

State

State of a physical system at a given time is basically all information that identifies the particular state the system is in. For example, “I am reading these notes” is a description of your current state, but not a complete one. To describe your state accurately, you also need to give the position of your hands, head, etc., but should also give the position and velocity of each particle your body contains.

The state of a classical particle in 1D can be specified completely by giving its position x and momentum p (at a given time). By using this information, you can compute everything that can be computed. For example, you can find its energy, its state one second later (i.e., position and momentum 1s later) as a function of the current state (x, p) . The state for such a particle can be *mathematically represented* by a point in the 2-dimensional *phase space*. For the general case of a classical system composed of N classical particles moving in 3D, the state can be specified by giving a total of $6N$ coordinates of position and momenta. Equivalently, the state of that system at a given time, can be mathematically represented by a point in $6N$ -dimensional phase space.

For quantum mechanical systems, you already know that the classical phase space is not suitable for describing any state of the system. There is a more complicated (but simpler in certain respects) mathematical structure for representing the states of quantum mechanical systems. This has to be so, because new concepts not met in classical systems, such as *superposition*, arise in quantum mechanics and your mathematical apparatus should be appropriate to handle these.

It appears that, any state of a quantum mechanical system can be *mathematically represented* as a *normalized vector* in a certain *Hilbert space*. The classical phase space then has to be replaced by this Hilbert space. Such vectors satisfy the basic property you expect from states, i.e., they contain *all the information* about the actual physical state the system is in. Any property of the system in that particular state can be computed by using that vector. But it appears that, some certain properties familiar from classical mechanics are not meaningful and hence cannot be computed. For example, you cannot get a meaningful answer to “what is the exact position of this particle in this particular state” and you cannot compute it. But you can ask “what is the probability that this particle will appear around x when I measure its position” and you can get the answer by computing a certain expression involving the vector that contains the state information. In particular, the vector contains all information that can be produced in actual experiments.

This is why, we now turn to the description of Hilbert spaces, which will be our playground for the whole semester.

A Hilbert space is a

- (1) a vector space,
- (2) with an inner product defined on it,
- (3) which is also complete.

This is how a Hilbert space is defined mathematically. But, in quantum mechanics we also require the Hilbert spaces to have two additional properties. First, we must use complex numbers when handling these vectors. The complex numbers are needed for appropriately handling the phase information. So the space has to be a *complex Hilbert space*. We also need an additional property called *separability*, which is equivalent to saying that the Hilbert space should have a *countable* dimension. (If you don't know what this means, you can safely forget it.)

Vector Spaces

A **vector space** is a set satisfying certain properties. The elements of that set are called **vectors**. There are two operations defined on that set, the **vector addition** and **multiplication with a scalar**. You are probably already familiar with these, but here we give the following definitions for the sake of completeness.

In vector addition, you take two vectors and add them which gives you a third vector: $\vec{A} + \vec{B} = \vec{C}$. It satisfies commutativity, $\vec{A} + \vec{B} = \vec{B} + \vec{A}$, and associativity $\vec{A} + (\vec{B} + \vec{D}) = (\vec{A} + \vec{B}) + \vec{D}$. There is a unique element, $\vec{0}$, called *zero vector*, which we usually represent simply by 0. It satisfies $\vec{A} + 0 = 0 + \vec{A} = \vec{A}$. Finally for each vector \vec{A} , there is another vector $(-\vec{A})$ which satisfies $\vec{A} + (-\vec{A}) = 0$. In short, the vector addition is quite similar to the number addition; they are different only in the fact that you add vectors in vector addition.

In multiplication with a scalar, you take a vector \vec{A} and multiply it with a *scalar number* c that produces another vector $c\vec{A}$. If c are always have to be real numbers, then we have a **real vector space**. If c can be complex, then we have a **complex vector space**. In the following, we will represent the scalar numbers by lowercase roman letters a, b, c , etc. without explicitly saying that they are scalars. We don't differentiate between multiplication from left or right, so $c\vec{A} = \vec{A}c$. This operation satisfies a few properties: (i) $b(c\vec{A}) = c(b\vec{A}) = (cb)\vec{A}$, (ii) $c(\vec{A} + \vec{B}) = c\vec{A} + c\vec{B}$, (iii) $(b + c)\vec{A} = b\vec{A} + c\vec{A}$, (iv) $0\vec{A} = 0$, (v) $1\vec{A} = \vec{A}$, (vi) $(-1)\vec{A} = (-\vec{A})$ etc. In short, it has all the properties you can expect. Note that in the notation above we don't distinguish between number 0 and vector 0. They are completely different mathematical objects; one is a vector and the other is a number. But, since it does not cause big confusions, we can safely use the same notation for both.

Note that we have not defined a multiplication operation for two vectors. For a simple vector space, such operations are not required.

Examples

The 3 dimensional vectors \mathbb{R}^3 is a real vector space. Any vector \vec{A} can be expressed as $\vec{A} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$ where a_x, a_y and a_z are scalars.

The n -dimensional column matrices ($n \times 1$ matrices) forms a (real or complex) vector space. Any vector can be expressed as

$$\vec{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},$$

where a_i are (real or complex) scalars. Such matrices will be called **column vectors**. They are going to be important for us because any Hilbert space can be represented as column vectors. The zero vector is,

$$0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

The vector addition and multiplication with scalars are carried out for each row separately, i.e.,

$$\vec{A} + \vec{B} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} , \quad c\vec{A} = c \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{bmatrix} .$$

The vector spaces above should already be familiar to you. But, there are also some unconventional examples of vector spaces. Note that any set where you can define vector addition and multiplication with a scalar operations, satisfying the properties above, can be appropriately called a vector space. One example is functions. Consider, for example, complex-valued functions of a real variable, i.e., functions F where $F : \mathbb{R} \rightarrow \mathbb{C}$. Note our notation: $F(x)$ is the value of the function F at the point x . $F(x)$ is basically a scalar number. But, we use F to denote the function itself.

We can define vector addition and multiplication with a scalar operations as follows.

$$\begin{aligned} H &= F + G & \text{if } H(x) &= F(x) + G(x) \text{ for all } x \quad , \\ H &= cF & \text{if } H(x) &= cF(x) \text{ for all } x \quad . \end{aligned}$$

The zero vector is obviously that unique function F where $F(x) = 0$ for all x in which case we simply say $F = 0$. Note that we have not imposed any continuity or convergence conditions on functions. Addition of such conditions will define different vector spaces with elements formed from functions (Example: polynomials with degree less than 2 forms a vector space).

It might be hard to visualize functions as vectors as we usually tend to imagine vectors as pointed arrows. However, what is important in here is the basic operations and the implications they contain. For that reason, it is mathematically convenient to investigate the function spaces as vector spaces once you have learned all about vector spaces. Also, note the similarity of functions to column vectors. A function F is basically an infinite list of numbers $F(x)$ associated with all real numbers x . An n -dimensional column vector \vec{A} , on the other hand, is a finite list of numbers a_1, a_2, \dots, a_n . They are different only in the aspect that for functions we have an infinite number of “rows” and the “row index” x is a continuous variable.

These are not the exclusive list of vector spaces. There are vector spaces which are constructed on more complicated mathematical sets (e.g., the Fock space in QM). However, in all cases, it is usually sufficient to know that the set itself is a vector space (or a Hilbert space for QM) and this enables us to do a number of computations. For such cases, you don’t need to know the exact detailed nature of the vector space, i.e., you don’t need to know if your vector is a function, a column matrix or a pointed arrow.

In such cases, it is useful to follow a standard notation for vectors, independent of what their exact nature are. We will use the notation $|\psi\rangle$ for the vectors. We will also call these vectors as *kets*. The zero vector will again be represented by the same symbol, 0. These are the basic building blocks of the *Dirac notation* which appears to be quite valuable for quantum mechanics. We will appreciate the usefulness of the Dirac notation later. We will choose the labels of kets from the lowercase Greek letters, $\alpha, \beta, \psi, \phi$ etc. But in some cases, we may prefer to put a list of certain quantum numbers as the label. The context will tell you exactly what kind of vector you have. But, in special cases of column vectors, functions etc. we will leave the Dirac notation occasionally. In short, we have a vector space formed by vectors/kets where vector addition and multiplication with a scalar operations are defined: $|\alpha\rangle + |\beta\rangle = |\gamma\rangle, c|\psi\rangle = |\phi\rangle$.

A few things about vector spaces

For a given set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$, the vector that can be formed by the allowed operations

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + \dots + c_m |\alpha_m\rangle$$

is called a **linear combination** of the vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$. In quantum mechanics, we usually prefer the term **superposition** instead of linear combination. The set of all vectors, which can be written as a superposition of this kind will also form a vector space which is called a **subspace**. In that case, we say that $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ **spans** the subspace. (Example: $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ spans the two dimensional subspace, the xy -plane of \mathbb{R}^3 .) Note that subspaces always contain the zero vector, 0.

The set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is called **linearly dependent** if the equation

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + \dots + c_m |\alpha_m\rangle = 0$$

has a non-zero solution for c_i (i.e., any solution other than the trivial solution $c_1 = c_2 = \dots = c_m = 0$). In that case, we can express one of these vectors as a superposition of others. (Example: The set $\{\vec{A}, \vec{B}, \vec{C}\}$ where $\vec{A} = \hat{\mathbf{i}}, \vec{B} = \hat{\mathbf{j}}$ and $\vec{C} = 3\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ is linearly dependent. In that case we have $3\vec{A} + 2\vec{B} - \vec{C} = 0$. We can express any vector as a superposition of others: $\vec{A} = (\vec{C} - 2\vec{B})/3$ or $\vec{B} = (\vec{C} - 3\vec{A})/2$ etc.)

The set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is called **linearly independent** if the equation

$$c_1 |\alpha_1\rangle + c_2 |\alpha_2\rangle + \dots + c_m |\alpha_m\rangle = 0$$

has only the trivial solution for c_i , i.e., $c_1 = c_2 = \dots = c_m = 0$. In such a case, it is not possible to express any $|\alpha_i\rangle$ as a superposition of others.

The set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is called a **basis** if (1) it is linearly independent and (2) it spans the whole vector space. In that case, it can be shown that all bases have the same number of elements and that number (here m) is called the **dimension** of the vector space. In such a case, *any vector* $|\psi\rangle$ that belongs to the vector space, can be written as a superposition of these m vectors,

$$|\psi\rangle = b_1 |\alpha_1\rangle + b_2 |\alpha_2\rangle + \dots + b_m |\alpha_m\rangle \quad .$$

Moreover, as these m vectors are linearly independent, the numbers b_1, b_2, \dots, b_m are unique. In other words, you cannot express the same $|\psi\rangle$ as two different superpositions. There is only one superposition that gives $|\psi\rangle$. This enables us to identify the vector $|\psi\rangle$ by the m -tuple of scalars (b_1, b_2, \dots, b_m) or by $m \times 1$ column vectors. Therefore, any finite dimensional vector space has a *matrix representation*.

Examples

The three vectors $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is a basis for the 3-dimensional vector space \mathbb{R}^3 . It is the standard basis we use. But, the three vectors $\{\hat{\mathbf{i}} + \hat{\mathbf{j}}, 2\hat{\mathbf{i}} + \hat{\mathbf{j}} - 5\hat{\mathbf{k}}, 137\hat{\mathbf{j}} - \pi\hat{\mathbf{k}}\}$ is also a basis. Any vector of \mathbb{R}^3 can be expressed in either of these bases. The vectors $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\}$ is a basis for the 2-dimensional subspace it generates, the xy -plane.

For $n \times 1$ column matrices, the standard basis is the set of vectors $\{u_1, u_2, \dots, u_n\}$ where u_k is the column vector whose k th row is 1 and all other elements are 0, i.e.,

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad u_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Then there are infinite dimensional vector spaces. The functions of a real variable is an example. However, in such a case, constructing a basis is problematic. The basic reason is: You have a superposition of an infinite number of vectors, i.e., a series. In series sums however, we meet with convergence problems. You should be able to check whether your infinite superposition converges to a vector or it diverges. You can check convergence only if you have defined some “distance” between vectors. This implies that you can compute the “length” of all vectors. Therefore, we should add more structure on the vector spaces, only in such a case constructing bases for infinite dimensional spaces becomes meaningful. The inner product enables us to do that.

Inner Product

Inner product on a vector space is an operation that takes two vectors and produces a *scalar number* (real or complex depending on whether the vector space is real or complex). For two vectors $|\psi\rangle$ and $|\phi\rangle$, the inner product of these two vectors is denoted as $\langle\phi|\psi\rangle$. It is required to satisfy the following properties

- (1) $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$ where $*$ denotes the complex-conjugation operation.
- (2) Linearity in the right vector: If $|\psi\rangle = c|\alpha\rangle + d|\beta\rangle$, then $\langle\phi|\psi\rangle = c\langle\phi|\alpha\rangle + d\langle\phi|\beta\rangle$.
- (2') (2) together with (1) implies anti-linearity in the left vector: If $|\phi\rangle = a|\gamma\rangle + b|\lambda\rangle$, then $\langle\phi|\psi\rangle = a^*\langle\gamma|\psi\rangle + b^*\langle\lambda|\psi\rangle$.
- (3) For any vector $|\psi\rangle$, $\langle\psi|\psi\rangle \geq 0$. (Note that (1) implies that $\langle\psi|\psi\rangle$ is a real number.)

(4) $\langle \psi | \psi \rangle = 0$ only if the vector itself is zero, $|\psi\rangle = 0$. Otherwise, if $|\psi\rangle$ is a non-zero vector, then $\langle \psi | \psi \rangle$ is a positive number different from zero.

For a real vector space, we have $\langle \phi | \psi \rangle = \langle \psi | \phi \rangle$, which is a real number. In that case, the inner product is linear in both vectors. However, in quantum mechanics we must use the complex case.

Examples

The *dot product* or *scalar product* of vectors in \mathbb{R}^3 is an inner product. It is usually defined in a geometric way: $\vec{A} \cdot \vec{B} = |A| |B| \cos \theta$ where $|A|$ and $|B|$ are the lengths and θ is the angle between these vectors. It naturally follows that $\vec{A} \cdot \vec{A} = |A|^2$, the length square of the vector.

For $n \times 1$ column vectors, the inner product is usually defined as

$$\text{if } u = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad v = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \text{then } \langle u | v \rangle = a_1^* b_1 + a_2^* b_2 + \cdots + a_n^* b_n \quad .$$

The inner product is usually expressed as the matrix product, $\langle u | v \rangle = u^\dagger v$, where u^\dagger is the *hermitian conjugate* of u , i.e., the matrix obtained by converting all columns into rows and taking complex conjugate of all elements,

$$u^\dagger = [a_1^* \quad a_2^* \quad \cdots \quad a_n^*] \quad .$$

(If you are not familiar with these, then you need to learn some basic matrix algebra).

For two functions, F and G , of a real variable, the usual way to define their inner product is

$$\langle F | G \rangle = \int_{-\infty}^{+\infty} F^*(x) G(x) dx \quad .$$

Of course, such a definition is useful only if the integral above is convergent for all vectors in your vector space. We can achieve this if we choose the vector space from *square-integrable functions*, i.e., those functions F where

$$\int_{-\infty}^{+\infty} |F(x)|^2 dx < \infty \quad .$$

Such functions form a vector space. Moreover, the inner product of two such functions is a finite value. Completeness can be shown by advanced mathematical methods. As a result, the square-integrable functions form a Hilbert space.

A few things about the inner products

The quantity $\|\psi\| = \sqrt{\langle \psi | \psi \rangle}$ is called the **norm** or **length** of the vector $|\psi\rangle$. Therefore, the property (4) is basically saying that all non-zero vectors have non-zero lengths and only the zero vector has zero length.

A vector with norm 1, $\|\psi\| = 1$, is called a **normalized** or a unit vector. For any non-zero vector $|\psi\rangle$, the vector

$$|\psi'\rangle = \frac{1}{\|\psi\|} |\psi\rangle$$

is normalized and that process is called normalization.

Two vectors $|\psi\rangle$ and $|\phi\rangle$ are called **orthogonal** (or simply perpendicular) if their inner product is zero, $\langle\psi|\phi\rangle = 0$. Such geometric notions (length, orthogonality) are useful in imagining these vector spaces. Of course, saying that “two functions are perpendicular to each other” does not make much sense to you, but it enables you to see certain things clearly if you think of it as being similar to 3D vectors, “the pointed arrows”. This analogy usually works, but there are some complications due to the presence of complex numbers.

One example is the *Pythagorean theorem*. If $|\alpha\rangle$ and $|\beta\rangle$ are orthogonal and $|\gamma\rangle = |\alpha\rangle + |\beta\rangle$ then

$$\|\gamma\|^2 = \|\alpha\|^2 + \|\beta\|^2 \quad .$$

In that case, it is useful to think of $|\alpha\rangle$, $|\beta\rangle$ and $|\gamma\rangle$ as forming the sides of a right-angled triangle. Of course, there is a generalization of this theorem also: If $|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_n\rangle$ are mutually orthogonal n -vectors and $|\gamma\rangle = |\alpha_1\rangle + |\alpha_2\rangle + \dots + |\alpha_n\rangle$ then

$$\|\gamma\|^2 = \|\alpha_1\|^2 + \|\alpha_2\|^2 + \dots + \|\alpha_n\|^2 \quad .$$

You can also show that the triangle inequality is valid (for example by using the Schwarz inequality below): If $|\gamma\rangle = |\alpha\rangle + |\beta\rangle$ (whatever the angle between them is) then

$$\|\gamma\| \leq \|\alpha\| + \|\beta\| \quad ,$$

i.e., one side of a triangle is smaller than the sum of the other two sides. Such relations are useful in telling you that some familiar geometric notions are present in Hilbert spaces as well. However, most of these are not particularly useful in quantum mechanics.

Schwarz inequality

Schwarz inequality is one exception. It will be used in deriving the *uncertainty relations*. It says that for any two kets, the inner product cannot exceed the product of norms of these vectors,

$$|\langle\psi|\phi\rangle| \leq \|\psi\| \|\phi\| \quad .$$

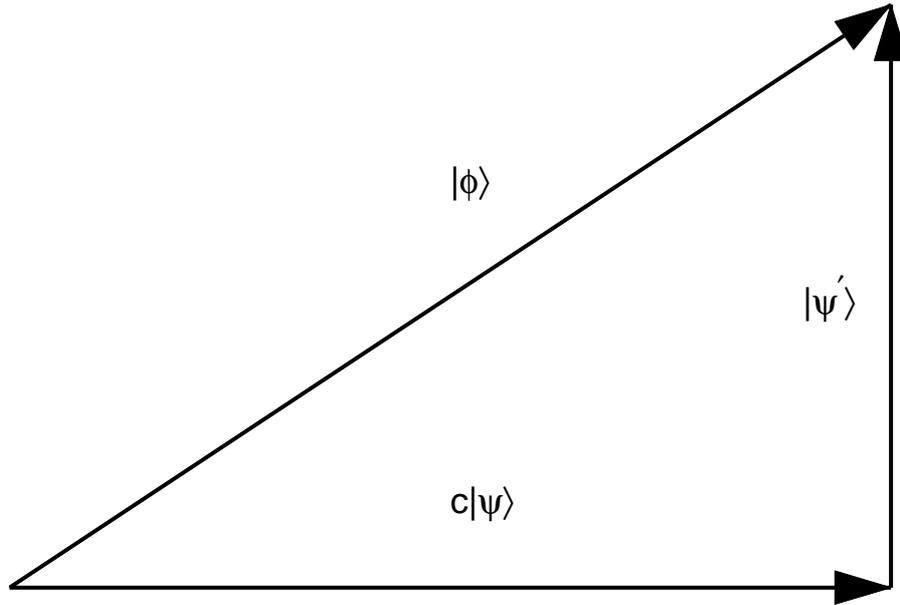
It can be proved as follows. Suppose that you want to express $|\phi\rangle$ as a superposition of a vector parallel to $|\psi\rangle$ and a vector perpendicular to $|\psi\rangle$, i.e.,

$$|\phi\rangle = c |\psi\rangle + |\psi'\rangle$$

where c is a scalar number and $|\psi'\rangle$ is a vector such that $\langle\psi|\psi'\rangle = 0$. You can easily see that

$$c = \frac{\langle\psi|\phi\rangle}{\|\psi\|^2} \quad ,$$

$$|\psi'\rangle = |\phi\rangle - c |\psi\rangle = |\phi\rangle - \frac{\langle\psi|\phi\rangle}{\|\psi\|^2} |\psi\rangle \quad ,$$



is the solution. Then, by Pythagorean theorem, we have

$$\|\phi\|^2 = |c|^2 \|\psi\|^2 + \|\psi'\|^2 \geq |c|^2 \|\psi\|^2 \quad .$$

Inserting the value of c we have found into the last inequality, we get the Schwarz inequality. The vector $c|\psi\rangle$ as defined above is called the component of $|\phi\rangle$ along the vector $|\psi\rangle$ (also called the projection of $|\phi\rangle$ onto $|\psi\rangle$). What the Schwarz inequality is saying is that the component of a vector has a smaller length than the length of the vector.

Note that, the inequality becomes an equality only if $|\psi'\rangle$ is zero-vector, in other words, when $|\phi\rangle$ and $|\psi\rangle$ are parallel: $|\phi\rangle = c|\psi\rangle$.

The Schwarz inequality can be used to define the angle θ between two vectors $|\phi\rangle$ and $|\psi\rangle$ as follows: $|\langle\phi|\psi\rangle| = \|\phi\| \|\psi\| \cos\theta$. The inequality basically tells us that $\cos\theta \leq 1$, in other words, θ can be computed. Due to the appearance of the inner product inside absolute value, $\cos\theta$ can never be negative, i.e., θ cannot be greater than 90° . This is one complication introduced by the complexity of the space. However, two special angles have clear meanings. When θ is 90° , we have $\cos\theta = 0$ and the vectors are orthogonal. When $\theta = 0$, we have $\cos\theta = 1$ and we say that the vectors are “parallel”.

A comment on parallelism: The use of complex numbers prevents a complete analogy in here. For real vector spaces we define parallelism as follows: Two vectors \vec{A} and \vec{B} are parallel to each other if $\vec{A} = c\vec{B}$ where c is a positive number. Vectors \vec{A} and \vec{B} are anti-parallel if $\vec{A} = -c\vec{B}$ for some $c > 0$. In other words, parallel vectors point in the same direction and anti-parallel vectors point in the opposite direction.

Such a classification is not appropriate for complex vector spaces. It might be tempting to say that $|\psi\rangle$ and $-|\psi\rangle$ are anti-parallel to each other, but what about $|\psi\rangle$ and $i|\psi\rangle$? Or an infinite number of others: $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$. For this reason, we say that the vectors are parallel in all such cases. In other words, $|\phi\rangle$ and $|\psi\rangle$ are parallel if $|\phi\rangle = c|\psi\rangle$ for some complex number c . But, depending on the phase angle of c , we say that there is a *phase difference*

between $|\phi\rangle$ and $|\psi\rangle$. For example, $|\psi\rangle$ and $i|\psi\rangle$ have a 90° phase difference; $|\psi\rangle$ and $-|\psi\rangle$ have 180° phase difference etc.

Orthonormal Bases

A set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is called **orthonormal** if (1) each vector in the set is normalized and (2) two different vectors are orthogonal to each other. We summarize this definition by the following compact notation

$$\langle\alpha_i|\alpha_j\rangle = \delta_{ij} \quad .$$

Here, δ_{ij} is the Kronecker delta, i.e., it is 1 if $i = j$ and 0 otherwise. You can see that an orthonormal set is necessarily linearly independent.

A set of vectors $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is called an **orthonormal basis** if the set is a basis and the set is orthonormal. We deal with such orthonormal bases frequently in QM. There are two reasons for why they are important. First, such bases appear naturally in QM, they appear as eigenstates of an experimentally measurable quantity and each vector can be given a clear physical meaning. Second reason is a mathematical one. It is much more easier to handle the orthonormal bases than the non-orthonormal ones. The treatment below concentrates on the mathematical convenience.

First, since $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is a basis, we can expand any vector $|\psi\rangle$ as a superposition of these vectors:

$$|\psi\rangle = \sum_{i=1}^m b_i |\alpha_i\rangle \quad . \quad (1)$$

Next, if we take the inner product of this expression with one of $|\alpha_j\rangle$ we get

$$\langle\alpha_j|\psi\rangle = \sum_{i=1}^m b_i \langle\alpha_j|\alpha_i\rangle = \sum_{i=1}^m b_i \delta_{ij} = b_j \quad .$$

Therefore the expansion coefficients b_j can be computed easily by using the inner product.

$$b_j = \langle\alpha_j|\psi\rangle \quad . \quad (2)$$

It is sometimes convenient to join the equations (1) and (2) in a single expression

$$|\psi\rangle = \sum_{i=1}^m \langle\alpha_i|\psi\rangle |\alpha_i\rangle = \sum_{i=1}^m |\alpha_i\rangle \langle\alpha_i|\psi\rangle \quad . \quad (3)$$

We will see later that the last form is particularly nice in Dirac notation.

Next, let us consider two arbitrary vectors $|\psi\rangle$ and $|\phi\rangle$ which have the following expansions in the orthonormal basis $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$.

$$|\psi\rangle = \sum_{i=1}^m b_i |\alpha_i\rangle \quad , \quad |\phi\rangle = \sum_{i=1}^m c_i |\alpha_i\rangle \quad .$$

In that case, the inner product of these two vectors can be expressed simply in terms of the expansion coefficients as follows

$$\langle \phi | \psi \rangle = \sum_{i=1}^m c_i^* b_i \quad .$$

In particular,

$$\langle \psi | \psi \rangle = \sum_{i=1}^m |b_i|^2 \quad .$$

What is nice about these expressions is that, when we represent these abstract vectors as $m \times 1$ column vectors, the inner product of the vectors coincides with the inner product of column vectors. In other words,

$$|\psi\rangle \rightarrow u_\psi = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad |\phi\rangle \rightarrow u_\phi = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix},$$

then $\langle \phi | \psi \rangle = \langle u_\phi | u_\psi \rangle = u_\phi^\dagger u_\psi$. This makes the matrix representation attractive. Suppose you have an m -dimensional Hilbert space, constructed from vectors with a very complicated mathematical structure. In some cases, you may not even know what these vectors are. But in all cases, you can represent these vectors as $m \times 1$ column vectors. Moreover, all operations you can carry out with these vectors can be done simply with the corresponding matrices. Note also that this is true independent of how the orthonormal basis $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots, |\alpha_m\rangle\}$ is chosen. Of course, if you choose a different orthonormal basis, you will get a different representation (different column vectors for the same vector $|\psi\rangle$), but as long as you use the same basis consistently, the representation will work.

Example: Show that the following three vectors

$$\begin{aligned} \vec{A} &= (\hat{\mathbf{i}} + \hat{\mathbf{j}} + \hat{\mathbf{k}})/\sqrt{3} \quad , \\ \vec{B} &= (\hat{\mathbf{i}} - \hat{\mathbf{j}})/\sqrt{2} \quad , \\ \vec{C} &= (\hat{\mathbf{i}} + \hat{\mathbf{j}} - 2\hat{\mathbf{k}})/\sqrt{6} \quad , \end{aligned}$$

form an orthonormal basis for \mathbb{R}^3 . Expand the vector $\vec{P} = 2\hat{\mathbf{i}} + \hat{\mathbf{j}} - \hat{\mathbf{k}}$ in that basis.

Solution: It is easy to check that each vector is normalized and different vectors are orthogonal. We have an orthonormal set with three vectors. For this reason, they are linearly independent. Any three linearly-independent vectors form a basis for \mathbb{R}^3 since the dimension of this space is 3.

For the expansion of \vec{P} , let

$$\vec{P} = a\vec{A} + b\vec{B} + c\vec{C} \quad .$$

Then we have

$$\begin{aligned} a &= \vec{A} \cdot \vec{P} = 2/\sqrt{3} \quad , \\ b &= \vec{B} \cdot \vec{P} = 1/\sqrt{2} \quad , \\ c &= \vec{C} \cdot \vec{P} = 5/\sqrt{6} \quad . \end{aligned}$$

Therefore,

$$\vec{P} = \frac{2}{\sqrt{3}}\vec{A} + \frac{1}{\sqrt{2}}\vec{B} + \frac{5}{\sqrt{6}}\vec{C} \quad .$$

Check that this expression indeed gives the same \vec{P} . Note also that $\|\vec{P}\|^2$ can be computed by using either basis which gives the same results:

$$\vec{P} \cdot \vec{P} = P_x^2 + P_y^2 + P_z^2 = a^2 + b^2 + c^2 = 6 \quad .$$

Completeness

Completeness of Hilbert spaces basically tells you that a sequence of vectors which looks like converging converges to an actual vector (you are not supposed to understand anything from this sentence). As a rule, all finite-dimensional vector spaces with an inner product are complete. Therefore, these are Hilbert spaces. The complication arises only when the dimensional is infinite.

For us, the following definition of completeness will be sufficient. Let the infinite set $\{|\alpha_1\rangle, |\alpha_2\rangle, \dots\}$ be an orthonormal basis for an infinite dimensional Hilbert space. Consider the following series

$$|\psi\rangle = \sum_{i=1}^{\infty} b_i |\alpha_i\rangle \quad . \quad (4)$$

Then this series converges to a finite vector $|\psi\rangle$ if and only if the series

$$\sum_{i=1}^{\infty} |b_i|^2 \quad (5)$$

is convergent (i.e., it does not diverge to infinity). In that case, the norm of $|\psi\rangle$ is given as

$$\|\psi\|^2 = \langle\psi|\psi\rangle = \sum_{i=1}^{\infty} |b_i|^2 \quad .$$

Moreover, you can manipulate the series of the form (4) just like the way you manipulate them for finite dimensional case.

Completeness basically tells you that you can form *infinite superpositions* with the condition that (5) is convergent. The condition is equivalent to saying that the result of the series is not a vector with infinite length. However, physicists are reputed for not being bound by such restrictions. In some cases, we find it easier to work with vectors with infinite length (position and momentum eigenstates for example). Although mathematically incorrect, manipulation of such infinite vectors does not cause much problem if they are handled appropriately. Moreover, such infinite vectors are met only in intermediate stages of the calculations. For realistic situations we always have finite vectors. For example when $|\psi\rangle$ in (4) represents the state of a physical system, it has to be normalized, so the convergence is necessary in such cases.

States of a quantum mechanical system

This is the first postulate of quantum mechanics. For any physical system, we claim the existence of a certain Hilbert space such that

- (1) Every physical state of the system is represented by a normalized vector (say $|\psi\rangle$ with $\|\psi\| = 1$). For this reason, such normalized vectors are also called **states**, although they are actually just mathematical representations of the states.
- (2) Two normalized vectors $|\psi\rangle$ and $|\psi'\rangle$ that differ by an *overall phase factor* (i.e., $|\psi\rangle = e^{i\theta} |\psi'\rangle$ for some angle θ) represent the same physical state.
- (3) If two normalized vectors $|\psi\rangle$ and $|\phi\rangle$ are not parallel, they represent different physical states. (i.e., some physical property of the system is different in these states.)

Remember that by the word “state” we imply a mathematical structure that contains all of the information about the actual state. So, the normalized vectors of the Hilbert space somehow contain this information. But, we have not yet mentioned about how the physical properties of the system can be computed by using these. To do that, we need to introduce the operators on the Hilbert space and associate them with the observables. Until that time, it is sufficient to know that these vectors somehow contain this information.

The rule 2 actually says that parallel vectors represent the same state ($|\psi\rangle$ and $|\phi\rangle$ are parallel if $|\psi\rangle = c|\phi\rangle$ for some complex number c). But, since we require the state vectors to be normalized, parallelism reduces to an overall phase difference (c is a complex number with modulus 1.) Therefore, parallel vectors somehow yield same physical properties.

The state of the system can change with time. As the state at a given time contains all information about what would the state be after some time, there should also be a rule enabling us to compute the state at an arbitrary time t , once an initial state is given. In other words, we should have a relationship between $|\psi(t)\rangle$ and $|\psi(t=0)\rangle$ where $|\psi(t)\rangle$ represents the state at time t . To do this, we again need operators. As a result, that part of the first postulate will be covered later.

By “system” we sometimes imply certain logical parts of some bigger systems. For example, to describe only the spin degree of freedom of an electron, we need a 2-dimensional Hilbert space. So, normalized vectors in that space give us all possible spin-states of an electron, but does not give us any information about the “positional state” of the electrons. To describe only the positional state of an electron, we need an infinite dimensional Hilbert space, which turns out to be square-integrable complex-valued functions of position. Of course, to describe the state of an electron, we need to specify both the spin state and positional state, so a larger Hilbert space is needed which somehow contains the two described above (more on this later). Similarly, a 2-dimensional Hilbert space is needed for polarization states of a photon etc.

Polarization states of a photon

Here we give an example of a physical system and its associated Hilbert space that contains the states. For the polarization states of a photon, the associated state vectors contain elements which can be interpreted easily. This is just an example to show you how the state

vector actually contains all the information about the actual physical state. (For similar interpretations on other systems, we need to utilize the observables, so we need to wait for these.)

Consider a classical electromagnetic wave which is completely polarized. The polarization (i.e., the direction of the electric field) might be plane, circular or elliptical; we consider all possibilities in here. We will suppose that the wave is a plane wave propagating along $+z$ axis, has wavenumber k and angular frequency ω . (As we are not interested in the “positional state” of photons, the generality of our conclusions will not be lost by these assumptions.) As the electric field must be perpendicular to the propagation direction, $\vec{E}(\vec{r}, t)$ will be pointing in some direction on the xy -plane. Using the fact that the wave is plane, we can write the most general expression for the components of the electric field as

$$\begin{aligned} E_x(\vec{r}, t) &= E_{0x} \cos(kz - \omega t + \phi_1) \quad , \\ E_y(\vec{r}, t) &= E_{0y} \cos(kz - \omega t + \phi_2) \quad , \\ E_z(\vec{r}, t) &= 0 \quad . \end{aligned}$$

Here E_{0x} and E_{0y} are the maximum values (amplitude) of the corresponding components of the E -field, ϕ_1 and ϕ_2 are some angles representing the phase shifts of the x and y components of the wave. The fact that we have used \cos instead of \sin does not change the final results, so this is unimportant.

Now, consider a single photon that this classical wave contains. We represent the *polarization state* of such a photon, by the following 2×1 column matrix with complex entries,

$$\psi = \frac{1}{\sqrt{E_{0x}^2 + E_{0y}^2}} \begin{bmatrix} E_{0x} e^{i\phi_1} \\ E_{0y} e^{i\phi_2} \end{bmatrix} .$$

A few remarks are needed in here:

- (1) We have called this ψ , because this is the actual quantum mechanical state used for describing the polarization state of photons.
- (2) The prefactor with the square-root is introduced to produce a normalized vector, i.e., $\langle \psi | \psi \rangle = 1$. But, note that the prefactor also eliminates the unnecessary information contained in the magnitude of the E -field. The magnitude of the E -field is basically related to the total energy contained in the wave and has nothing to do with polarization. For example, if the E -field is doubled, i.e., $E_{0x} \rightarrow 2E_{0x}$ and $E_{0y} \rightarrow 2E_{0y}$, then the total energy carried by the wave increases 4 times (equivalently, number of photons increases 4 times). But in either case, all photons have the same polarization and therefore should have same ψ . The prefactor exactly achieves that.
- (3) Note that a uniform shift of both phase angles changes only the overall phase factor of ψ . For example, if ϕ_1 and ϕ_2 are both increased by θ , i.e., $\phi_1 \rightarrow \phi_1 + \theta$ and $\phi_2 \rightarrow \phi_2 + \theta$ then ψ changes by $\psi \rightarrow e^{i\theta} \psi$. By the postulates of QM , both ψ and $e^{i\theta} \psi$ should represent the same state (even though they are mathematically different vectors). On the physical side, we can easily convince ourselves that a uniform shift of both phase angles does not

change the polarization of the wave (same polarization whatever the value of θ is). For example, such a uniform shift can be achieved by relocating the origin (where $z = 0$ is) or by redefining the initial time (when $t = 0$ is). The place of origin and the time $t = 0$ depends on the observer and has nothing to do with the actual polarization of the wave. Therefore, in such situations we have the same physical state (polarization) but different mathematical states (ψ and $e^{i\theta}\psi$) which differ only by an overall phase factor.

- (4) Therefore, the polarization of the wave can be described by giving the value of two parameters: The ratio of x and y amplitudes, E_{0x}/E_{0y} and the relative phase angle between these directions, $\phi_2 - \phi_1$. The complex vector ψ contains this information as well. It has two complex components, which means that it contains 4 real parameters. But the restriction that ψ has to be normalized reduces this count to 3. Finally, the overall phase angle does not contain an essential information, so the count reduces to 2 real parameters.
- (5) In the specific case of plane polarization, the electric field points along a specific direction (or opposite to it) at every point and at every time. In this case, both phase angles can be chosen equal: $\phi_1 = \phi_2$ (both components of E -field are in phase). For example, if the polarization direction is along $\hat{\mathbf{n}} = \cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}$, then $E_{0x} = E_0 \cos\theta$ and $E_{0y} = E_0 \sin\theta$ where E_0 is the amplitude of E -field along $\hat{\mathbf{n}}$. In such a case the state is

$$\psi_{\hat{\mathbf{n}}} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} ,$$

perhaps multiplied with an overall phase factor. It is nice to see that the components of $\psi_{\hat{\mathbf{n}}}$ contains the components of the polarization direction $\hat{\mathbf{n}}$. We can read the polarization direction easily by just looking at the state vector. Note also that if θ is increased by 180° , i.e., $\theta \rightarrow \theta + \pi$, then $\hat{\mathbf{n}}$ changes to $-\hat{\mathbf{n}}$. But this cannot change the polarization as the electric field oscillates along $+\hat{\mathbf{n}}$ and $-\hat{\mathbf{n}}$. You can see that such a change in θ changes $\psi_{\hat{\mathbf{n}}}$ only by an overall phase factor.

- (6) The two special cases are the polarization along- x called horizontal polarization (H for short) and the polarization along- y which is called vertical polarization (V for short). Their respective state vectors are given by

$$\psi_H = \begin{bmatrix} 1 \\ 0 \end{bmatrix} , \quad \psi_V = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

These states form a basis (in fact an orthonormal basis) for the Hilbert space. Therefore, all vectors can be written as a superposition of these two vectors. In the particular case of $\psi_{\hat{\mathbf{n}}}$ we have

$$\psi_{\hat{\mathbf{n}}} = \cos\theta\psi_H + \sin\theta\psi_V .$$

- (7) The other special polarization type is the circular polarization. In that case we have $E_{0x} = E_{0y}$ (E -field has same amplitude in these two, in fact all, directions) but the phase angles differ by 90° . Depending on which component lags behind, we get the right circular (R for short) and left circular (L for short) polarizations. The corresponding state vectors are

$$\psi_L = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} , \quad \psi_R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} .$$

These two vectors also form an orthonormal basis. Therefore any vector can also be expanded in this basis. In particular $\psi_{\hat{\mathbf{n}}}$ can be expanded as

$$\psi_{\hat{\mathbf{n}}} = \frac{e^{-i\theta}}{\sqrt{2}}\psi_L + \frac{e^{+i\theta}}{\sqrt{2}}\psi_R \quad .$$

The need to expand a given state in different orthonormal bases arise frequently. Basically, an orthonormal basis is associated with a certain measurement type you carry out on a system.

- (8) All other states which cannot be represented as in (5) or (7) are called elliptically polarized. In such states, the physical electric field rotates on the xy plane (not plane polarized) as the time changes or the position is varied. Moreover, the amplitude of the field along different directions varies with direction. The tip of the electric field does not draw a circle, but draws an ellipse (not circularly polarized). For a given state vector of the form,

$$\psi = \begin{bmatrix} a \\ b \end{bmatrix} \quad ,$$

where a and b are complex numbers with $|a|^2 + |b|^2 = 1$, the behavior of the E -field as a function of position and time can be analyzed by using

$$\begin{aligned} E_x(\vec{r}, t) &= E_0 \operatorname{Re} (ae^{i(kz-\omega t)}) \quad , \\ E_y(\vec{r}, t) &= E_0 \operatorname{Re} (be^{i(kz-\omega t)}) \quad , \end{aligned}$$

where E_0 is an E -field amplitude.

Superposition

Let us add a few final remarks on superpositions. These are valid for all quantum mechanical systems, but the actual example will be taken from the polarization state of photons. The state $\psi_{\hat{\mathbf{n}}} = \cos\theta\psi_H + \sin\theta\psi_V$ is a superposition of ψ_H and ψ_V (horizontal and vertical polarization states). The expansion coefficients in the superposition will be related to some experimental probabilities later. Although such superpositions are quite simple mathematically as they are vector sums (just like $\hat{\mathbf{n}} = \cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}}$), in the context of QM some wrong interpretations can be met frequently.

- * “The photon in state $\psi_{\hat{\mathbf{n}}}$ is either horizontally polarized or vertically polarized, but we don’t know which one” is a commonly met misinterpretation. I want to consider this as equivalent to saying that “the vector $\hat{\mathbf{n}}$ is either horizontal or vertical, but we don’t know which one” which is clearly nonsense. Although such interpretations are met in alternative interpretations of quantum mechanics, they are different theories. The states in these theories are also different from the ones in quantum mechanics: One should specify not only the normalized vector in the Hilbert space, but also some other parameters which contain some additional information about the state. They are theories completely different from QM.

* The correct quantum mechanical interpretation should be something like this: “The photon in state $\psi_{\hat{\mathbf{n}}}$ is both horizontally and vertically polarized at the same time.” It contains both polarizations in a peculiar superposition expressed above. The equivalent statement for vectors is “the vector $\hat{\mathbf{n}}$ is both horizontal or vertical, it is in a direction that contains both of these components.”

Although $\psi_{\hat{\mathbf{n}}}$ can be expressed as a superposition of ψ_H and ψ_V , it can also be expressed as a superposition in an alternative basis as well. Same interpretation is also valid for these cases as well. For example, in terms of the circularly polarized states,

$$\psi_{\hat{\mathbf{n}}} = \frac{e^{-i\theta}}{\sqrt{2}}\psi_L + \frac{e^{+i\theta}}{\sqrt{2}}\psi_R \quad .$$

Therefore $\psi_{\hat{\mathbf{n}}}$ is both right and left circularly polarized at the same time etc. The exact superposition will give you the complete information about the actual state. For example, all plane polarized states will be expressed as above, differing only in the value of θ . That angle appears as a *relative phase angle* in the expression above (relative phase is the difference of phase angles of expansion coefficients for ψ_L and ψ_R . In our case, it is 2θ). Such relative phase angles (as opposed to overall phase angles) contain some physical information (in this case, the exact polarization direction of plane polarized wave).

Consider the special case $\theta = 0$, which correspond to the horizontal polarization. The expression above yields

$$\begin{aligned} \psi_H &= \frac{1}{\sqrt{2}}\psi_L + \frac{1}{\sqrt{2}}\psi_R \\ &= \frac{1}{2} \begin{bmatrix} 1 \\ i \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad . \end{aligned}$$

Notice how the second component adds up to 0. Both left and right circularly polarized states contain vertical polarization (they are superpositions of H and V). But, when we take a particular superposition of L and R, the vertical component is eliminated. This phenomenon is known as *destructive interference*. Similarly H component adds up to 1 with *constructive interference*. Interference phenomenon is frequently met in waves, but we also meet them even in this simple example of a 2-dimensional Hilbert space. They appear as a result of *superposition*, which is also conventionally associated with waves.

In this particular case, one might be inclined to think that these phenomena arises from our association of the state ψ with the electromagnetic wave (vertical electric fields of L and R are completely out of phase, so they interfere destructively when superposed. Therefore, in ψ same should happen). However, such phenomena appear also in other physical systems, such as the spin of an electron, where we do not have a clear association of ψ with some physical wave. Interference phenomena appear in every quantum mechanical system, whatever the physical nature of the system is. They arise only from the superpositions of states (of these kets in the Hilbert space) and may not always have a clear explanation in terms of some physical waves.