. Peculiar quantum effects associated with the superposition of states are not to be seen, but the properties of matter in general, and indeed the properties of the forces of nature, are all intrinsically quantum in origin.