1 Mathematical preliminaries

The mathematical language of quantum mechanics is that of vector spaces and linear algebra. In this preliminary section, we will collect the various definitions and mathematical results that we will use below.

1.1 Hilbert space

Complex vector spaces

We start with the idea of a *complex vector space*. This is a space of objects that we call *vectors*, such that any complex number times a vector gives another vector and any sum of vectors gives another vector. In quantum mechanics, we will use the notation $|v\rangle$ (called a "ket") to denote a vector. All the properties familiar from real vector spaces still hold in the complex case.[1]

Inner product

For the vector spaces in quantum mechanics, we have an additional structure called an *inner product* which generalizes the dot product for real vector spaces. The inner product is a map that takes a pair of vectors to a complex number that we denote by

$$(|v_1\rangle, |v_2\rangle) \to \langle v_1|v_2\rangle \tag{1}$$

The inner product obeys the following properties:

- $\langle v_2|v_1\rangle = \langle v_1|v_2\rangle^*$
- If $|w\rangle = z_1|v_1\rangle + z_2|v_2\rangle$ then $\langle v_3|w\rangle = z_1\langle v_3|v_1\rangle + z_2\langle v_3|v_2\rangle$
- If $|w\rangle = z_1|v_1\rangle + z_2|v_2\rangle$ then $\langle w|v_3\rangle = z_1^*\langle v_1|v_3\rangle + z_2^*\langle v_2|v_3\rangle$
- $\langle v|v\rangle \geq 0$ with equality only for $|v\rangle = \mathbf{0}$.

Here, the third property follows from the first two. The last property allows us to associated a real "length" to a vector and also gives us a measure of distance between vectors, defined as the length of the difference vector.

[1] Specifically, we have $(|v_1\rangle + |v_2\rangle) + |v_3\rangle = |v_1\rangle + (|v_2\rangle + |v_3\rangle)$, $|v_1\rangle + |v_2\rangle = |v_2\rangle + |v_1\rangle$, there exists $\mathbf{0}$ such that $|v\rangle + \mathbf{0} = |v\rangle$ for all $|v\rangle$, $-|v\rangle + |v\rangle = \mathbf{0}$, $z_1(z_2|v\rangle) = (z_1 z_2)|v\rangle$, $1|v\rangle = |v\rangle$, $z(|v_1\rangle + |v_2\rangle) = z|v_1\rangle + z|v_2\rangle$, $(z_1 + z_2)|v\rangle = z_1|v\rangle + z_2|v\rangle$.

Dual vectors

We can also give meaning to the object $\langle v|$, (known as a "bra"). It represents a linear map from the vector space to complex numbers, known mathematically as a *dual vector*. The map is defined by

$$|w\rangle \to \langle v|w\rangle \ . \tag{2}$$

Hilbert space

Finally we can define a complex *Hilbert space*: it is a complex vector space with inner product satisfying one more technical criterion known as *completeness*. This is the statement that any sequence of vectors whose elements become arbitrarily close to one another as the sequence progresses (a *Cauchy sequence*) must converge to a vector in the space; this is automatically satisfied for finite dimensional examples.

1.2 Orthonormal bases, finite and infinite dimensional spaces

It is often convenient (and will be physically relevant) to define a basis for a vector space. We define an *orthonormal basis* to be a set of vectors $|e_n\rangle$ which span the space (i.e. any vector can be expressed as a linear combination of them) and such that

$$\langle e_m|e_n\rangle = \delta_{mn} \ . \tag{3}$$

Here δ_{mn} vanishes for m not equal to n and is 1 otherwise. Given any vector $|v\rangle$, we can then write

$$|v\rangle = \sum_n c_n|e_n\rangle \tag{4}$$

where

$$c_n = \langle e_n|v\rangle \ . \tag{5}$$

We say that the set of coefficients $(c_1, c_2, c_3, \dots)$. represent the vector $|v\rangle$ in the basis $|e_n\rangle$. There are an infinite number of possible bases, and the set of coefficients representing a given vector will generally be different for each choice. Thus, it is important to keep in mind that the list $(c_1, c_2, c_3, \dots)$ has no intrinsic meaning unless we specify which basis we are talking about.

Dimension of Hilbert space

The number of basis elements can be finite or infinite. When the number is finite, every orthonormal basis for the vector space will have the same number of vectors. We call this the *dimension* of the Hilbert space.

The infinite dimensional case

In the infinite dimensional case, we will typically be able to define a set of basis vectors $|e_n\rangle$ that is "countable," i.e. labeled by positive integers. However, it is sometimes convenient to define basis vectors labeled by real numbers. In this case, the statement of orthonormality becomes

$$\langle x|y\rangle = \delta(x-y) \tag{6}$$

where the Dirac delta function $\delta(x)$ (mathematically not actually a function but a *distribution*) is defined to vanish for $x \neq 0$ and satisfy[2]

$$\int dx\delta(x) = 1 \ . \tag{7}$$

This implies that for any function f,

$$\int dx\delta(x-a)f(x) = f(a) \ . \tag{8}$$

In this continuous case, we can represent a general vector as

$$|v\rangle = \int dx v(x)|x\rangle \tag{9}$$

where the function $v(x)$ takes the place of the coefficients c_i as the basis-dependent representation of the vector.

1.3 Operators

Given a Hilbert space, a *linear operator* or simply *operator* is defined as a map

$$|v\rangle \to \hat{\mathcal{O}}|v\rangle \tag{10}$$

satisfying

$$\hat{\mathcal{O}}(z_1|v_1\rangle + z_2|v_2\rangle) = z_1\hat{\mathcal{O}}|v_1\rangle + z_2\hat{\mathcal{O}}|v_2\rangle \ . \tag{11}$$

The set of operators has the mathematical structure of an *algebra*. This means that they form a complex vector space (i.e. we can add operators and multiply them by a constant), with the additional structure of being able to multiply operators. This multiplication is defined by

$$(\hat{\mathcal{O}}_1\hat{\mathcal{O}}_2)|v\rangle = \hat{\mathcal{O}}_1(\hat{\mathcal{O}}_2|v\rangle) \ . \tag{12}$$

The algebra of operators includes an "identity" element, the operator $\mathbf{1}$ that takes every vector to itself. This makes it a *unital algebra*.

[2]Here, x could be a single coordinate on the real line, could also represent coordinates in a higher-dimensional space, or on some other geometry.

Matrix representation of an operator

By the linearity property, the action of an operator is completely specified by its action on any basis of vectors: if $|v\rangle = \sum c_n |e_n\rangle$ then $\hat{\mathcal{O}}|v\rangle = \sum c_n \hat{\mathcal{O}}|e_n\rangle$.

Acting on any basis vector, the operator must give some linear combination of basis vectors. We can express this as

$$\hat{\mathcal{O}}|e_n\rangle \equiv \sum_m \mathcal{O}_{mn}|e_m\rangle \ , \tag{13}$$

where

$$\mathcal{O}_{mn} \equiv \langle e_m|\mathcal{O}|e_n\rangle \ . \tag{14}$$

are called the matrix elements of the operator $\hat{\mathcal{O}}$.

The matrix $\mathcal{O}_{mn}$ gives a representation of the operator $\hat{\mathcal{O}}$ in a particular basis. If a vector $|v\rangle$ is represented in the same basis by coefficients c_n, the vector $\mathcal{O}|v\rangle$ is represented by coefficients c'_m given by[3]

$$c'_m = \sum_n \mathcal{O}_{mn} c_n \ ; \tag{15}$$

That is, we write the coefficients c_n as a column vector and multiply by the matrix O_{mn} to get the new coefficients.

It often helps to think of these linear operators as performing some type of geometrical transformation on the vector space. We can have various qualitatively different types including operators that stretch things in various ways, operators that preserve lengths similar to rotations, or operators that project onto some lower dimensional subspace.

Eigenvectors and eigenvalues

The action of an operator may be particularly simple on certain vectors. We say that $v\rangle$ is an *eigenvector* of $\hat{\mathcal{O}}$ with *eigenvalue* λ if

$$\hat{\mathcal{O}}|v\rangle = \lambda|v\rangle \ . \tag{16}$$

Sometimes there can be different eigenvectors with the same eigenvalue. In this case, any linear combination of these eigenvectors is an eigenvector with this eigenvalue.

For some operators, it is possible to choose a basis of vectors which are all eigenvectors of $\hat{\mathcal{O}}$. In this case, $\hat{\mathcal{O}}$ is represented simply as a diagonal matrix.

[3] To see this, we note that $c'_m = \langle e_m|\mathcal{O}|v\rangle = \langle e_m|\mathcal{O}(\sum_n c_n|e_n\rangle) = \sum_n \mathcal{O}_{mn}c_n$.

Hermitian operators

A particular example of this type with special physical relevance in quantum mechanics is the Hermitian operator. For these operators, it is possible to choose the basis of eigenvectors to be orthonormal, and all of the eigenvalues are real. We can visualize the action of a Hermitian operator as stretching the space along each eigenvector direction by a multiplicative factor given by the eigenvalue.

For a Hermitian operator, if we denote the eigenvalues by λ_n and the eigenvectors by $|\lambda_n\rangle$, the matrix elements in this basis are simply

$$\mathcal{O}_{mn} = \lambda_m \delta_{mn} \ , \tag{17}$$

i.e. a diagonal matrix with the eigenvalues along the diagonal.

In another basis, a Hermitian operator will generally not be diagonal. But we can recognize it as Hermitian since in an orthonormal basis, the *adjoint* of the corresponding matrix (defined as the complex conjugate of the transpose) must be equal to the matrix itself:

$$\mathcal{O}_{nm} = \mathcal{O}^*_{mn} \ . \tag{18}$$

We can express this property in a basis independent language: for an operator $\hat{\mathcal{O}}$, we define the adjoint operator $\hat{\mathcal{O}}^\dagger$ to be the operator satisfying

$$\langle \chi | \hat{\mathcal{O}}^\dagger | \psi \rangle = \langle \psi | \hat{\mathcal{O}} | \chi \rangle^* \ . \tag{19}$$

With this definition, a Hermitian operator is an operator satisfying $\hat{\mathcal{O}}^\dagger = \hat{\mathcal{O}}$. A basic theorem in linear algebra states that this is true if and only if there exists an orthonormal basis of eigenvectors with real eigenvalues.

Properties of the Adjoint

From the definition of the adjoint, we can immediately see the following properties that will be useful below

- $(z\hat{\mathcal{O}})^\dagger = z^* \hat{\mathcal{O}}^\dagger$
- If $|w\rangle = \hat{\mathcal{O}}|v\rangle$, then $\langle w|u\rangle = \langle v|\hat{\mathcal{O}}^\dagger|u\rangle$
- $(\hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2)^\dagger = \hat{\mathcal{O}}_2^\dagger \hat{\mathcal{O}}_1^\dagger$
- If $\hat{\mathcal{O}}|v\rangle = \lambda |v\rangle$, then $\langle v|\mathcal{O}^\dagger|w\rangle = \lambda^* \langle v|w\rangle$

From the third property here, we see that the product of two Hermitian operators is generally not Hermitian. However, it follows from the properties of the adjoint that the combination $i[\hat{\mathcal{A}}, \hat{\mathcal{B}}]$ is Hermitian if $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ are Hermitian.

Unitary operators

Another type of operator that will be important in quantum mechanics is the *unitary operator*. These are operators that preserve the inner product (and therefore all lengths and distances). Preserving the inner product means that if $|v\rangle$ and $|w\rangle$ are two vectors, and if $\hat{\mathcal{U}}$ is an operator, then the inner product between $\hat{\mathcal{U}}|v\rangle$ and $\hat{\mathcal{U}}|w\rangle$ is the same as the inner product between $|v\rangle$ and $|w\rangle$. This gives the condition

$$\langle v|\mathcal{U}^\dagger\mathcal{U}|w\rangle = \langle v|w\rangle \ . \tag{20}$$

If this is true for all states, it means that the matrix representation of $\mathcal{U}^\dagger\mathcal{U}$ in any basis is the same as the matrix representation of the identity operator $\mathbf{1}$, so they must be the same operator:

$$\mathcal{U}^\dagger\mathcal{U} = \mathbf{1} \ . \tag{21}$$

We can think of a unitary operator as a complex generalization of a rotation: unitary operators take an orthonormal basis to another orthonormal basis, and any operator with this property is unitary.

Relation between unitary operators and Hermitian operators

The simplest unitary operator is the identity operator $\mathbf{1}$. If we consider a family of unitary operators $\hat{\mathcal{U}}(\epsilon)$ with $\hat{\mathcal{U}}(0) = \mathbf{1}$, then for very small ϵ, we can write

$$\hat{\mathcal{U}}(\epsilon) = \mathbf{1} + \epsilon\hat{\mathcal{B}} + \dots \tag{22}$$

where the dots indicate terms at higher orders in epsilon.[4]

If the relation

$$\hat{\mathcal{U}}^\dagger(\epsilon)\hat{\mathcal{U}}(\epsilon) = \mathbf{1} \tag{23}$$

holds for all ϵ, the Taylor expansion of the left side must equal the Taylor expansion of the right side. Inserting the expansion (22) and looking at the terms of order ϵ, we find that

$$\hat{\mathcal{B}}^\dagger + \hat{\mathcal{B}} = 0 \ . \tag{24}$$

Defining $\mathcal{A} = i\mathcal{B}$, we see that the operator $\mathcal{A}$ satisfies $-\mathcal{A}^\dagger + \mathcal{A} = 0$. Thus, if $\mathcal{U}(\epsilon)$ represents a family of unitary operators near with

$$\hat{\mathcal{U}}(\epsilon) = \mathbf{1} - i\epsilon\hat{\mathcal{A}} + \dots \tag{25}$$

[4]If you are not comfortable with Taylor expanding an operator, think of the matrix elements of that operator in some basis. Each of these is an ordinary function of ϵ which can be Taylor expanded. The operator B is the one whose matrix elements are the terms at order ϵ in the Taylor expansion of the matrix elements of $\hat{\mathcal{U}}(\epsilon)$.

then $\mathcal{A}$ must be Hermitian.

We can say that $-i\hat{\mathcal{A}}|v\rangle$ gives the small change in the vector under an *infinitesimal* unitary transformation.

Given any Hermitian operator, we can also go the other direction and define a family of unitary operators "generated" by this operator. From the discussion above, the operator

$$(1 - i\epsilon\mathcal{A}) \tag{26}$$

should be almost a unitary operator when ϵ is very small. It's not quite, because we need to add certain higher order terms in ϵ for (23) to be satisfied. In However, we can generate a family of unitary operators by the following trick. We consider a limit where $\epsilon \to 0$ so that we can ignore any higher order terms, but we also apply the operator multiple times, so that we don't just end up with **1**. Specifically, we define

$$\hat{\mathcal{U}}_{\mathcal{A}}(a) = \lim_{N\to\infty}(1 - i\frac{a}{N}\hat{\mathcal{A}})^N \tag{27}$$

This limit is well known as one definition of the exponential function, so we can write simply

$$\hat{\mathcal{U}}_{\mathcal{A}}(a) = e^{-ia\hat{\mathcal{A}}}\ . \tag{28}$$

We can think of the exponential of an operator as being defined by the power series

$$e^x = 1 + x + \frac{1}{2}x^2 + \ldots \tag{29}$$

or we can work in a basis where the matrix $\mathcal{A}$ is diagonal with eigenvalues λ_n and say that in this basis, the matrix for $\hat{\mathcal{U}}_{\mathcal{A}}(a)$ is diagonal with matrix elements $e^{-ia\lambda_n}$.

In summary, we can associate a Hermitian operator to any one-parameter family of unitary operators near the identity operator, and we can associate a family of unitary operators to any Hermitian operator by exponentiation.

2 Basics of Quantum mechanics

We are now ready to outline the basic structure of quantum mechanics.

2.1 States

The physical configuration of a system at some time is known as a *state*. In quantum mechanics, a state is represented mathematically as a nonzero vector in some Hilbert space.

We will often denote this as $|\Psi\rangle$. Two vectors represent the same physical state if they are related to each other through multiplication by a complex number.[5] Since the zero vector is not allowed, any state vector is equivalent to a state with $\langle\Psi|\Psi\rangle = 1$, so we will often assume this condition and refer to such states as *normalized*. There is still freedom to multiply by a complex number $e^{i\phi}$ with norm 1 (known as a *phase*) without changing the physical configurations.

Generally, this state will evolve with time, so we can write $|\Psi(t)\rangle$ to describe the state vector at time t.

2.2 Observables

In quantum mechanics, physical quantities such as energy, position, or angular momentum (we will generally call them *observables*) do not generally have definite values in a given state. However, a basic assumption is that there exists for each observable $\mathcal{O}$ some orthonormal basis of states $|\lambda_n\rangle$, each of which has a definite value λ_n for $\mathcal{O}$. We call these the eigenvectors (or eigenstates) and eigenvalues associated with $\mathcal{O}$. In some cases, two or more of the eigenvalues can be the same; in this case, the whole space of states spanned by the corresponding eigenvectors is understood to have this same definite value for $\mathcal{O}$, and we have some freedom in choosing which orthonormal basis of states to choose.

What about states which are not eigenstates? Since the eigenstates form a basis, we can write any state $|\Psi\rangle$ as a linear combination of these basis elements,

$$|\Psi\rangle = \sum_n c_n |\lambda_n\rangle \ . \tag{30}$$

For such a state, we say that $|\Psi\rangle$ does not have a definite value for $\mathcal{O}$; in a measurement of $\mathcal{O}$, we might find any of the values λ_n, with probability $|c_n|^2$, and the state becomes the corresponding eigenstate after the measurement. Here, we are assuming that the state is normalized; otherwise $|c_n|^2$ gives the relative probability.

To see that the probabilities add up to 1 in the normalized case, we note that by the orthonormality of the basis,

$$1 = \langle\Psi|\Psi\rangle = \sum_n \sum_m c_n^* c_m \langle\lambda_n|\lambda_m\rangle = \sum_n |c_n|^2 = 1 \ . \tag{31}$$

We can give a more direct formula for the probability by noting that

$$c_n = \langle\lambda_n|\Psi\rangle \ . \tag{32}$$

[5] Mathematically, the set of nonzero vectors in an N-dimensional complex vector space $\mathbb{C}^N$ together with this equivalence relation defines a space called complex projective space $\mathbb{CP}^{N-1}$.

which follows using the orthonormality of the basis by taking the inner product of $|\lambda_n\rangle$ with the two sides of (30). Thus if we measure $\mathcal{O}$ in the state $|\Psi\rangle$, the probability that we will obtain the result λ_n is simply

$$p_n = |\langle\lambda_n|\Psi\rangle|^2 \ . \tag{33}$$

If we make a large number of measurements of the same observable on equivalent states, we will obtain a distribution of possible results governed by these probabilities. The average of these results, known as the *expectation value* of the observable is

$$\bar{\mathcal{O}} = \sum_n p_n \lambda_n = \sum_n \lambda_n |\langle\lambda_n|\Psi\rangle|^2 \tag{34}$$

We may also be interested in how spread out the distribution of results is. A useful measure of this is to compute the standard deviation of the distribution, defined by taking the average value of $(\mathcal{O} - \bar{\mathcal{O}})^2$ and then taking the square root. In quantum mechanics, we call this the *uncertainty* $\Delta\mathcal{O}$ of the observable $\mathcal{O}$ in the state $|\Psi\rangle$. It is given explicitly by

$$\begin{aligned} \Delta\mathcal{O} &= (\sum_n p_n(\lambda_n - \bar{\mathcal{O}})^2)^{\frac{1}{2}} \\ &= (\sum_n p_n \lambda_n^2 - (\sum_n p_n \lambda_n)^2)^{\frac{1}{2}} \ . \end{aligned} \tag{35}$$

2.3 Hermitian operators for observables

Given any observable $\mathcal{O}$, we can define a Hermitian operator $\hat{\mathcal{O}}$ associated with it, by defining

$$\hat{\mathcal{O}}|\lambda_n\rangle = \lambda_n|\lambda_n\rangle \tag{36}$$

for the eigenstates of $\mathcal{O}$ and using linearity to define the action on any other state. By our definition, the eigenvectors of $\hat{\mathcal{O}}$ are orthonormal and the eigenvalues are real, so the operator is Hermitian.

We will see below what the physical interpretation of the state $\hat{\mathcal{O}}|\Psi\rangle$ is; for now we point out the useful result that the expectation value of $\mathcal{O}$ defined above can be calculated for the state Ψ as

$$\bar{\mathcal{O}} = \langle\Psi|\hat{\mathcal{O}}|\Psi\rangle \ . \tag{37}$$

This can be checked using the definition (36) to reproduce the result (36).

Commuting vs noncommuting observables

For two different observables $\mathcal{O}_1$ and $\mathcal{O}_2$, it is typically the case that the eigenvectors of one are not eigenvectors of the other. Physically, this means that a state with a definite value for $\mathcal{O}_1$ will not have a definite value for $\mathcal{O}_2$ and vice versa.

On the other hand, if the two observables do share a common basis of eigenstates, it is possible to know the value of both observables at the same time.

It is not hard to show that two Hermitian operators will share a common basis of eigenstates if and only if the corresponding operators *commute* with each other, i.e. that in acting with the two operators in succession on any state, we get the same result regardless of the order. This will be true if and only if the *commutator* $[\hat{\mathcal{O}}_1, \hat{\mathcal{O}}_2] \equiv \hat{\mathcal{O}}_1\hat{\mathcal{O}}_2 - \hat{\mathcal{O}}_2, \hat{\mathcal{O}}_1$ is equal to zero.

In the case where the operators share a common basis, both $\mathcal{O}_1$ and $\mathcal{O}_2$ will have matrix representations that are diagonal in this basis, so it is clear that the corresponding matrices also commute with each other.

The fact that some observables cannot have definite values simultaneously leads directly to the idea of an *uncertainty principle* in quantum mechanics. Starting from the definition (35) of uncertainty and making use of the Cauchy-Schwarz inequality (see Griffiths for details) it is possible to prove the *generalized uncertainty principle*

$$\Delta\mathcal{O}_1\Delta\mathcal{O}_2 \geq \frac{1}{2}|\langle i[\mathcal{O}_1, \mathcal{O}_2]\rangle| \,. \tag{38}$$

We will see that this leads to the more familiar Heisenberg Uncertainty Principle when applied to the position and momentum operators.

2.4 Physical transformations as unitary operators

In quantum mechanics, physical transformations such as rotations, translations, and time evolution correspond to maps that take states to other states. These maps should be *linear* and *preserve the inner product*, that is they should act as unitary operators.

The linearity property means that if we have a superposition of two states, the transformation of this state will just be the superposition of the superposition of the transformed states. For example, if we have the state $\frac{1}{2}|\uparrow\rangle + \frac{1}{2}|\downarrow\rangle$ for a spin half particle, and we rotate by $\pi/2$ about the y axis so that $|\uparrow\rangle$ goes to $|\rightarrow\rangle$ and $|\downarrow\rangle$ goes to $|\leftarrow\rangle$, the state $\frac{1}{2}|\uparrow\rangle + \frac{1}{2}|\downarrow\rangle$ should be transformed to $\frac{1}{2}|\rightarrow\rangle + \frac{1}{2}|\leftarrow\rangle$. This seems very plausible, but in the end, it is just a basic assumption of quantum mechanics. Mathematically, it means that physical transformations are represented by linear operators.

Preserving the inner product is equivalent to the requirement that a normalized state $|\Psi\rangle$ should transform to another normalized state. This is essentially the statement that probabilities must be conserved: given any orthonormal basis corresponding to a physical observable, if the norm of a state decreased after some physical transformation, the probabilities for the various possible outcomes of a measurement would no longer sum to 1.

2.5 Infinitesimal physical transformations

Physical transformations such as rotations or translations come in continuous families. For example, we can rotate about an axis by any angle θ or translate in the x direction by any distance a. For a quantum system, these families of transformations correspond to families of unitary operators. If $\hat{\mathcal{T}}(a)$ is such a family of operators depending on some parameter a, with $\hat{\mathcal{T}}(0) = \mathbf{1}$, i.e. $a = 0$ corresponds to doing nothing, then it is useful to consider the possible *infinitesimal transformations*, that is, the transformations for infinitesimal values of a. For example, we can consider a rotation by an infinitesimal angle or translation by an infinitesimal distance. These are important, because many more general transformations can be built up by doing a series of these infinitesimal transformations.[6]

According the general discussion of unitary operators above, any infinitesimal transformation can be expressed as

$$\hat{\mathcal{T}}(\epsilon) = \mathbf{1} - i\epsilon\hat{\mathcal{O}} + \dots \tag{39}$$

where $\hat{\mathcal{O}}$ is a hermitian operator. This can be associated some physical observable $\mathcal{O}$. Conversely, given any physical observable $\mathcal{O}$, there is an associated Hermitian operator $\hat{\mathcal{O}}$. From this, we can define an infinitesimal physical transformation as in (39). Applying this transformation many times, we generate a family of physical transformations described by the unitary operators (see discussion in section 1)

$$\hat{\mathcal{T}}(a) = e^{-ia\hat{\mathcal{O}}} \ . \tag{40}$$

Thus we see that there is a one-to-one correspondence between infinitesimal physical transformations (or the families of transformations that they generate) and physical observables. As we will see below, the connection is the familiar(?) one from classical mechanics:

- energy $\leftrightarrow$ time translation/evolution
- momentum in direction $\hat{n} \leftrightarrow$ translation in direction $\hat{n}$
- angular momentum around the axis in direction $\hat{n} \leftrightarrow$ rotations about the axis in direction $\hat{n}$

Normally this connection is discussed as the connection between symmetries and conserved quantities. But only some physical quantities will be conserved and only some transformations are symmetries. We will now see that in quantum mechanics, the physical observables that are conserved quantities are exactly the ones associated with physical transformations that are symmetries.

[6] There are some physical transformations, such as mirror reflection, that cannot be built up from a series of infinitesimal transformations. These transformations are often referred to as *discrete*.

Energy and time evolution

Perhaps the most important physical transformation in quantum mechanics is the operation of *time evolution* that takes us from the state of the system at some time t_0 to the state of the system at some later time $t_0 + \delta t$. This is a unitary operator that we will call

$$\hat{\mathcal{T}}(\delta t, t_0) \; . \tag{41}$$

In many cases, we have a system with *time-translation invariance*: this means that the evolution operator is the same for any time t_0, so $\hat{\mathcal{T}}$ is only a function of δt.

As above, we can think of an infinitesimal version of this operator, where δt is taken to be very small. In this case, we can write

$$\hat{\mathcal{T}}(\delta t, t_0) = \mathbf{1} - i\delta t \frac{\hat{H}}{\hbar} + \dots \tag{42}$$

where $\hat{H}$ is a Hermitian operator that we call the *Hamiltonian*; in general, it can depend on t_0. We will see below that in the time-translation invariant case where it does not, the physical quantity associated with $\hat{H}$ is always conserved, this conserved quantity is what we call *energy*. The constant $\hbar$ introduced above is included so that $\hat{H}$ will have the ordinary units of energy instead of inverse time.

The result (42) gives the infinitesimal form of the time evolution operator. Applying this to a state, we learn that

$$|\Psi(t_0 + \delta t)\rangle = (\mathbf{1} - i\delta t \frac{\hat{H}}{\hbar} + \dots)|\Psi(t_0)\rangle \tag{43}$$

Rearranging this, we get

$$\frac{1}{\delta t}(|\Psi(t_0 + \delta t) - |\Psi(t_0)\rangle) = -i\frac{\hat{H}}{\hbar}|\Psi(t_0)\rangle + \dots \; , \tag{44}$$

where the dots indicate terms of order δt and higher. Finally, in the limit $\delta t \to 0$, the left side becomes the derivative of the state, so we obtain a differential equation for the time evolution of a state

$$i\hbar \frac{d}{dt}|\Psi\rangle = \hat{H}|\Psi\rangle \tag{45}$$

This is known as the *Schrödinger equation*. In this general form, it applies to any quantum system.

Conserved quantities in quantum mechanics

Now that we have some understanding of time evolution, we can discuss what is meant by a conserved quantity in quantum mechanics. Since physical observables don't even have definite values for most states, it is less obvious what we might mean by something being conserved. However, we might consider the following possibilities

- A physical observable $\mathcal{O}$ is conserved if and only if its expectation value is unchanging in time for all states of the system.
- A physical observable $\mathcal{O}$ is conserved if and only if for any state of the system, the probabilities of finding the various eigenvalues λ_n are all unchanging in time.

The latter statement clearly implies the former and appears to be a stronger condition, but we will now see that they are equivalent. Furthermore both are equivalent to the statement that the operator $\hat{\mathcal{O}}$ associated with $\mathcal{O}$ commutes with the Hamiltonian operator, i.e.

$$[\hat{\mathcal{O}}, \hat{H}] = 0 \; . \tag{46}$$

We are assuming here that $\mathcal{O}$ itself is some fixed physical quantity with no inherent time dependence (i.e, we assume $d\hat{\mathcal{O}}/dt = 0$). For the proof, see the homework 3 solutions.

Symmetries in quantum mechanics

A *symmetry* in quantum mechanics is a physical transformation represented by some unitary operator $\hat{\mathcal{T}}$ with the property that if $|\Psi(t)\rangle$ is any solution to the Schrödinger equation, then $\hat{\mathcal{T}}|\Psi(t)\rangle$ is a solution to the Schrödinger equation. This definition is in line with our intuition that symmetries acting on a system give us configurations "equivalent" to the original configuration.

To understand which physical transformations have this property, consider any $|\Psi(t)\rangle$ satisfying the Schrödinger equation (45). Then $\hat{\mathcal{T}}|\Psi(t)\rangle$ will also satisfy the Schrödinger equation for some specific operator $\hat{\mathcal{T}}$ if and only if

$$\begin{aligned} & i\hbar \frac{d}{dt}(\hat{\mathcal{T}}|\Psi\rangle) = \hat{H}(\hat{\mathcal{T}}|\Psi\rangle) \\ \iff \quad & i\hbar \hat{\mathcal{T}} \frac{d}{dt}|\Psi\rangle = \hat{H}\hat{\mathcal{T}}|\Psi\rangle \\ \iff \quad & \hat{\mathcal{T}}\hat{H}|\Psi\rangle = \hat{H}\hat{\mathcal{T}}|\Psi\rangle \\ \iff \quad & [\hat{\mathcal{T}}, \hat{H}]|\Psi\rangle = 0 \; . \end{aligned} \tag{47}$$

This will be true for any possible state $|\Psi\rangle$ if and only if

$$[\hat{\mathcal{T}}, \hat{H}] = 0 \; . \tag{48}$$

Thus, a physical transformation is a symmetry if and only if the corresponding unitary operator commutes with the Hamiltonian.

If we have a continuous family of symmetries such as rotations around an axis or translations, the operators $\hat{\mathcal{T}}(\epsilon)$ representing infinitesimal transformations must also obey the relation (48). From the expression (39) we see that this will be true if and only if the Hermitian operator $\hat{\mathcal{O}}$ associated with this transformation also commutes with the Hamiltonian.

Thus, the condition that a Hermitian operator gives an infinitesimal transformation that is a symmetry is

$$[\hat{\mathcal{O}}, \hat{H}] = 0 \; . \tag{49}$$

Symmetries $\leftrightarrow$ conservation laws

We have already seen that any Hermitian operator can be associated with a physical observable and also a physical transformation. We now see that the condition (46) that the physical observable associated with $\hat{\mathcal{O}}$ is conserved is the same as the condition (49) that the physical transformation associated with $\hat{\mathcal{O}}$ is a symmetry. This establishes the relation between symmetries and conservation laws in quantum mechanics.

Time translations and energy

The most universal example of a operator that commutes with the Hamiltonian is the Hamiltonian itself. Thus, for any quantum mechanical system with time-translation invariance (i.e. for which a single time-independent operator governs the time evolution at all times), the observable associated with the Hamiltonian operator must be conserved. From our experience with classical mechanics, we know that the conserved quantity associated with time-translation invariance is the *total energy* of the system. Thus, the Hamiltonian operator appearing in the Schrödinger equation is the energy operator.

A useful consequence of this is that to understand time-evolution in quantum mechanics, it is very useful to find the eigenstates of the energy operator. If $|E_n\rangle$ represent these energy eigenstates, we have by definition

$$\hat{H}|E_n\rangle = E_n|E_n\rangle \tag{50}$$

so the time evolution operator acts as[7]

$$\hat{\mathcal{T}}(t)|E_n\rangle = e^{-\frac{i}{\hbar}\hat{H}t}|E_n\rangle = e^{-\frac{i}{\hbar}E_n t}|E_n\rangle \tag{51}$$

We see that the state of the system after time t is physically equivalent to the original state, since it is just the original state multiplied by a phase factor. In particular, the probabilities for any possible measurements will be the same as for the initial state. For this reason, we call the energy eigenstates *stationary states*.

Using the linearity property of time evolution in quantum mechanics, it is now simple to understand the evolution of any state. We first take the state at $t = 0$ and express it as a

[7] We are using here that if a state is an eigenstate of some operator, it will be an eigenstate of any power of that operator and more generally any (analytic) function of that operator. Alternatively, we can just check that our final result satisfies the Schrödinger equation.

combination of energy eigenstates:

$$|\Psi(t=0)\rangle = \sum c_n |E_n\rangle \qquad\qquad c_n = \langle E_n|\Psi(t=0)\rangle \ . \tag{52}$$

By linearity, we can immediately say that the state at a later time is

$$|\Psi(t)\rangle = \hat{\mathcal{T}}(t)|\Psi(t=0)\rangle = \sum c_n \hat{\mathcal{T}}(t)|E_n\rangle = \sum c_n e^{-\frac{i}{\hbar}E_n t}|E_n\rangle \ . \tag{53}$$

Thus, the problem of understanding time evolution is reduced to the problem of finding the energy eigenstates, i.e. solving (50). For this reason, (50) is often called the time-independent Schrödinger equation.

Translations and momentum

Another important example of an observable being related to some physical transformation is the connection between translations and momentum in quantum systems. Let $\hat{\mathcal{T}}(a)$ be the operator that translates a system by an amount a in some direction. Then we can write the infinitesimal version of this transformation as

$$\hat{\mathcal{T}}(a) = \mathbf{1} - ia\frac{1}{\hbar}\hat{P} + \dots \ . \tag{54}$$

By our general discussion above, the Hermitian operator $\hat{P}$ will correspond to a physical quantity that is conserved for any system where these translations are a symmetry. From classical mechanics, we know that this is the defining property of *momentum*, so we can say that $\hat{P}$ is the quantum operator associated with momentum. We have again included the constant $\hbar$ into the definition above in order that $\hat{P}$ will have the usual units of momentum.

For a system with a single spatial direction labeled by coordinate x, we can understand better the properties of the momentum operator by noting that if a state $|\Psi\rangle$ is defined by wavefunction $\psi(x)$ the the state $\hat{\mathcal{T}}(a)|\Psi\rangle$ must have wavefunction $\psi(x-a)$ i.e. the same function translated in the x direction by an amount a. This is implied by the definition of $\hat{\mathcal{T}}(a)$. Mathematically, this gives

$$\langle x|\hat{\mathcal{T}}(a)|\Psi\rangle = \psi(x-a) \ . \tag{55}$$

This equation is true for all a, so it must be true for very small a where (54) holds. in this case, we can expand both sides in a Taylor series in a, and all the terms must be equal. Inserting (54) into (55) and equating the terms with one power of a, we find

$$-ia\frac{1}{\hbar}\langle x|\hat{P}|\Psi\rangle = -a\psi'(x) \tag{56}$$

or

$$\langle x|\hat{P}|\Psi\rangle = -i\hbar\frac{d}{dx}\psi(x) \ . \tag{57}$$

We see that acting with the momentum operator on a state $|\Psi\rangle$ is equivalent to acting with $-id/dx$ on the wavefunction for that state.

Using this result, one application is to derive the commutator between the position operator and the momentum operator corresponding to that same direction. To do this, we note that the position operator $\hat{X}$ obeys $\hat{X}|x\rangle = x|x\rangle$, so for any $|\Psi\rangle$,

$$\langle x|\hat{X}\hat{P}|\Psi\rangle = x\langle x|\hat{P}|\Psi\rangle = -i\hbar x\frac{d}{dx}\psi(x) \ . \tag{58}$$

On the other hand, the wavefunction for the state $\hat{X}|\Psi\rangle$ is

$$\langle x|\hat{X}|\Psi\rangle = x\langle x|\Psi\rangle = x\psi(x) \tag{59}$$

so by our result above, the wavefunction $\langle x|\hat{P}\hat{X}|\Psi\rangle$ for the state $\hat{P}(\hat{X}|\Psi\rangle)$ is

$$-i\frac{d}{dx}(x\psi(x)) = -i\hbar\psi(x) - i\hbar x\frac{d}{dx}\psi(x) \ . \tag{60}$$

Thus, we finally have that

$$\langle x|[\hat{X},\hat{P}]|\Psi\rangle = i\hbar\psi(x) = i\hbar\langle x|\Psi\rangle \ . \tag{61}$$

This implies that the matrix elements of the operator $[\hat{X},\hat{P}]$ are the same as the matrix elements of the operator $i\hbar\mathbf{1}$, so it must be that

$$[\hat{X},\hat{P}] = i\hbar\mathbf{1} \ . \tag{62}$$

This fundamental relation will be very useful in understanding a variety of quantum mechanical systems. Based on the generalized uncertainty principle above, this tells us immediately that

$$\Delta X\Delta P \geq \frac{\hbar}{2} \ . \tag{63}$$

Example: a particle in one dimension

As an example for how to use all of this formalism, consider a physical system describing a particle in one dimension subject to forces corresponding to a potential energy function $V(x)$. In classical mechanics, the energy for such a system is given by

$$E = \frac{p^2}{2m} + V(x) \ . \tag{64}$$

For the quantum system, time evolution is governed by the Hamiltonian, which is the same as the energy operator. In quantum mechanics, position and momentum can't have definite values at the same time, so at best, the relation (64) could be true about the expectation

values of these quantities. We can arrange this by translating (64) to a statement about operators

$$\hat{H} = \frac{\hat{P}^2}{2m} + V(\hat{X}) \ . \tag{65}$$

With this definition, we can now translate the general form (45) for the Schrödinger equation to a differential equation describing the evolution of the position-space wavefunction $\psi(x)$. We'll call the wavefunction at time t $\psi(x,t)$. Starting from (45), we can take the inner product on both sides with the position space basis element $|x\rangle$ to obtain

$$\begin{aligned}
& \langle x|i\hbar\frac{d}{dt}|\Psi\rangle = \langle x|\hat{H}|\Psi\rangle \\
\iff \quad & i\hbar\frac{d}{dt}\langle x|\Psi\rangle = \langle x|\frac{\hat{P}^2}{2m} + V(\hat{X})|\Psi\rangle \\
\iff \quad & i\hbar\frac{\partial}{\partial t}\psi(x,t) = \frac{1}{2m}\langle x|\hat{P}^2|\Psi\rangle + \langle x|V(\hat{X})|\Psi\rangle \\
\iff \quad & i\hbar\frac{\partial}{\partial t}\psi(x,t) = \frac{1}{2m}(-i\hbar\frac{d}{dx})^2\langle x|\Psi\rangle + \langle x|V(x)|\Psi\rangle \\
\iff \quad & i\hbar\frac{\partial}{\partial t}\psi(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) + V(x)\psi(x,t)
\end{aligned} \tag{66}$$

Thus, we have derived the usual position-space form of the 1D Schrödinger equation, appearing on page 1 of Griffiths.

3 Multipart quantum systems

It is often the case that a quantum mechanical system that we are interested in has various independent degrees of freedom i.e. separate parts that each have their own state. This is the case in a system with many particles, but also in systems where a single particle has both x motion and y motion, or x, y, z motion and spin.

To understand the quantum description of multipart systems, consider an example where we have two particles in one dimension. Classically, each would have a definite value for position. We can describe these by introducing variables x_1 and x_2 describing the locations of the two particles. But in the quantum system, states with definite values for x_1 and x_2 are very special states - they are eigenstates for the two position operators $\hat{X}_1$ and $\hat{X}_2$. If we label these states as $|x_1 x_2\rangle$, then the general state will be a linear combination of these. We need a function $\psi(x_1, x_2)$ to describe the coefficients in such a linear combination, so we see that the wavefunction for the two-part system is a function of two variables rather than just two functions of one variable.

The key point here is that the basis for the combined system is in one-to-one correspondence with pairs $(|x_1\rangle, |x_2\rangle)$ of basis elements, one from each part. Mathematically, a vector space

with basis constructed in this way is known as the *tensor product* or direct product of the the two smaller vector spaces.[8].

In a case where both parts are represented by finite dimensional Hilbert spaces, if we have d_1 basis elements $|n\rangle$ for the first part and d_2 basis elements $|N\rangle$ for the second part, there are $d_1 d_2$ basis elements for the tensor product Hilbert space. These are labeled either as $|nN\rangle$ or $|n\rangle \otimes |N\rangle$, where the $\otimes$ in the latter notation indicates that there is a bilinear operation that takes one vector from the first Hilbert space and one vector from the second Hilbert space and gives us a vector in the larger Hilbert space. We can define this operation on any two vectors $|\psi\rangle = \sum_n c_n |n\rangle$ and $|\Psi\rangle = \sum c_N |N\rangle$ by

$$|\psi\rangle \otimes |\Psi\rangle = (\sum_n c_n |n\rangle) \otimes (\sum c_N |N\rangle) \equiv \sum_{nN} c_n C_N |n\rangle \otimes |N\rangle \ . \tag{67}$$

Physically, this tensor product state represents a state where the first subsystem is definitely in the state $|\psi\rangle$ while the second part is definitely in the state $|\Psi\rangle$. A very important point is that only very special states can be written in this way. For example, in a system with two spins, the state

$$\frac{1}{\sqrt{2}} \left(|\uparrow\rangle \otimes |\uparrow\rangle + |\downarrow\rangle \otimes |\downarrow\rangle \right) \tag{68}$$

cannot be written as a tensor product; in this case we say that the first subsystem is *entangled* with the second subsystem.

We emphasize that the two vector spaces need not have the same dimension. For example, when describing a particle with position degrees of freedom and also spin degrees of freedom, we would have basis elements $|x s_z\rangle \equiv |x\rangle \otimes |s_z\rangle$ where the first part is infinite dimensional and the second part is finite dimensional. For a system like this, the general state would be a superposition

$$\int dx (\psi_\uparrow(x) |x \uparrow\rangle + \psi_\downarrow(x) |x \downarrow\rangle) \tag{69}$$

so we would have an "up" wavefunction and a "down" wavefunction to describe the state.

For a multipart system, given any operator acting on one of the individual Hilbert spaces, we can promote it to an operator acting on the full Hilbert space simply by declaring that it has no effect (i.e. that it acts as the identity operator) on the remaining parts. In our example with two finite-dimensional systems, if $\hat{\mathcal{O}}^1$ is an operator acting on the first part as

$$\hat{\mathcal{O}}^1 |n\rangle = \sum_m \mathcal{O}^1_{mn} |m\rangle \ , \tag{70}$$

we can promote it to an operator acting on the whole system as

$$\hat{\mathcal{O}}^1 |nN\rangle \equiv \sum_m \mathcal{O}^1_{mn} |mN\rangle \ . \tag{71}$$

[8]This should be distinguished with a direct sum, in which the set of basis elements for the full space is taken to be the union of basis elements for the subspaces

Sometimes the full operator is denoted by $\hat{\mathcal{O}}_1 \otimes \mathbb{1}$ to emphasize that it is acting as the identity operator on the second part.

It is easy to check from the definition that operators defined in this way but acting on the different parts automatically commute with one another. Thus, we can write the product of operators acting on the different parts as $\mathcal{O}^1\mathcal{O}^2$ or as $\mathcal{O}^2\mathcal{O}^1$; in the $\otimes$ notation, both of these are equivalent to $\mathcal{O}^1 \otimes \mathcal{O}^2$. As an example, consider the operator $\hat{x}\hat{S}_x$ in our system describing the position and the spin of a particle. This is the product of the operator obtained by promoting $\hat{x}$ to an operator on the full system with the operator obtained by promoting $\hat{S}_x$ to an operator acting on the full system. Acting on an (x, S_z) basis state $|x \downarrow\rangle$, this gives

$$\hat{x}\hat{S}_x|x \downarrow\rangle = \frac{\hbar}{2}x|x \uparrow\rangle \ . \tag{72}$$

Just as the states of the combined system cannot usually be written in the form $|\Psi_1\rangle \otimes |\Psi_2\rangle$, general operators of the combined system cannot usually be written as $\mathcal{O}^1 \otimes \mathcal{O}^2$. Typically, the best we can do is write general states/operators as linear combinations of things that can be written in this way.