The Mathematical Language of Quantum Mechanics

I. Probability and Statistics

We need to develop several concepts of probability theory in order to understand many aspects of quantum theory.

Let us imagine that we have a box which contains N balls, each marked with some number, which we denote generically by ν. In general, the same ν-number may appear on more than one ball. We let n_k be the number of balls on which there appears the particular ν-number ν_k . The box of balls is therefore described by two sets of numbers

$$\{\nu_i\} = \nu_1, \nu_2, \nu_3, \nu_4, \ldots\ldots\ldots\ldots \quad , \quad \{n_i\} = n_1, n_2, n_3, n_4, \ldots\ldots\ldots\ldots$$

Since the total number of balls is N we must have

$$\sum_k n_k = n_1 + n_2 + n_3 + n_4 + \ldots\ldots\ldots\ldots = N$$

Suppose that we select a ball at random from the box; what is the probability p_k that the selected ball will show the value ν_k? Since out of N possible selections, n_k of these would yield the ν-number ν_k, we conclude that

$$p_k = \frac{n_k}{N} \tag{01}$$

Thus, if $n_k = 0$ it would be **impossible** to select a ball showing ν_k , and we would have $p_k = 0$; on the other hand, if $n_k = N$ it would be an **absolute certainty** that the selected ball would show ν_k, and we would have $p_k = 1$. These results all agree with our **common sense** notions of probability. In general, the numbers $\{p_k\}$ satisfy the conditions

$$0 \le p_k \le 1 \quad \text{for all } k \tag{02a}$$

and

$$\sum_k p_k = 1 \tag{02b}$$

Proof: we have

$$p_k = \frac{n_k}{N} \quad , \quad \sum_k n_k = N \quad , \quad n_k \ge 0$$

$$or \quad 0 \le n_k \le N \rightarrow 0 \le p_k = \frac{n_k}{N} \le 1$$

$$\sum_k p_k = \sum_k \frac{n_k}{N} = \frac{\sum_k n_k}{N} = \frac{N}{N} = 1$$

Connection to quantum theory: As we mentioned earlier, it will turn out, however, that in the quantum world I can, after making a very **large** number of measurements on these "**identical**" systems, state a "**probability**" that the position will have a particular value.....that is, for the next measurement to be made of the position x, we have

$$probability(x) = \frac{\text{number of times value x was measured}}{\text{total number of measurements}}$$

This type of statement is one that is based on the results of **all previous measurements**. The goal of quantum mechanics is to predict these probabilities before the measurements are done.

This type of probability statement is one that is based on the results of **all previous measurements** and is fundamental to all of science. It is called the frequency model and is strictly valid only if the number if identical measurements tends to infinity.

Example:

Suppose we measure the heights of Swarthmore students and we find for $N = 1300$ = total number of measured heights(in cm)

n_h	h
50	150
100	160
200	170
300	180
300	190
200	200
100	210
50	220

where n_h = number of times height h was measured. Then the probability that we will measure $h = 190$ cm if another student (if we missed one) walks in is

$$probability(h = 190cm) = \frac{\text{number of times value 190 was measured}}{\text{total number of measurements}} = \frac{3}{13}(23\%)$$

This is a very simple and intuitive definition and it works very well (it is exact in the limit of total number = $N \to \infty$, **which we cannot really do in any real experiment**).

Gambling Probabilities

The probability of rolling a 2 when tossing a die is

$$p = \frac{\text{number of ways to roll a two}}{\text{total number of ways to roll anything}} = \frac{1}{6}$$

The probability of getting heads when flipping a coin is

$$p = \frac{\text{number of ways to flip a heads}}{\text{total number of ways to flip anything}} = \frac{1}{2}$$

In general,

$$p(success) = \frac{\text{number of successful outcomes}}{\text{number of possible outcomes}}$$

Does this rule apply to surgery where the only two possible outcomes are survival and death?

The answer clearly is no, i.e., we hope that $p(survival) >> 1/2$.

To use the rule for surgery, one must delineate a wider range of outcomes and then do the counting.

For example, consider a lottery. There are only two outcomes .... winning or losing, but the probability of winning is not $1/2$. In fact,

$$p(winning) = \frac{\text{number ways to win}}{\text{number of possible tickets}} = \frac{1}{\#\ combinations}$$

Suppose we are to pick the correct 6 numbers out of a selection of 40 numbers(no repeats). We buy 1 ticket. The number of possible combinations, however, is

40(39)(38)(37)(36)(35)=2,763,633,600 !!!!!

So the probability of winning is so small that you have a larger chance of being hit by a meteorite!

Now let us calculate the probability that a single random selection from the box will yield a ball showing **either** ν_k **or** ν_j. Since out of N possible selections, a total of $(n_k + n_j)$ would yield one of these ν-numbers, we conclude that

$$p(\nu_k \ or \ \nu_j) = \frac{n_k + n_j}{N} = p_j + p_k \qquad (03a)$$

This allows us to interpret Eq.(02b) as simply stating that it is an absolute certainty that a randomly selected ball will show **some** ν_k number, i.e., something must happen!

Example: Toss a die. What is the probability of rolling either a 1 **or** a 3?

In this case, there are six possible outcomes and two successful outcomes. We then get

$$p = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

In general, the word **or** is a signal to **add** probabilities.

Suppose we now make **two** random selections, taking care to return the first selected ball back to the box before making the second selection (thus it is possible to pick the same ball both times). What is the probability that the first ball will show the value ν_k **and** the second ball will show the value ν_j?

There are n_k ways to select a ν_k-ball, and for each of these ways there are n_j ways to select a ν_j-ball; thus, there are a total of $n_k \cdot n_j$ ways to select a ν_k-ball and then a ν_j-ball. There are, however, N

possible selections for the first ball and for each of these there are N possible selections for the second selections; thus, there are a total of $N \cdot N$ possible double selections. We conclude then that

$$p(\nu_k \text{ and } \nu_j) = \frac{n_k \cdot n_j}{N \cdot N} = p_j \cdot p_k \qquad \text{(03b)}$$

Eqs.(03a) and (03b) will form the basis for almost all of our considerations involving probability theory.

Example: Toss a die **and** simultaneously flip a coin. What is the probability of getting a 2 **and** a tails?

In this case, there are 12 possible outcomes

1H,2H,3H,4H,5H,6H,1T,2T,3T,4T,5T,6T

The probability of success here is then

$$p = \frac{1}{6} x \frac{1}{2} = \frac{1}{12}$$

In general, the word **and** is a signal to **multiply** probabilities.

Example: Flip 3 coins(or flip 1 coin 3 times).

Outcome Table

possible outcomes	probability of this outcome	number of heads
HHH	1/8	3
HHT	1/8	2
HTH	1/8	2
THH	1/8	2
HTT	1/8	1
THT	1/8	1
TTH	1/8	1
TTT	1/8	0

Therefore

$$p(3\,heads) = \frac{1}{8}\ ,\ p(2\,heads) = \frac{3}{8}\ ,\ p(1\,head) = \frac{3}{8}\ ,\ p(0\,heads) = \frac{1}{8}$$

Note that the sum of all terms must = 1, i.e., the probability of something happening must = 1, which it does.

Example: Toss 2 dice(or toss 1 die 2 times). What is the probability that the sum of the face-up dots is 4?

If the first die lands on 4 and the second on 3 we will denote the outcome [4,3].

$$p([1,1]) = \frac{1}{6}x\frac{1}{6} = \frac{1}{36} = p(2)\,x\,p(2)$$

or

$$p([n,m]) = \frac{1}{6}x\frac{1}{6} = \frac{1}{36} = p(n)\,x\,p(m) \quad , \quad n,m = 1,2,3,4,5,6$$

The we have

$$p(\text{sum is } 4) = p([2,2]) + p([1,3]) + p([3,1]) = \frac{1}{12}$$

Suppose now that we subject our box of N balls to M samplings, that is, we select a ball at random from the box, record its ν-number and return it to the box a total of M times. We denote by $\nu^{(i)}$ the ν-value recorded on the i^{th} sampling, and we make the following two definitions:

The **mean** or **average** value of the ν-values recorded is

$$\langle \nu \rangle = \frac{\sum_{i=1}^{M} \nu^{(i)}}{M} \tag{04}$$

which is the standard definition of the average or "best value" of a series of measurements. One must be careful not to imply that $\langle \nu \rangle$ has any truth or legitimacy beyond that of any of the individual $\nu^{(i)}$ values.

The **root-mean-square** (or rms) **deviation** of these values is

$$\Delta\nu = \sqrt{\frac{\sum_{i=1}^{M} \left(\nu^{(i)} - \langle \nu \rangle\right)^2}{M}} \tag{05}$$

To calculate this quantity, we must first calculate the deviation from the mean $\nu^{(i)} - \langle \nu \rangle$, of each ν-number obtained; we next compute the average of the squares of these deviations (the squares taken to keep the positive and negative deviations from cancelling each other) and finally, to counteract, to some extent, the squaring, we take the square root of this average. Thus $\Delta\nu$ is the **square root** of the **mean** of the **squares** of the **deviations** of the $\nu^{(i)}$ values from $\langle \nu \rangle$. This quantity is also called the **rms dispersion**, since it clearly measures the extent to which the $\nu^{(i)}$ values are dispersed about $\langle \nu \rangle$.

We can rewrite this expression in a more useful form that is easier to calculate. We have

$$(\Delta\nu)^2 = \frac{\sum_{i=1}^{M} \left(\nu^{(i)} - \langle \nu \rangle\right)^2}{M} = \frac{\sum_{i=1}^{M} \left(\left(\nu^{(i)}\right)^2 - 2\langle \nu \rangle \nu^{(i)} + \langle \nu \rangle^2\right)}{M} = \frac{\sum_{i=1}^{M} \left(\nu^{(i)}\right)^2}{M} - 2\langle \nu \rangle \frac{\sum_{i=1}^{M} \nu^{(i)}}{M} + \langle \nu \rangle^2 \frac{\sum_{i=1}^{M} 1}{M}$$

$$= \left\langle \nu^2 \right\rangle - 2\langle \nu \rangle\langle \nu \rangle + \langle \nu \rangle^2 \frac{M}{M} = \left\langle \nu^2 \right\rangle - \langle \nu \rangle^2$$

or

$$\Delta v = \sqrt{\langle v^2 \rangle - \langle v \rangle^2} \qquad (06)$$

In words, the rms deviation of the $v^{(i)}$ values is equal to the square root of the difference between the **average of the square** and the **square of the average**. We note that $\langle v^2 \rangle = \langle v \rangle^2$ **only if** every $v^{(i)}$ value coincides with $\langle v \rangle$ so that there is no dispersion or no deviation from the mean.

If we have knowledge of the two sets of numbers $\{v_k\}$ and $\{p_k\}$, it would seem that we ought to be able to **predict** approximately what values would be obtained for $\langle v \rangle$ and Δv. The key to making such a prediction is the following assumption: since n_k of the N balls have the number v_k, then in M random samplings of these balls we ought to obtain the value v_k approximately m_k times where

$$\frac{m_k}{M} = \frac{n_k}{N}$$

Using Eq.(01) we then find the approximate number of times the value v_k should appear in the set of values $v^{(1)}, v^{(2)}, \ldots\ldots, v^{(M)}$ is

$$m_k = \frac{n_k}{N} M = p_k M$$

With this result, the sum in Eq.(04) can be written

$$\langle v \rangle = \frac{1}{M} \sum_{i=1}^{M} v^{(i)} = \frac{1}{M} \sum_k m_k v_k = \frac{1}{M} \sum_k (p_k M) v_k = \sum_k p_k v_k \qquad (07)$$

Eq.(07) expresses $\langle v \rangle$ as a "weighted sum" of the possible v_k values; the weight assigned to any particular value v_k is just the probability of its occurrence p_k . This value is the "theoretically expected" value; the "experimental" value in Eq.(04) will generally differ somewhat from this theoretical value owing to the randomness involved. However, in the limit of very many experimental samplings $(M \to \infty)$, the value in Eq.(04) may be expected to get arbitrarily closes to the value in Eq.(07), that is, the rms deviation from $\langle v \rangle$ in Eq.(07) should approach zero.

Equation (07) may be generalized quite easily as shown in the following exercise.

Let f be a given function of v, and let this function be evaluated for each of the $v^{(i)}$-values. The average or mean of the resulting set of $f(v^{(i)})$-values is

$$\langle f(v) \rangle = \sum_k p_k f(v_k) \qquad (08)$$

We note that by putting $f(v) = v$ we get Eq.(07).

Proof: we have

$$\langle f(v)\rangle = \frac{1}{M}\sum_{i=1}^{M} f(v^{(i)}) = \frac{1}{M}\sum_{i=1}^{M} m_k f(v_k) = \frac{1}{M}\sum_{i=1}^{M} (p_k M) f(v_k) = \sum_{i=1}^{M} p_k f(v_k)$$

By putting $f(v) = v^2$ in Eq.(08) we see that

$$\langle v^2 \rangle = \sum_k p_k v_k^2$$

Using this result and Eq.(07), we can then write Eq.(06) as

$$\Delta v = \sqrt{\left(\sum_k p_k v_k^2\right) - \left(\sum_k p_k v_k\right)^2} \qquad (09)$$

We now see that Eqs.(07) and (09) express the two basic quantities $\langle v \rangle$ and Δv wholly in terms of the numbers $\{v_k\}$ and $\{p_k\}$. Thus, given a set of values $\{v_k\}$ distributed with probabilities $\{p_k\}$ Eqs.(07) and (09) allows us to calculate the theoretically expected **mean** and **rms deviation** to be obtained by any random sampling of these v-values.

In terms of this definition of the average value for a set of measurements, the average height of our earlier sample of heights is given by

$$\langle height \rangle = \langle h \rangle = \frac{1}{L}\sum_{j=1}^{L} H_j$$

and we get

$$\begin{aligned}\langle height \rangle &= \langle h \rangle \\ &= \frac{1}{1300}\left(\begin{array}{r} 50x150 + 100x160 + 200x170 + 300x180 + 300x190 \\ + 200x200 + 100x210 + 50x220 \end{array}\right) \\ &= \frac{240500}{1300} = 185.0\end{aligned}$$

or

$$\langle height \rangle = \langle h \rangle = \frac{1}{N}\sum_h n_h h = \sum_h h \frac{n_h}{N} = \sum_h h \cdot prob(h)$$

and

$$\begin{aligned}\langle (height)^2 \rangle &= \langle h^2 \rangle \\ &= \frac{1}{1300}\left(\begin{array}{r} 50x150^2 + 100x160^2 + 200x170^2 + 300x180^2 + 300x190^2 \\ + 200x200^2 + 100x210^2 + 50x220^2 \end{array}\right) \\ &= \frac{44845000}{1300} = 34496.2\end{aligned}$$

Finally,

$$\begin{aligned}(\Delta h)^2 &= \langle h^2 \rangle - \langle h \rangle^2 = 271.2 \\ \Delta h &= 16.5\end{aligned}$$

Meaning of the Standard Deviation

For normal experimental data with random errors, the standard deviation has the following meaning. If we do a set of N measurements and obtain a mean value $\bar{x}$ and a standard deviation σ then

(1) the probability is 0.68 that any subsequent measurement of the average value, say $\bar{x}_1$ lies in the range $\bar{x} \pm \sigma_m$, where $\sigma_m = \sigma / \sqrt{N}$

(2) the interval $\bar{x} \pm r$, where $r = 0.67\sigma_m = \text{probable error}$, is such that the probability is 1/2 that a new measurement of the mean value would lie in this interval (and so also that the probability is 1/2 that it would lie outside the interval), that is, a 50% **confidence level**.

Exercise 4 illustrates the overall significance of $\langle \nu \rangle$ and $\Delta\nu$. Certainly a **complete** description of the expected results of a "multiple sampling" experiment requires the specification of **all** the numbers $(\nu_1, p_1), (\nu_2, p_2), (\nu_3, p_3), \ldots\ldots$. However, if we are asked to describe the results with **only two** numbers, we would evidently do well to state the values $\langle \nu \rangle$ and $\Delta\nu$: $\langle \nu \rangle$ is essentially a "collective value" for the set of ν-numbers, while $\Delta\nu$ (or the smallness thereof) provides a quantitative measure of the degree to which it is actually meaningful to so characterize the set of ν-values by the **single** value $\langle \nu \rangle$.

Probability Concepts (more details)

Quantum mechanics will necessarily involve probability in order for us to make the connection with experimental measurements.

We will be interested in understanding the quantity

$P(A|B)$ = probability of event A given that event B is true

In essence, event B sets up the conditions or an environment and then we ask about the (conditional) probability of event A given that those conditions exist. The | symbol means "given" so that items to the right of this "conditioning" symbol are taken as being true.

In other words, we set up an experimental apparatus, which is expressed by properties B and do a measurement with that apparatus, which is expressed by properties A. We generate numbers (measurements) which we use to give a value to the quantity $P(A|B)$.

We start with the **standard** mathematical formalism based on axioms. We define these events

A = *occurrence of* A (denotes that proposition A is true)

$\sim A$ = *NOT* A = *nonoccurrence of* A (denotes that proposition A is false)

$A \,\&\, B = A$ *AND* B = *occurrence of* both A *and* B (denotes proposition A and B is true)

$A \vee B = A$ *OR* B = *occurrence of* at least one of the events A *and* B

(denotes proposition A or B is true)

and standard Boolean logic as shown below:

Boolean logic uses the basic statements AND, OR, and NOT. Using these and a series of Boolean expressions, the final output would be one TRUE or FALSE statement.

This is illustrated below:

If A is true AND B is true, then (A AND B) is true

If A is true AND B is false, then (A AND B) is false

If A is true OR B is false, then (A OR B) is true

If A is false OR B is false, then (A OR B) is false

or written as a "truth" table:

A	B	$(A \wedge B))$	$(A \vee B)$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	0	0

where $1 = TRUE, 0 = FALSE$.

We then set up a theory of probability with these axioms:

(1) $P(A|A) = 1$

This is the probability of the occurrence A given the occurrence of A. This represents a **certainty** and, thus, the probability must = 1. This is clearly an obvious assumption that we must make if our probability ideas are to make any sense at all.

In other words, if I set the experimental apparatus such that the meter reads A, then it reads A with probability = 1.

(2) $0 \le P(A|B) \le P(B|B) = 1$

This just expresses the sensible idea that no probability is greater than the probability of a certainty and it make no sense to have the probability be less than 0.

(3) $P(A|B) + P(\sim A|B) = 1 \quad or \quad P(\sim A|B) = 1 - P(A|B)$

This just expresses the fact that the probability of something (anything) happening ($A \; or \; \sim A$) given B is a certainty (=1), that is, since the set $A \; or \; \sim A$ includes everything that can happen, the total probability that one or the other occurs must be the probability of a certainty and be equal to one.

(4) $P(A \& B|C) = P(A|C)P(B|A \& C)$

This says that the probability that 2 events A, B both occur given that C occurs equals the probability of A given C multiplied by the

probability of B given (A&C), which makes sense if you think of them happening **in sequence**.

All other probability relationships can be derived from these axioms.

The nonoccurrence of A given that A occurs must have probability = 0. This is expressed by

$$P(\sim A \mid A) = 0$$

This result clearly follows from the axioms since

$$P(A \mid B) + P(\sim A \mid B) = 1$$
$$P(A \mid A) + P(\sim A \mid A) = 1$$
$$P(\sim A \mid A) = 1 - P(A \mid A) = 1 - 1 = 0$$

Example: Let us evaluate $P(X \& Y \mid C) + P(X \& \sim Y \mid C)$.

We use axiom (4) in the 1st term with $A = X$, $B = Y$ *and* $C = C$ and in the 2nd term with $A = X$, $B = \sim Y$ *and* $C = C$ to get

$$P(X \& Y \mid C) + P(X \& \sim Y \mid C) = P(X \mid C)P(Y \mid X \& C) + P(X \mid C)P(\sim Y \mid X \& C)$$
$$= P(X \mid C)\left[P(Y \mid X \& C) + P(\sim Y \mid X \& C)\right] = P(X \mid C)[1] \quad \text{using axiom (3)}$$

and finally

$$P(X \& Y \mid C) + P(X \& \sim Y \mid C) = P(X \mid C)$$

which says probability that X is true regardless of whether Y is true, is the sum of the probabilities of X and Y for all possibilities associated with Y (Y and $\sim Y$ in this case).

Now let us use this result with $X = \sim A$, $Y = \sim B$. This gives

$$P(\sim A \& \sim B \mid C) = P(\sim A \mid C) - P(\sim A \& B \mid C) = 1 - P(A \mid C) - P(\sim A \& B \mid C)$$

Expanding the last term using $X = B$, $Y = A$ we then have

$$P(B \& \sim A \mid C) + P(B \& A \mid C) = P(B \mid C)$$

or

$$P(\sim A \& B \mid C) = P(B \mid C) - P(B \& A \mid C)$$

which gives

$$P(\sim A \& \sim B \mid C) = 1 - P(A \mid C) - P(B \mid C) + P(A \& B \mid C)$$

Now

$$P(A \vee B) = 1 - P(\sim (A \vee B) \mid C) = 1 - P(\sim A \& \sim B) \mid C)$$

and since

$$(\sim (A \vee B)) = (\sim A \& \sim B)$$

i.e., we can construct a 'truth table" as shown below, which illustrates the equality directly

A	B	$(\sim(A \vee B))$	$(\sim A \& \sim B)$	
1	1	0	0	
1	0	0	0	(this is the "truth table")
0	1	0	0	
0	0	1	1	

we finally get

$$P(A \vee B) = P(A|C) + P(B|C) - P(A \& B|C)$$

This is a very important and useful result.

If we have $P(A \& B|C) = 0$, then events A and B are said to be **mutually exclusive** given that C is true and the relation then reduces to

$$P(A \vee B) = P(A|C) + P(B|C)$$

This is the rule of **addition of probabilities for exclusive events**.

Some other important results are:

$$If\ \ A \& B = B \& A\ ,\ then\ \ \mathrm{P(A|C)P(B|A \& C) = P(B|C)P(A|B \& C)}$$

$$If\ \ P(A|C) \neq 0\ ,\ \ then\ \ \ P(B|A \& C) = P(A|B \& C)\frac{P(B|C)}{P(A|C)}$$

which is **Baye's theorem**. It relates the probability of B given A to the probability of A given B.

When we say that B is **independent** of A, we will mean

$$P(B|A \& C) = P(B|C)$$

or the occurrence of A has **NO influence** on the probability of B given C. Using axiom (4) we then have the result:

if A and B are independent given C,
then $P(A \& B|C) = P(A|C)P(B|C)$

This is called **statistical** or **stochastic** independence. The result generalizes to a set of events $\{A_i\ ,\ \mathrm{i} = 1,2,....,\mathrm{n}\}$. All these events are independent if and only if

$$P(A_1 \& A_2 \& .... \& A_m|C) = P(A_1|C)P(A_2|C).....P(A_m|C)$$

for all $m \leq n$.

Now let us think about these ideas in another way that has fundamental importance in modern approaches to quantum theory. The fundamental result in this view will turn out to be the Bayes formula and its relationship to measurements.

Suppose that we have an experimental measurement, M, that can yield either $A\ or\ \sim A$ as results, with a probability for result A given by

$$P(A \mid M) = p$$

In general, we let any sequence of n independent measurements be labelled as event M^n and we define n_A as the number of times A occurs, where $0 \le n_A \le n$.

Now imagine we carry out a sequence of n independent measurements and we find that A occurs r times. The probability for a sequence of results that includes result A r times and $\sim A$ $(n-r)$ times (independent of their order in the sequence) is given by

$$p^r q^{n-r}$$

where

$$q = P(\sim A \mid M) = 1 - P(A \mid M) = 1 - p$$

The different sequence orderings are mutually exclusive events and thus we have

$$P(n_A = r \mid M^n) = \sum_{\substack{all\ possible \\ orderings}} p^r q^{n-r}$$

The sum $\sum_{\substack{all\ possible \\ orderings}}$ just counts the number of ways to distribute r A's and $(n-r)$ $\sim A$'s, where all the terms contain the common factor $p^r q^{n-r}$. This result is given by the Binomial probability distribution as

$$\frac{n!}{r!(n-r)!}$$

so that

$$P(n_A = r \mid M^n) = \frac{n!}{r!(n-r)!} p^r q^{n-r}$$

These ideas will play a very important role in understanding basic concepts in quantum theory.

II. Complex Numbers

The equations of quantum theory will involve complex numbers. Let us review a few elementary properties of complex numbers.

Complex numbers involve the quantity i where

$$i^1 = i, i^2 = -1, i^3 = -i, i^4 = 1, i^5 = i, etc$$

You will notice that I never said that $i = \sqrt{-1}$ which is a nonsense statement mathematically since the square root function is **not defined** for negative numbers.

A **complex number** is then defined as $z = p + iq$ where p = real part of z = $\mathrm{Re}z$ and q = imaginary part of z = $\mathrm{Im}z$.

The **complex conjugate** of z is defined by $z^* = p - iq$

Complex arithmetic is straightforward.

If $z_1 = p + iq$ and $z_2 = r + is$ then

$$z_1 - z_2 = (p - r) + i(q - s)$$
$$z_1 z_2 = (p + iq,)(r + is) = (pr - qs) + i(qr + ps)$$

In addition, we define

$$\textit{absolute value of} \quad z = |z| = \sqrt{p^2 + q^2} \geq 0$$

or

$$\left|z^2\right| = p^2 + q^2 = (p + iq)(p - iq) = zz^* = z^* z$$

Various functions are defined by power series, i.e.,

$$e^{\alpha x} = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} x^n$$
$$\sin \alpha x = \sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n+1}}{(2n+1)!} x^{2n+1}$$
$$\cos \alpha x = \sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n}}{(2n)!} x^{2n}$$

These expansions are valid even if the parameter α is a complex number.

Complex Exponentials

We then have the very important relation

$$e^{i\alpha x} = \sum_{n=0}^{\infty} \frac{i^n \alpha^n}{n!} x^n = 1 + i\alpha x - \frac{\alpha^2}{2!} x^2 - i\frac{\alpha^3}{3!} x^3 + \frac{\alpha^4}{4!} x^4 + i\frac{\alpha^5}{5!} x^5 - .....$$
$$= \left(1 - \frac{\alpha^2}{2!} x^2 + \frac{\alpha^4}{4!} x^4 - .....\right) + \left(i\alpha x - i\frac{\alpha^3}{3!} x^3 + i\frac{\alpha^5}{5!} x^5 - .....\right)$$
$$= \cos \alpha x + i \sin \alpha x$$

This is called the **Euler relation**.

Using the Euler relation we have the important results

$$\sin \alpha x = \frac{e^{i\alpha x} - e^{-i\alpha x}}{2i} \quad \text{and} \quad \cos \alpha x = \frac{e^{i\alpha x} + e^{-i\alpha x}}{2}$$

The Euler relation allows us to define i in a better manner. We have

$$e^{i\pi} = \cos \pi + i \sin \pi = -1$$

We also have the properties

$$e^{a+b} = e^a e^b$$

$$e^a = e^{a/2}e^{a/2} \rightarrow e^{a/2} = \sqrt{e^a}$$

$$\left(e^a\right)^n = e^{na}$$

Therefore

$$\sqrt{e^{i\pi}} = e^{i\pi/2} = \cos\pi/2 + i\sin\pi/2 = i$$

Another useful relationship:

$$e^{\ell n(Q)} = Q$$

Now suppose we let

$$a = r\cos\theta \quad , \quad b = r\sin\theta$$
$$z = a + ib = r\cos\theta + ir\sin\theta$$

We can then write

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}$$

where

$$\sqrt{a^2 + b^2} = r = |z| \quad , \quad \tan\theta = \frac{b}{a}$$

Using these relations we then have

$$\ell nz = \ell nre^{i\theta} = \ell nr + \ell ne^{i\theta} = \ell nr + i\theta$$

Some Useful Mathematics

- **Taylor Series**

Functions can be expanded in power series. Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$. Then, we get by differentiation $f(0) = a_0$, $f'(0) = a_1$, $\frac{1}{2!}f''(0) = a_2$ and so on..... or in general $\frac{1}{n!}f^{(n)}(0) = a_n$. Therefore,

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0)x^n$$

which is called a **Maclaurin series** for $f(x)$ or a **Taylor series** for $f(x)$ about the origin.

A Taylor series, in general, means a power series in powers of $(x-a)$ where a = some constant. The derivation of the coefficients is identical to the last derivation except that we use $x = a$ instead of $x = 0$. Let

$$f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$$

Then, we get by differentiation $f(a) = a_0$, $f'(a) = a_1$, $\frac{1}{2!}f''(a) = a_2$ and so on.....

or in general $\frac{1}{n!} f^{(n)}(a) = a_n$. Therefore,

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)(x-a)^n$$

- **Binomial Series**

Now consider the following function $f(x) = (1+x)^n$. If we expand this as a Taylor series we get

$$f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) x^n$$

where

$$f(0) = 1 \quad , \quad f'(0) = n \quad , \quad \frac{1}{2!} f''(0) = n(n-1) \quad \text{and so on.....}$$

so that

$$f(x) = (1+x)^n = 1 + nx + \frac{n(n-1)}{2!} x^2 + ... = \sum_{m=0}^{\infty} \frac{n!}{m!(n-m)!} x^m = \sum_{m=0}^{\infty} \binom{n}{m} x^m$$

which is the **Binomial series**. For n = integer, the series terminates and we have an n^{th} order polynomial. The expression $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ is the Binomial coefficient.

Examples:

Taylor Series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ...$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ...$$

$$e^{\alpha x} = 1 + \frac{\alpha x}{1!} + \frac{(\alpha x)^2}{2!} + \frac{(\alpha x)^3}{3!} + ...$$

Binomial Series

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!} x^2 + ...$$

$$\frac{1}{(1 \pm x)^n} \approx 1 \mp nx \qquad x << 1$$

Complex Exponential

$$\begin{aligned} e^{i\alpha x} &= 1 + \frac{i\alpha x}{1!} + \frac{(i\alpha x)^2}{2!} + \frac{(i\alpha x)^3}{3!} + \frac{(i\alpha x)^4}{4!} ..... \\ &= 1 + \frac{i\alpha x}{1!} - \frac{(\alpha x)^2}{2!} - \frac{i(\alpha x)^3}{3!} + \frac{(\alpha x)^4}{4!} ..... \\ &= (1 - \frac{(\alpha x)^2}{2!} + \frac{(\alpha x)^4}{4!} .....) + i(\frac{\alpha x}{1!} - \frac{(\alpha x)^3}{3!} .....) \\ &= \cos \alpha x + i \sin \alpha x \end{aligned}$$

$$e^{\pm i\alpha x} = \cos\alpha x \pm i\sin\alpha x$$

$$\cos\alpha x = \frac{e^{+i\alpha x} + e^{-i\alpha x}}{2} \quad , \quad \sin\alpha x = \frac{e^{+i\alpha x} - e^{-i\alpha x}}{2i}$$

III. Hilbert Space Vectors and Dirac Language

There is an **algorithm**, named Quantum Mechanics, for predicting the behavior of physical systems. It correctly predicts all of the strange behaviors of the electron that we have already described in the fictitious world of **color and hardness**.

There is a standard way of interpreting this algorithm - a way of confronting the meaning of **superposition** - which comes from Neils Bohr and Max Born. It is called the **Copenhagen Interpretation**.

During this course, we will develop a language, which is due to Dirac, to describe the algorithm and the standard interpretation. The language of quantum mechanics that we must use in our description is mathematics; in particular, a branch of mathematics called "vector spaces". As mentioned earlier, as we attempt to delve into the workings of any new world or culture, we must make some attempt to learn the language of the new culture. Otherwise, we would have no hope of ever understanding the inner workings of this new culture. Science, especially physics, is another culture for almost all human beings. In the case of physics, especially quantum physics, the language of the new culture is mathematics. It is important to understand that mathematics is just another language....another way of thinking about things and that this language of mathematics happens to be better at explaining the inner workings of quantum physics than other languages developed from everyday experiences. Although, I will sometimes use words that are new to you, which is something that always happens when learning a new language, you are already familiar with all of the concepts and ideas I will use, i.e., you have come across these ideas before in other mathematics courses.

It will take us some time to work out the kinks in your knowledge, but only because you have never seen these ideas used in this context nor have you tried to use this language to explain complex physical phenomena before.

We will start this discussion using the language of vectors in 2- and 3-dimensional space that is familiar to you and then I will shift to Dirac language and repeat some of our earlier discussion so that you can get used to the new language.

During this discussion we will learn many of the fundamental ideas of linear algebra. Let us start with objects called **vectors**.

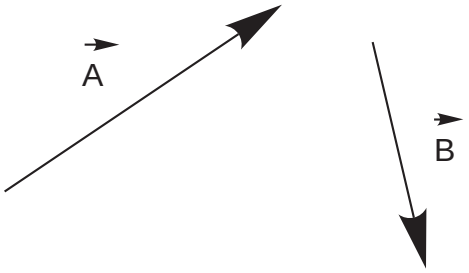
A vector has many levels of complexity and is a very abstract mathematical object. A vector is a mathematical(geometrical) object that is representable by two numbers in two dimensions, three numbers in three dimensions, and so on. One characterization is to specify its magnitude or length and orientation or direction - imagine that it is a directed line segment. As we shall see, quantum mechanics will be formulated in terms of vectors, but they will not be directed line segments.

The Standard Language of Vectors

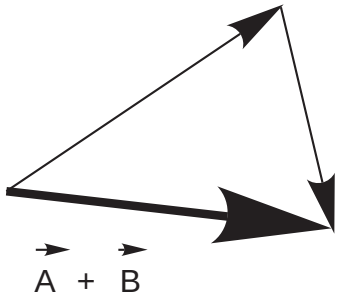
As we said, in ordinary space, we can represent a vector by a directed line segment(an arrow).

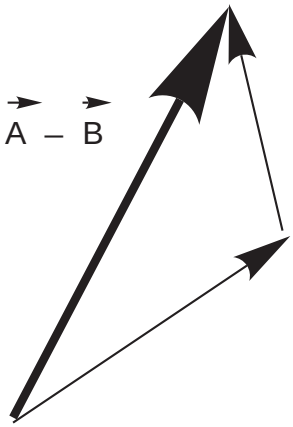
A straightforward property of a vector is multiplication of the vector by a scalar (a real number) $\vec{C} = \alpha\vec{A}$. In this case the magnitude of the vector changes and the direction stays the same (it might reverse if $\alpha < 0$).

Now given two vectors as shown below



we **define** the sum and difference of the two vectors or the general property **vector addition** by the diagrams shown below:





Clearly, vector addition as defined above, i.e.,

$$\vec{C} = \vec{A} + \vec{B} \quad , \quad \vec{D} = \vec{A} - \vec{B} = \vec{A} + (-\vec{B})$$

yields a new vector in each case. This new vector can have both a different direction and a different magnitude than either of the two vectors that are used to create it.

These two properties allow us to define a **linear combination** of vectors as

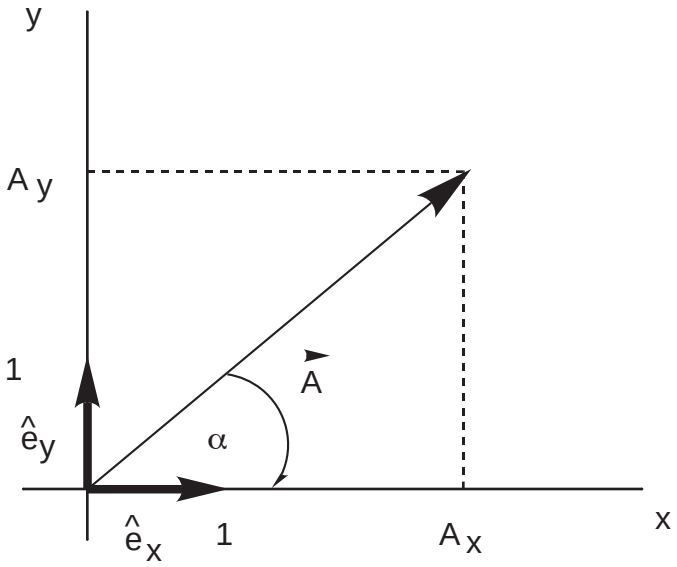
$$\vec{C} = \alpha\vec{A} + \beta\vec{B} \qquad (11)$$

which is also a well-defined vector.

Although this is a perfectly good way to proceed, it will not allow us to generalize the notion of a vector beyond ordinary physical space, which is an arena that will turn out to be much too confining in our effort to understand quantum mechanics.

We need to formulate these same concepts in another way.

Consider the vector shown below:



In this figure, we have also defined two special vectors, namely,

$$\begin{aligned} \hat{e}_x &= \text{unit(length = 1) vector in x - direction} \\ \hat{e}_y &= \text{unit(length = 1) vector in y - direction} \end{aligned} \qquad (12)$$

In terms of these **unit vectors** we can write

$$\vec{A} = A_x\hat{e}_x + A_y\hat{e}_y \qquad (13)$$

where

$$\begin{aligned} A_x\hat{e}_x &= \text{vector of length } A_x \text{ in the x - direction} \\ A_y\hat{e}_y &= \text{vector of length } A_y \text{ in the y - direction} \end{aligned} \qquad (14)$$

and the sum of these two vectors equals $\vec{A}$ because of the rule for adding vectors that we defined earlier.

We now define

$$\begin{aligned} A_x &= \text{component of vector } \vec{A} \text{ in the x - direction} \\ A_y &= \text{component of vector } \vec{A} \text{ in the y - direction} \end{aligned} \qquad (15)$$

From the diagram it is also clear that

$$A_x = A\cos\alpha \quad \text{and} \quad A_y = A\sin\alpha \qquad (16)$$

where

A = **length** of the vector $\vec{A} = \sqrt{A_x^2 + A_y^2}$ (17)
(by Pythagorous theorem)

We can then redefine vector addition in terms of components and unit vectors as follows:

$$\begin{aligned}\vec{A} &= A_x\hat{e}_x + A_y\hat{e}_y \\ \vec{B} &= B_x\hat{e}_x + B_y\hat{e}_y \\ \vec{A}+\vec{B} &= (A_x+B_x)\hat{e}_x + (A_y+B_y)\hat{e}_y \\ \vec{A}-\vec{B} &= (A_x-B_x)\hat{e}_x + (A_y-B_y)\hat{e}_y\end{aligned} \qquad (18)$$

i.e., we can just **add and subtract components**.

We now define a very important and useful new mathematical object using unit vectors. It is the **scalar or inner product** and its symbol is a . (a dot). We define this operation with a set of rules involving the unit vectors:

$$\hat{e}_x\cdot\hat{e}_x = 1 = \hat{e}_y\cdot\hat{e}_y \quad , \quad \hat{e}_x\cdot\hat{e}_y = 0 = \hat{e}_y\cdot\hat{e}_x$$

$$or\ \hat{e}_i\cdot\hat{e}_j = \delta_{ij} = \begin{cases}1 & \text{i = j} \\ 0 & \text{i} \neq \text{j}\end{cases} = \text{Kronecker delta} \qquad (19)$$

The inner product satisfies the following relations:

$$\begin{aligned}&(\alpha\hat{e}_i)\cdot(\beta\hat{e}_j) = \alpha\beta\hat{e}_i\cdot\hat{e}_j \\ &(\alpha\hat{e}_i+\gamma\hat{e}_k)\cdot(\beta\hat{e}_j+\eta\hat{e}_m) = \alpha\beta\hat{e}_i\cdot\hat{e}_j + \alpha\eta\hat{e}_i\cdot\hat{e}_m + \gamma\beta\hat{e}_k\cdot\hat{e}_j + \gamma\eta\hat{e}_k\cdot\hat{e}_m\end{aligned} \qquad (20)$$

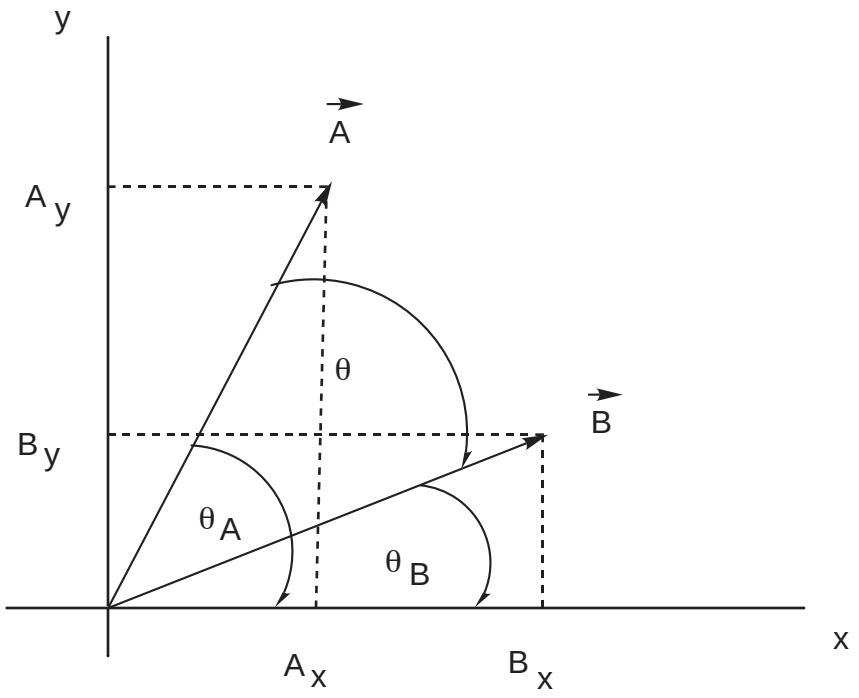
Using these defining relations we can now determine the scalar product of any two vectors as follows

$$\begin{aligned}\vec{A} &= A_x\hat{e}_x + A_y\hat{e}_y \\ \vec{B} &= B_x\hat{e}_x + B_y\hat{e}_y \\ \vec{A}\cdot\vec{B} &= (A_x\hat{e}_x + A_y\hat{e}_y)\cdot(B_x\hat{e}_x + B_y\hat{e}_y) \\ &= A_xB_x\hat{e}_x\cdot\hat{e}_x + A_xB_y\hat{e}_x\cdot\hat{e}_y + A_yB_x\hat{e}_y\cdot\hat{e}_x + A_yB_y\hat{e}_y\cdot\hat{e}_y \\ &= A_xB_x(1) + A_xB_y(0) + A_yB_x(0) + A_yB_y(1) = A_xB_x + A_yB_y\end{aligned} \qquad (21)$$

We note that

$$\begin{aligned}&\vec{A}\cdot\vec{A} = A_xA_x + A_yA_y = A_x^2 + A_y^2 = A^2 = norm\ of\ \vec{A} \\ &A = \sqrt{\vec{A}\cdot\vec{A}} = \text{length of the vector } \vec{A}\end{aligned} \qquad (22)$$

Now looking at the diagram below, we can derive another important result.



We have

$$\vec{A}\cdot\vec{B} = A_x B_x + A_y B_y = AB\left(\cos(\theta_A)\cos(\theta_A) + \sin(\theta_A)\sin(\theta_A)\right)$$
$$= AB\cos(\theta_A - \theta_B) = AB\cos\theta \qquad (23)$$

so that

$$\vec{A}\cdot\vec{B} = AB\cos\theta$$
$$= (\textit{length of } \vec{A})(\textit{length of } \vec{B})\cos(\textit{angle between } \vec{A} \textit{ and } \vec{B})$$
$$= (\textit{length of } \vec{A})(\textit{length of } \vec{B} \textit{ in the direction of } \vec{A})$$
$$= (\textit{length of } \vec{A})(\textit{projection of } \vec{B} \textit{ onto the direction of } \vec{A})$$

Therefore, we have

$$\vec{B} = \vec{A} \rightarrow \theta = 0 \rightarrow \vec{A}\cdot\vec{A} = A^2 \quad \text{as before}$$
$$\vec{B} \text{ perpendicular(orthogonal) to } \vec{A} \rightarrow \theta = \frac{\pi}{2} = 90^\circ \rightarrow \vec{A}\cdot\vec{B} = 0 \qquad (24)$$
$$\textit{or vice versa if } \vec{A}\cdot\vec{B} = 0\,, \textit{then } \vec{A} \text{ is orthogonal to } \vec{B}$$

If two vectors satisfy Eq.(19), then they are said to **orthonormal** = **orthogonal** (inner product = 0) + **normalized** to one (length = 1).

We also have for any vector

$$\vec{A} = A_x\hat{e}_x + A_y\hat{e}_y$$
$$\vec{A}\cdot\hat{e}_x = \left(A_x\hat{e}_x + A_y\hat{e}_y\right)\cdot\hat{e}_x = A_x = \text{x - component}$$
$$\vec{A}\cdot\hat{e}_y = \left(A_x\hat{e}_x + A_y\hat{e}_y\right)\cdot\hat{e}_y = A_y = \text{y - component} \qquad (25)$$
$$\vec{A} = (\vec{A}\cdot\hat{e}_x)\hat{e}_x + (\vec{A}\cdot\hat{e}_y)\hat{e}_y$$

Generalizing to 3 dimensions we have

$$\vec{A} = A_x\hat{e}_x + A_y\hat{e}_y + A_z\hat{e}_z \;=\; \text{any vector in the vector space} \qquad (26)$$

where the set of three orthonormal vectors $\{\hat{e}_x, \hat{e}_y, \hat{e}_z\}$ are called a **basis** for the vector space (any vector can be written as a linear combination of the basis vectors) and we have

$$\hat{e}_i \cdot \hat{e}_j = \begin{cases} 1 & \text{i = j} \\ 0 & \text{i} \neq \text{j} \end{cases} \qquad \text{i , j = x,y,z or 1,2,3}$$

The number of required basis vectors is the number of numbers needed to characterize a general vector = the **dimension** of the space.

The entire collection of vectors we can generate from a basis set is called a **vector space**.

So in this room, I would need 3 numbers to characterize each vector. This room is a small part of a 3-dimensional vector space, which is called the **universe** at an instant of time.

Completely removing (x,y,z) from our notation(because it limits us to a maximum of 3 dimensions) we have

$$\vec{A} = \sum_{j=1}^{3} A_j\hat{e}_j$$

$$\hat{e}_k \cdot \vec{A} = \hat{e}_k \cdot \sum_{j=1}^{3} A_j\hat{e}_j \;=\; \sum_{j=1}^{3} A_j\hat{e}_k \cdot \hat{e}_j = \sum_{j=1}^{3} A_j\delta_{kj} = A_k = k^{th}\ component \qquad (27)$$

so that

$$\vec{A} = \sum_{j=1}^{3} A_j\hat{e}_j = \sum_{j=1}^{3} \left(\hat{e}_j \cdot \vec{A}\right)\hat{e}_j \qquad (28)$$

Dirac Language

We now introduce the standard way of writing(from Dirac) vectors in Quantum Physics.

The **ket** $|A\rangle$ will be the symbol representing the vector $\vec{A}$ (the old notation) where the **ket-label** A contains **all the information** we know about the ket vector.

For example, we might have labels like (if we were talking about ordinary 2-dimensional vectors) $|Q\rangle = |r_Q, \theta_Q\rangle = |Q_x, Q_y\rangle$.

Vector addition is written as $|A\rangle + |B\rangle = |C\rangle$. This addition property is very important.

These ket vectors will represent **real physical systems** in the universe. The addition of ket vectors will be connected with the idea of **superposition** of physical systems that will be central to our understanding of quantum physics.

This property of two vectors that they can be added together to form a third vector in the same space will allow us, later on, to construct physical systems that are **sums** of other physical systems

with strange properties. In particular, it will allow us to construct **magenta electrons as superpositions of hard and soft electrons**!

So this very simple and basic property of vectors will be connected to powerful properties of real physical systems.

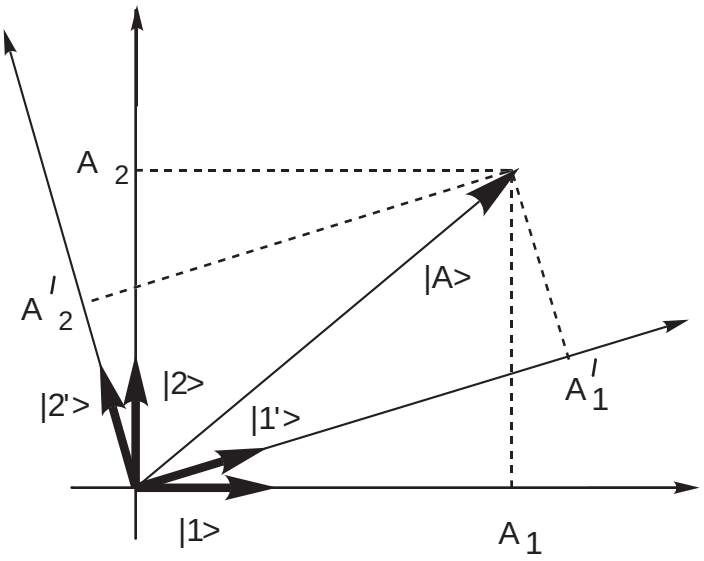
For the moment, however, the Dirac language **simply** looks like a change of notation.

Let us continue.

The equations that we have developed so far are rewritten in the following way in Dirac language:

$$\hat{e}_i \rightarrow |i\rangle$$
$$\hat{e}_i \cdot \hat{e}_j = \left(|i\rangle, |j\rangle\right) = \langle i|j\rangle = \text{"}bracket\text{"}$$
$$\vec{A} = \sum_{j=1}^{3} A_j \hat{e}_j \rightarrow |A\rangle = \sum_{j=1}^{3} A_j |j\rangle \qquad (29)$$
$$\hat{e}_k \cdot \vec{A} = A_k \rightarrow \langle k|A\rangle = A_k$$
$$\vec{A} = \sum_{j=1}^{3} \left(\hat{e}_j \cdot \vec{A}\right)\hat{e}_j \rightarrow |A\rangle = \sum_{j=1}^{3} \langle j|A\rangle |j\rangle$$

The actual choice of a basis set is arbitrary as shown below



where we have

$$|A\rangle = A_1|1\rangle + A_2|2\rangle = A'_1|1'\rangle + A'_2|2'\rangle \qquad (30)$$

Clearly, the vector is the same for different bases - only the components change; any **orthonormal** basis set is equivalent in quantum mechanics, i.e., any basis set can be used to express any vector (or physical system).

As we shall see, in quantum physics, the A_i numbers (the vector components) will represent **real physically measurable** quantities.

To generalize to an N-dimensional space(this is no longer any real

space that you can move around in), we must use N mutually orthonormal vectors to describe it, say the set $|1\rangle,|2\rangle,|3\rangle,.....,|N-1\rangle,|N\rangle$ where the orthonormality is expressed by the relations $\langle i|j\rangle = \delta_{ij}$.

Property : for scalar products we have $\langle A|\big(|B\rangle + |C\rangle\big) = \langle A|B\rangle + \langle A|C\rangle$

We now use this property in an ordinary 2-dimensional space example, i.e., suppose we have the pair of vectors $|x\rangle, |y\rangle$ where $\langle x|x\rangle = 1 = \langle y|y\rangle$, $\langle x|y\rangle = 1 = \langle y|x\rangle$ and where the (x,y) labels indicate the direction of the vector in real 2-dimensional space - a real physically measurable thing. Since the vectors $|x\rangle, |y\rangle$ form a orthonormal pair of vectors, we can use the pair of vectors $|x\rangle, |y\rangle$ as basis vectors for the 2-dimensional plane (all vectors in the plane can be written in terms of them).

Now suppose we have two vectors

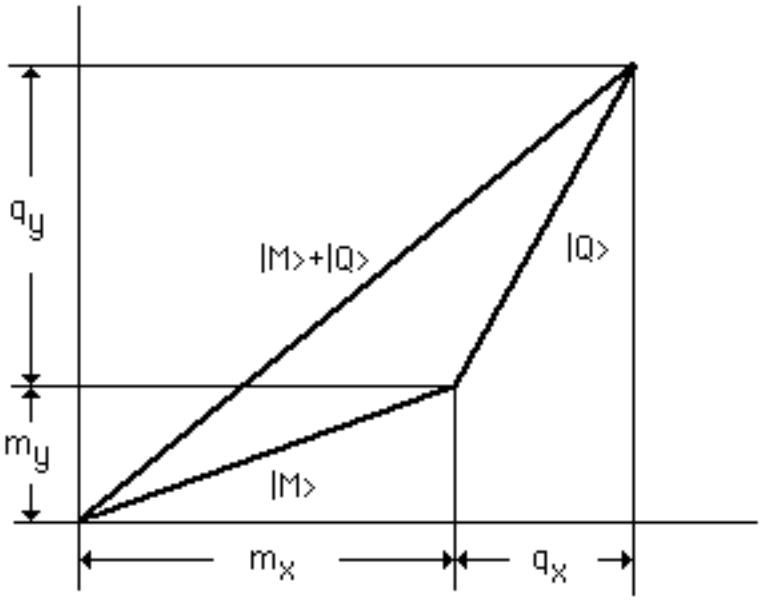
$$|M\rangle = m_x|x\rangle + m_y|y\rangle$$

$$m_x = \sqrt{\langle M|M\rangle}\cos\theta_M = \langle x|M\rangle = L_M\cos\theta_M \,,\; m_y = \sqrt{\langle M|M\rangle}\sin\theta_M = \langle y|M\rangle = L_M\sin\theta_M$$

$$|Q\rangle = q_x|x\rangle + q_y|y\rangle$$

$$q_x = \sqrt{\langle Q|Q\rangle}\cos\theta_Q = \langle x|Q\rangle = L_Q\cos\theta_Q \,,\; q_y = \sqrt{\langle Q|Q\rangle}\sin\theta_Q = \langle y|Q\rangle = L_Q\sin\theta_Q$$

as shown below:



Now

$$|M\rangle + |Q\rangle = (m_x + q_x)|x\rangle + (m_y + q_y)|y\rangle$$

where we have just added components.

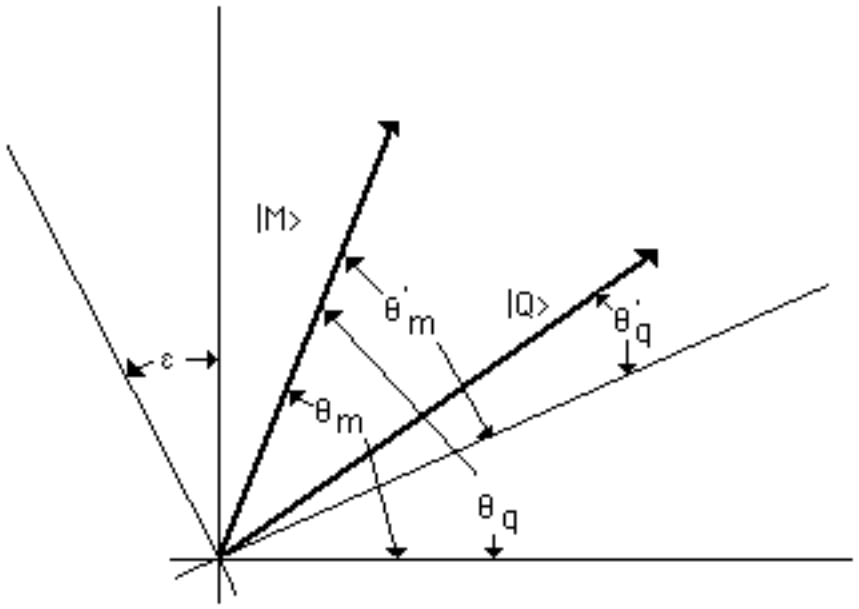
The scalar product of these two vectors is given by

$$\langle M|Q\rangle = m_x q_x + m_y q_y \qquad (31)$$

= **invariant** (**independent** of basis choice)

Proof:

If we choose a rotated axis pair (rotate by angle ε as shown below),

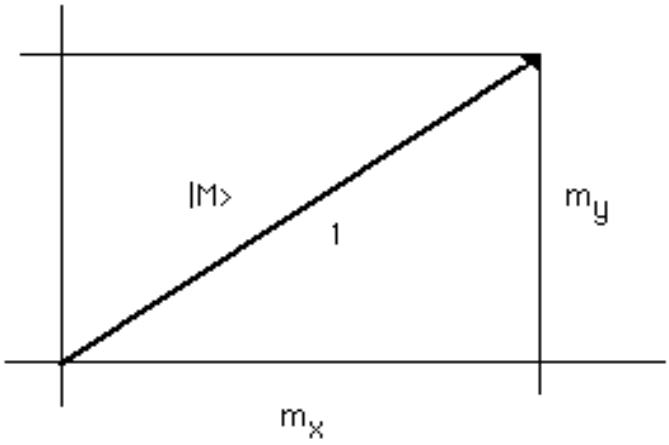


then we get

$$\langle M'|Q'\rangle = L_M L_Q \cos\left(\theta'_M - \theta'_Q\right) = L_M L_Q \cos\left((\theta_M - \varepsilon) - (\theta_Q - \varepsilon)\right) = L_M L_Q \cos\left(\theta_M - \theta_Q\right) = \langle M|Q\rangle$$

so that the scalar product is **independent of choice of basis vectors**.

Now suppose we want $\langle M|M\rangle = 1$ (vector normalized to 1). What does that mean? If we let $Q \to M$ in Eq.(31), then we must have $\langle M|M\rangle = m_x^2 + m_y^2 = 1$ or the sum of the squares of the components = 1. This is just the Pythagorean theorem(see diagram)



If $\langle M|M\rangle = m_x^2 + m_y^2 \neq 1$, then we can **normalize** the vector by dividing as shown below:

$$|M\rangle = \frac{1}{\sqrt{m_x^2 + m_y^2}}(m_x|x\rangle + m_y|y\rangle) \qquad (32)$$

So we only need to know the two numbers (the two orthogonal components) for each vector and everything else can be determined.

The components seem to contain all the possible information about a vector.

Now, generalizing to N dimensions (where we would need N mutually orthonormal basis vectors). Let us designate the basis vectors by

$$|1\rangle, |2\rangle, |3\rangle, ..., |N-1\rangle, |N\rangle$$

Suppose we have the two vectors

$$|M\rangle = m_1|1\rangle + m_2|2\rangle + ...m_N|N\rangle$$
$$|Q\rangle = q_1|1\rangle + q_2|2\rangle + ...q_N|N\rangle$$

then(as before),

$$|M\rangle + |Q\rangle = (m_1 + q_1)|1\rangle + (m_2 + q_2)|2\rangle + ... + (m_N + q_N)|N\rangle$$

and

$$\langle M|Q\rangle = m_1q_1 + m_2q_2 + ... + m_Nq_N$$

= **invariant** (**independent** of basis choice)

Our discussion up to this point has involved what is called a **real vector space**, i.e., the vector components are real numbers. Quantum mechanics involves a complex vector space, so we will need a generalization to create the correct vectors and vector properties appropriate for quantum theory.

In particular, the coefficients (components) of the basis vectors will need to be complex numbers, i.e., in the expression

$$|M\rangle = m_x|x\rangle + m_y|y\rangle$$

m_x and m_y will need to be complex numbers.

To allow for the possibility that quantum vectors might have complex components, we **generalize** the definition of the scalar product to

$$\langle M|Q\rangle = m^*_x\, q_x + m^*_y\, q_y \qquad (33)$$

and

$$\langle M|M\rangle = m^*_x\, m_x + m^*_y\, m_y = |m_x|^2 + |m_y|^2 \qquad (34)$$

which means that the length is still a real number!

An alternative way of looking at the inner products goes as follows:

given two **ket** vectors $|A\rangle, |B\rangle$ we define new mathematical objects given by the symbols $\langle A|, \langle B|$ called **bra** vectors. If the ket vectors $|A\rangle, |B\rangle$ belong to a vector space, then the bra vectors $\langle A|, \langle B|$ also belong to a vector space called the **dual** space and

$$\langle A| \quad \text{is the dual vector for the vector} \quad |A\rangle$$

If we have the ket vector $|A\rangle = A_1|1\rangle + A_2|2\rangle$ then its dual or bra vector is

$$\langle A| = A_1^*\langle 1| + A_2^*\langle 2| \tag{35}$$

and the bra(-c-)ket or inner product is defined as

$$\langle A|B\rangle = \big(\langle A|\big)|B\rangle = \big(|A\rangle, |B\rangle\big) \tag{36}$$

The bra vector is called a **linear functional.** When acting on a ket vector it produces a number (the inner product as in Eq.(36)). Using this idea, Eq.(34) above can be written, explicitly, in this way

$$\begin{aligned}
\langle M|M\rangle &= \Big(m^*_x\,\langle x| + m^*_y\,\langle y|\Big)\Big(m_x|x\rangle + m_y|y\rangle\Big) \\
&= m^*_x\,m_x\langle x|x\rangle + m^*_x\,m_y\langle x|y\rangle + m^*_y\,m_x\langle y|x\rangle + m^*_y\,m_y\langle y|y\rangle \quad \text{(property \#1)} \\
&= m^*_x\,m_x + m^*_y\,m_y \qquad \text{(using orthonormality)} \\
&= |m_x|^2 + |m_y|^2
\end{aligned}$$

We also have that

$$\langle V|W\rangle = \langle W|V\rangle^* \tag{37}$$

Illustration using Color and Hardness Properties

The above formalism is easily able to describe the world of color and hardness.

Suppose that we choose the ket vector $|g\rangle$ to represent the state of an electron with color = green, the ket vector $|m\rangle$ to represent the state of an electron with color = magenta, the ket vector $|h\rangle$ to represent the state of an electron with hardness = hard, and the ket vector $|s\rangle$ to represent the state of an electron with hardness = soft.

We will assume that

$$\begin{aligned}
\langle g|g\rangle &= 1 = \langle m|m\rangle \\
\langle g|m\rangle &= 0 = \langle m|g\rangle
\end{aligned} \tag{38}$$

i.e., they are an orthonormal basis set and

$$\begin{aligned}
\langle h|h\rangle &= 1 = \langle s|s\rangle \\
\langle h|s\rangle &= 0 = \langle s|h\rangle
\end{aligned} \tag{39}$$

so that they are an orthonormal basis set also.

Since we observed that hard electrons can appear to be both green and magenta electrons we might make **a first guess** (this is what theoretical physicists do) that

$$|hard\rangle = |h\rangle = a|g\rangle + b|m\rangle \quad \text{= SUPERPOSITION of both color states} \qquad (40)$$
$$\text{= LINEAR COMBINATION of both color states}$$

where

$$a = \langle g|h\rangle \quad \text{and} \quad b = \langle m|h\rangle \qquad (41)$$

Can this simple correspondence really work? Is it possible that SUPERPOSITION is just a linear combination or just vector addition!

We shall see later that it does work!!

We will find many different mathematical ways to represent these fundamental vector objects.

If we know the basis we are using (we shall see later that it will be chosen according to what we are **trying to measure**) then we only need the two numbers a and b, i.e., the components .

Two different ways to represent the superposition are

$$|hard\rangle = a|g\rangle + b|m\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \qquad (42)$$

where $\begin{pmatrix} a \\ b \end{pmatrix}$ is a column vector or 1 x 2 matrix (more about this later).

That is almost all we will need to know about vectors in order to do quantum theory.

So lots of new words, but really just high school ideas of vectors, components, length(Pythagorus), perpendicular, etc along with an open mind willing to accept new uses for these old ideas and new ways of expressing them.

As we shall see, **ket vectors will represent physical systems in quantum mechanics.**

IV. Operators in a Vector Space

There is one other kind of mathematical object we will need to know about to do quantum theory. They are called **operators**. This one may not be familiar to you from H.S. but the concept should be, as we will now see.

In a vector space, an **operator** is a **definite rule for taking every vector in the space into some other vector in the space**.

An operator will be represented by the ^ symbol, i.e., $\hat{Q}$, so that the action of the operator $\hat{Q}$ is given by the relation

$$|B'\rangle = \hat{Q}|B\rangle \qquad (43)$$

Think about this moving from **right to left** ---

vector $|B\rangle$ acted upon by operator $\hat{Q}$ is changed into vector $|B'\rangle$

Think of an operator as some kind of box, where you put a vector in and get another(**either different or the same**) vector out. This is similar to the definition of a function for the space of numbers.

Some properties of operators in QM:

$$\hat{Q}\left(|A\rangle + |B\rangle\right) = \hat{Q}|A\rangle + \hat{Q}|B\rangle \quad \text{(linearity)}$$

$$\hat{Q}\left(c|A\rangle\right) = c\hat{Q}|A\rangle$$

$$\langle C|\left(\hat{Q}|B\rangle\right) = \langle C|B'\rangle = number \equiv \langle C|\hat{Q}|B\rangle \quad \text{(matrix element)} \qquad (44)$$

$$(\hat{Q}_1 + \hat{Q}_2)|A\rangle = \hat{Q}_1|A\rangle + \hat{Q}_2|A\rangle \quad \text{(linearity)}$$

$$(\hat{Q}_1\hat{Q}_2)|A\rangle = \hat{Q}_1(\hat{Q}_2|A\rangle) \quad \text{(order matters)}$$

These properties imply that all of our operators are what the mathematicians call **LINEAR** operators. The fact that quantum mechanics can be understood using only linear operators is truly amazing because they are the simplest kind of operator that mathematicians can think of.

All observables or "quantities that we can measure" will be represented by operators in quantum mechanics.

Now we introduce a **completely new property** using vectors and operators.

This new property of operators involves mathematical objects called **eigenvalues and eigenvectors**.

In some special cases it turns out that we find the following result:

$$\hat{B}|B\rangle = b|B\rangle \qquad (45)$$

where b = number, that is, we get the **exact same vector** back when using the operator (possibly multiplied by a number $\neq 1$).

In this case, $|B\rangle$ is an **eigenvector** of $\hat{B}$ with **eigenvalue** b.

These ideas will be very important in quantum mechanics because the **only possible results that can be obtained from the measurement of a physical quantity or observable** that can be represented by the operator $\hat{W}$ will be the **eigenvalues** of $\hat{W}$ (**a postulate we will make later**)!

The set of eigenvalues is called the **spectrum** of the operator.

It will also turn out that the **eigenvectors** of the operator representing an **observable** can **always** be used as the **basis** for the vector space.

Average Values in Quantum Mechanics
or Another Way to Represent Operators

Suppose that we have N **identically prepared** physical systems, each represented by the **same** state vector $|\psi\rangle$.

Suppose that we make measurements of an observable represented by the operator $\hat{B}$ with eigenvalues/eigenvectors given by

$$\hat{B}|b_j\rangle = b_j|b_j\rangle \quad \text{j} = 1,2,3,4,.....$$

Suppose the measurement results(remember they must be the eigenvalues as we shall see later) are

$$b_k \quad , \quad n_k \text{ times} \qquad \text{k} = 1,2,3,4,.....$$

where

$$\sum_k n_k = N = \text{ total number of measurements}$$

Now from the definition we derived earlier, the average value of $\hat{B}$ is

$$\begin{aligned} \langle \hat{B} \rangle &= \text{average or expectation value of } \hat{B} \\ &= \frac{1}{N}\sum_k n_k b_k = \sum_k \frac{n_k}{N} b_k = \sum_k b_k \,\text{Prob}(b_k) \end{aligned} \tag{46}$$

In quantum theory, this will given by (**a postulate**) the expression

$$\langle \hat{B} \rangle = \sum_k b_k \,\text{Prob}(b_k) = \sum_k b_k |\langle b_k|\psi\rangle|^2 \tag{47}$$

Since the set $\{|b_k\rangle\}$ is a basis we can write

$$|\psi\rangle = \sum_k \langle b_k|\psi\rangle |b_k\rangle \tag{48}$$

so that we will find the important result

$$\text{Prob}(b_k) = |\langle b_k|\psi\rangle|^2 = \text{absolute value squared of the component of } |\psi\rangle \text{ along } |b_k\rangle \tag{49}$$

Now by definition

$$|\langle b_k|\psi\rangle|^2 = \langle b_k|\psi\rangle^* \langle b_k|\psi\rangle \tag{50}$$

and from Eq.(37) and the definition of the bra vector we have

$$\langle b_k|\psi\rangle^* = \langle \psi|b_k\rangle \tag{51}$$

Therefore

$$|\langle b_k|\psi\rangle|^2 = \langle \psi|b_k\rangle\langle b_k|\psi\rangle \tag{52}$$

and

$$\left\langle \hat{B} \right\rangle = \sum_k b_k \langle \psi | b_k \rangle \langle b_k | \psi \rangle = \langle \psi | \left(\sum_k b_k | b_k \rangle \langle b_k | \right) | \psi \rangle \tag{53}$$

Now the quantity $\sum_k b_k |b_k\rangle\langle b_k| = \hat{Q}$ must be some operator in the vector space.

Proof:

Now we can always write

$$|a\rangle = \sum_s d_s |b_s\rangle = \text{arbitrary vector}$$

since the set $\{|b_s\rangle\}$ is a basis. This gives

$$\begin{aligned} \hat{Q}|a\rangle &= \left(\sum_k b_k |b_k\rangle\langle b_k| \right) |a\rangle = \left(\sum_k b_k |b_k\rangle\langle b_k| \right) \left(\sum_s d_s |b_s\rangle \right) \\ &= \sum_{k,s} d_s b_k |b_k\rangle\langle b_k | b_s\rangle = \sum_{k,s} d_s b_k |b_k\rangle \delta_{ks} = \sum_k b_k d_k |b_k\rangle \end{aligned}$$

so that the operation by $\hat{Q}$ on an arbitrary vector returns a vector indicating that $\hat{Q}$ is an operator.

But we also have

$$\hat{B}|a\rangle = \sum_k d_k \hat{B}|b_k\rangle = \sum_k b_k d_k |b_k\rangle$$

This is the same result as for the operator $\hat{Q}$, which says (since the vector $|a\rangle$ is arbitrary) that we can always write

$$\hat{B} = \sum_k b_k |b_k\rangle\langle b_k| \tag{54}$$

This is another(very important) way of representing the operator $\hat{B}$, i.e., **any operator can be written in terms of its eigenvalues and eigenvectors**.

Therefore, the average value in Eq.(53) becomes

$$\left\langle \hat{B} \right\rangle = \langle \psi | \left(\sum_k b_k |b_k\rangle\langle b_k| \right) | \psi \rangle = \langle \psi | \hat{B} | \psi \rangle \tag{55}$$

Therefore, when we do an experiment of this type, i.e., identical measurements on many identical systems, the (expected) average of those measurements is just this **average** or **expectation** value $\langle \psi | \hat{B} | \psi \rangle$.

Projection Operators

Operators of the form

$$\hat{P} = |b_k\rangle\langle b_k| \tag{56}$$

where

$$\hat{B}|b_j\rangle = b_j|b_j\rangle \quad \text{j} = 1,2,3,4,.....$$

are called **projection operators**. They have some very interesting and useful properties.

First, what are the eigenvalues of a projection operator? We have

$$\hat{P}^2 = (|b_k\rangle\langle b_k|)(|b_k\rangle\langle b_k|) = |b_k\rangle\langle b_k|b_k\rangle\langle b_k| = |b_k\rangle\langle b_k| = \hat{P}$$

or if

$$\hat{P}|\alpha\rangle = \lambda|\alpha\rangle$$

so that $|\alpha\rangle = \text{eigenvector}$ and $\lambda = \text{corresponding eigenvalue}$, then we have

$$\hat{P}^2|\alpha\rangle = \lambda\hat{P}|\alpha\rangle = \lambda^2|\alpha\rangle = \hat{P}|\alpha\rangle = \lambda|\alpha\rangle$$

$$(\lambda^2 - \lambda)|\alpha\rangle = 0$$

$$\lambda^2 - \lambda = 0 \rightarrow \lambda = 0 \text{ or } 1 \rightarrow \text{eigenvalues}$$

Second, consider the following operator $\hat{I} = \sum_k |b_k\rangle\langle b_k|$. We have

$$\hat{I}|\beta\rangle = \sum_k |b_k\rangle\langle b_k|\beta\rangle = \sum_k \langle b_k|\beta\rangle|b_k\rangle = |\beta\rangle$$

or

$$\hat{I} = \sum_k |b_k\rangle\langle b_k| \tag{57}$$

is the i**dentity** operator. The identity operator is defined as that operator which does not change anything!

Our earlier result (Eq.(54)) now makes sense since

$$\hat{B} = \hat{B}I = \hat{B}\sum_k |b_k\rangle\langle b_k| = \sum_k \hat{B}|b_k\rangle\langle b_k| = \sum_k b_k|b_k\rangle\langle b_k|$$

Operator Examples:

Suppose we define an operator $\hat{G} = |g\rangle\langle g| = |green\rangle\langle green|$. We then have these properties

$$\hat{G}|g\rangle = |g\rangle\langle g|g\rangle = |g\rangle$$

$$\hat{G}|m\rangle = |g\rangle\langle g|m\rangle = 0$$

or the states representing the green and magenta electrons are the eigenvectors with eigenvalues 1 and 0, respectively.

We find that

$$\langle g|\hat{G}|g\rangle = |\langle g|g\rangle|^2 = 1 = \text{expectation value of } \hat{G} \text{ in the green state}$$

and we also have

$$\langle m|\hat{G}|m\rangle = \left|\langle g|m\rangle\right|^2 = 0 = \text{expectation value of } \hat{G} \text{ in the magenta state}$$

These results make sense if we **interpret** $\hat{G} = |g\rangle\langle g|$ as the operator corresponding to a measurement of the green property of electrons, i.e., an observer looking at the output of a green aperture of a color box.

The first result then says if we measure the probability that the color of a green electron is green we get the value 1 as expected and the second result then says if we measure the probability that the color of a green electron is magenta we get the value 0 as expected.

Pushing this strange idea even further, if we assume that

$$|hard\rangle = |h\rangle = \frac{1}{\sqrt{2}}|g\rangle + \frac{1}{\sqrt{2}}|m\rangle \text{ = SUPERPOSITION of green and magenta}$$

then we have

$$\begin{aligned}
\langle h|\hat{G}|h\rangle &= \text{expectation value of } \hat{G} \text{ in the hard state} \\
&= \left(\frac{1}{\sqrt{2}}\langle g| + \frac{1}{\sqrt{2}}\langle h|\right)\hat{G}\left(\frac{1}{\sqrt{2}}|g\rangle + \frac{1}{\sqrt{2}}|m\rangle\right) \\
&= \frac{1}{2}\langle g|\hat{G}|g\rangle + \frac{1}{2}\langle g|\hat{G}|m\rangle + \frac{1}{2}\langle m|\hat{G}|g\rangle + \frac{1}{2}\langle m|\hat{G}|m\rangle \\
&= \frac{1}{2}\langle g|\hat{G}|g\rangle = \frac{1}{2}
\end{aligned}$$

Another way of saying this is (using Eq.(47))

$$\begin{aligned}
\langle h|\hat{G}|h\rangle &= \text{expectation value of } \hat{G} \text{ in the hard state} \\
&= \sum(\text{eigenvalue g})(\text{probability of g in}|h\rangle) \\
&= (1)\left|\langle eigenvalue = 1|h\rangle\right|^2 + (0)\left|\langle eigenvalue = 0|h\rangle\right|^2 \\
&= (1)\left|\langle g|h\rangle\right|^2 + (0)\left|\langle m|h\rangle\right|^2 \\
&= (1)\frac{1}{2} + (0)\frac{1}{2} = \frac{1}{2}
\end{aligned}$$

which again makes sense i.e., if we have a beam of hard electrons, then we will measure an electron to be green 1/2 of the time as we observed earlier!

Clearly, this formalism is both neat and very powerful and certainly seems to have the potential to describe our earlier observations.

More Useful Mathematical Ideas

Matrices

We have been representing an operator by the way it acts on vectors.

An operator can also be represented by an array of numbers. In

mathematics such an array is called a **matrix**.,i.e., a matrix is a mathematical object which takes the form

$$[O] = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \text{2x2 matrix representing operator } \hat{O} \qquad (58)$$

where the numbers

$$\alpha, \beta, \gamma, \delta \text{ are called matrix elements of the operator } \hat{O}$$

and their numerical values depend on which set of basis vectors we are using and the corresponding operator $\hat{O}$.

Let us assume the two basis vectors are labelled by $|1\rangle, |2\rangle$.

We then define

$$\alpha = O_{11} = \langle 1|\hat{O}|1\rangle = O_{\text{row label, column label}}$$

$$\beta = O_{12} = \langle 1|\hat{O}|2\rangle$$

$$\gamma = O_{21} = \langle 2|\hat{O}|1\rangle$$

$$\delta = O_{22} = \langle 2|\hat{O}|2\rangle$$

or

$$[O] = \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} = \begin{pmatrix} \langle 1|\hat{O}|1\rangle & \langle 1|\hat{O}|2\rangle \\ \langle 2|\hat{O}|1\rangle & \langle 2|\hat{O}|2\rangle \end{pmatrix} \qquad (59)$$

A vector in the space is then represented by a 2-element column matrix (vector) of the form

$$|a\rangle = a_1|1\rangle + a_2|2\rangle = \langle 1|a\rangle|1\rangle + \langle 2|a\rangle|2\rangle \equiv \begin{pmatrix} \langle 1|a\rangle \\ \langle 2|a\rangle \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \qquad (60)$$

This implies that the basis vectors themselves are represented by the particular column matrices

$$|1\rangle = \begin{pmatrix} \langle 1|1\rangle \\ \langle 2|1\rangle \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |2\rangle = \begin{pmatrix} \langle 1|2\rangle \\ \langle 2|2\rangle \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (61)$$

Now consider the color basis. We have

$$|1\rangle = |g\rangle = \begin{pmatrix} \langle g|g\rangle \\ \langle m|g\rangle \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |2\rangle = |m\rangle = \begin{pmatrix} \langle g|m\rangle \\ \langle m|m\rangle \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$|hard\rangle = |h\rangle = \frac{1}{\sqrt{2}}|g\rangle + \frac{1}{\sqrt{2}}|m\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad (62)$$

We also have

$$|soft\rangle = |s\rangle = \frac{1}{\sqrt{2}}|g\rangle - \frac{1}{\sqrt{2}}|m\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The operator $\hat{G}$, was defined to be

$$G = |g\rangle\langle g| \tag{63}$$

and it represented measuring the **"green"** property of electrons.

The matrix representing $\hat{G}$ can now be written down:

$$(\hat{G})_{11} = \langle g|\big(|g\rangle\langle g|\big)|g\rangle = 1$$
$$(\hat{G})_{12} = \langle g|\big(|g\rangle\langle g|\big)|m\rangle = 0$$
$$(\hat{G})_{21} = \langle m|\big(|g\rangle\langle g|\big)|g\rangle = 0$$
$$(\hat{G})_{22} = \langle m|\big(|g\rangle\langle g|\big)|m\rangle = 0$$

or

$$\left[\hat{G}\right] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{64}$$

We then have

$$\hat{G}|g\rangle = \big(|g\rangle\langle g|\big)|g\rangle = |g\rangle$$

or

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \text{matrix multiplication}$$

where matrix multiplication is defined by these relations: if

$$[A] = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad , \quad [B] = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

then

$$[A][B] = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

or

$$[AB]_{ij} = \sum_k a_{ik}b_{kj} \tag{65}$$

So we have **two equivalent ways** of doing calculations, i.e., using **operator algebra** or **matrix algebra**.

Another Example in a Finite Dimensional Vector Space

Let us consider the 2-dimensional vector space spanned by the orthonormal basis set

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad , |2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

We can define two projection operators as

$$\hat{P}_1 = |1\rangle\langle 1| \quad , \quad \hat{P}_2 = |2\rangle\langle 2|$$

The matrix representation of these two projection operator is easily found using $\langle 1|1\rangle = \langle 2|2\rangle = 1$ and $\langle 1|2\rangle = \langle 2|1\rangle = 0$ and $\hat{Q}_{ki} = \langle k|\hat{Q}|i\rangle$. We have

$$\left(\hat{P}_1\right) = \begin{pmatrix} \langle 1|\hat{P}_1|1\rangle & \langle 1|\hat{P}_1|2\rangle \\ \langle 2|\hat{P}_1|1\rangle & \langle 2|\hat{P}_1|2\rangle \end{pmatrix} = \begin{pmatrix} \langle 1|1\rangle\langle 1|1\rangle & \langle 1|1\rangle\langle 1|2\rangle \\ \langle 2|1\rangle\langle 1|1\rangle & \langle 2|1\rangle\langle 1|2\rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\left(\hat{P}_2\right) = \begin{pmatrix} \langle 1|\hat{P}_2|1\rangle & \langle 1|\hat{P}_2|2\rangle \\ \langle 2|\hat{P}_2|1\rangle & \langle 2|\hat{P}_2|2\rangle \end{pmatrix} = \begin{pmatrix} \langle 1|2\rangle\langle 2|1\rangle & \langle 1|2\rangle\langle 2|2\rangle \\ \langle 2|2\rangle\langle 2|1\rangle & \langle 2|2\rangle\langle 2|2\rangle \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Now consider an arbitrary vector in this space

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1|1\rangle + a_2|2\rangle$$

We then have (using both Dirac and matrix language)

$$\hat{P}_1|a\rangle = |1\rangle\langle 1|a\rangle = a_1|1\rangle\langle 1|1\rangle + a_2|1\rangle\langle 1|2\rangle = a_1|1\rangle$$

or

$$\left(\hat{P}_1\right)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ 0 \end{pmatrix} = a_1\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and the projection operator performs as advertised.

We note that (at least in this **special case**)

$$\left(\hat{P}_1\right) + \left(\hat{P}_2\right) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{I} = \textit{identity operator}$$

or

$$(\hat{P}_1 + \hat{P}_2)|a\rangle = (|1\rangle\langle 1| + |2\rangle\langle 2|)|a\rangle = \sum_{j=1}^{2} |j\rangle\langle j|a\rangle = |a\rangle = \hat{I}|a\rangle$$

where we have made use of the expansion formula for an arbitrary state in an orthonormal basis.

Commutator

We define the **commutator** between two operators as

$$\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} \qquad (66)$$

For ordinary numbers the commutator is zero, but for operators it is not always zero.

If two operators have a **zero** commutator, then we say that they **commute**.

If we have

$$\hat{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad , \quad \hat{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

then

$$\left[\hat{A},\hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} - \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$$

$$= \begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix} - \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = -2\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \neq 0 \rightarrow \text{don't commute}$$

Let us devise a **general procedure** for finding the eigenvalues and eigenvectors of an operator. We will use the matrices that represent operators. In particular, we consider the matrices (operators)

$$\hat{A} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \quad , \quad \hat{B} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \quad , \quad \hat{C} = \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$$

We consider the operator $\hat{A}$ first. The eigenvalue/eigenvector problem can be written in the following way:

$$\hat{A}|\alpha\rangle = \alpha|\alpha\rangle$$

or

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \alpha\begin{pmatrix}a\\ b\end{pmatrix}$$

The last matrix equation corresponds to two simple algebraic equations

$$a = \alpha a$$

$$b = -\alpha b$$

The solutions to these equations are:

$$a \neq 0 \rightarrow \alpha = 1$$

$$\rightarrow b = -b \rightarrow b = 0$$

and

$$b \neq 0 \rightarrow \alpha = -1$$

$$\rightarrow a = -a \rightarrow a = 0$$

This says that the eigenvalues of $\hat{A}$ are $\alpha = \pm 1$ and the corresponding eigenvectors (normalized) to 1 are

$$|+1\rangle = \begin{pmatrix}1\\ 0\end{pmatrix} \quad , \quad |-1\rangle = \begin{pmatrix}0\\ 1\end{pmatrix}$$

So, for diagonal matrices (where the only non-zero matrix elements lie along the diagonal) the eigenvalue/eigenvector problem is trivial. The eigenvalues are just the diagonal matrix elements and the eigenvectors are just column vectors with a single entry = 1 and all other entries = 0. This generalizes to any number of dimensions.

Let us use this simple case to illustrate the general procedure for non-diagonal matrices. We had

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \alpha\begin{pmatrix}a\\ b\end{pmatrix}$$

This can be written as

$$\begin{pmatrix} 1-\alpha & 0 \\ 0 & -1-\alpha \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 0$$

This set of equations is called homogeneous (zeroes on the right-hand side). A trivial solution always exists - it is $a = b = 0$. This solution is of no interest to physicists. We will have a non-trivial solution if and only if

$$\text{determinant}\begin{pmatrix} 1-\alpha & 0 \\ 0 & -1-\alpha \end{pmatrix} = \det\begin{pmatrix} 1-\alpha & 0 \\ 0 & -1-\alpha \end{pmatrix} = \begin{vmatrix} 1-\alpha & 0 \\ 0 & -1-\alpha \end{vmatrix} = 0$$

where

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

In the above case we get

$$\begin{vmatrix} 1-\alpha & 0 \\ 0 & -1-\alpha \end{vmatrix} = 0 = (1-\alpha)(-1-\alpha) \rightarrow \alpha = \pm 1 = \text{eigenvalues}$$

Proof: Suppose that we have a pair of linear equations in two unknowns

$$a\alpha_1 + b\alpha_2 = q_1$$
$$c\alpha_1 + d\alpha_2 = q_2$$

We can write these equations in matrix notation as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$$

or

$$A\alpha = Q$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad , \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad , \quad Q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$$

The inverse of a matrix A^{-1} (or an operator) is defined by

$$A^{-1}A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = identity \qquad (67)$$

If the inverse exists, then we can write

$$A^{-1}A\alpha = I = \alpha = A^{-1}Q$$

which represents a solution for the unknowns α_1 and α_2.

For a 2x2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the inverse is given by

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \qquad \text{where} \qquad \det A = ad - bc$$

since

$$A^{-1}A = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{68}$$

The solution of the original equation is then given by

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \tag{69}$$

Example:

Suppose we have

$$x + 2y = 5$$
$$2x - 3y = 4$$

or

$$\begin{pmatrix} 1 & 2 \\ 2 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

We have

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -3 \end{pmatrix} \quad , \quad \det A = -7 \quad , \quad A^{-1} = -\frac{1}{7}\begin{pmatrix} -3 & -2 \\ -2 & 1 \end{pmatrix}$$

Check:

$$A^{-1}A = -\frac{1}{7}\begin{pmatrix} -3 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & -3 \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} -7 & 0 \\ 0 & -7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as it should.

We therefore have the solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} -3 & -2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 5 \\ 4 \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} -7 \\ -14 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

or

$$x = 1, y = 2$$

Now let us return to the case of the eigenvalue/eigenvector equation where we have

$$\hat{A}|\alpha\rangle = \alpha|\alpha\rangle = \alpha\hat{I}|\alpha\rangle$$

or

$$(\hat{A} - \alpha\hat{I})|\alpha\rangle = 0$$

If we define

$$\hat{R} = \hat{A} - \alpha\hat{I}$$

then we have the equation

$$\hat{R}|\alpha\rangle = 0$$

or

$$\left[\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} - \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}\right]\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} A_{11}-\alpha & A_{12} \\ A_{21} & A_{22}-\alpha \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Thus, we have a solution

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \frac{1}{\det\begin{pmatrix} A_{11}-\alpha & A_{12} \\ A_{21} & A_{22}-\alpha \end{pmatrix}}\begin{pmatrix} A_{11}-\alpha & A_{12} \\ A_{21} & A_{22}-\alpha \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The only possibility that this gives a nonzero(or nontrivial) result is when

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

which completes the proof.

A more general procedure(matrix not diagonal) for finding eigenvalues goes like the following:

$$\hat{G} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$$

$$\hat{G}|\lambda\rangle = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \lambda\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \lambda|\lambda\rangle$$

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} - \lambda\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} g_{11}-\lambda & g_{12} \\ g_{21} & g_{22}-\lambda \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = 0$$

Once again, this last set of equation is pair of homogeneous linear equations. Again, there is a always the trivial solution $(\lambda_1 = \lambda_2 = 0)$. A non-trivial solution exists only if

$$\text{determinant}\begin{pmatrix} g_{11}-\lambda & g_{12} \\ g_{21} & g_{22}-\lambda \end{pmatrix} = (g_{11}-\lambda)(g_{22}-\lambda) - g_{12}g_{21} = 0 \qquad (70)$$

which is a **quadratic** equation for the eigenvalues.

Let us apply this procedure to the other two operators. First we work with $\hat{B}$.

$$\hat{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \rightarrow \begin{vmatrix} -\alpha & 1 \\ 1 & -\alpha \end{vmatrix} = 0 = \alpha^2 - 1 \rightarrow \alpha = \pm 1 = \text{eigenvalues}$$

To find the eigenvectors we then proceed as follows:

$$\alpha = +1$$

$$\rightarrow \hat{B}|+1\rangle_B = |+1\rangle_B \rightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\rightarrow b = a$$

To normalize the eigenvector we must have

$$ {}_B\langle +1|+1\rangle_B = 1 = a^2 + b^2 = 2a^2 \rightarrow a = \frac{1}{\sqrt{2}} = b $$

so that

$$ |+1\rangle_B = \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} $$

and

$$ |-1\rangle_B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} $$

Second, we work with $\hat{C}$.

$$ \hat{C} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \rightarrow \begin{vmatrix} 1-\alpha & 2 \\ 2 & 1-\alpha \end{vmatrix} = 0 = (1-\alpha)(1-\alpha) - 4 \rightarrow \alpha^2 - 2\alpha - 3 = 0 \rightarrow \alpha = 3, -1 $$ = eigenvalues

To find the eigenvectors we then proceed as follows:

$$ \alpha = +3 $$

$$ \rightarrow \hat{C}|+3\rangle_C = 3|+3\rangle_C \rightarrow \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 3 \begin{pmatrix} a \\ b \end{pmatrix} $$

$$ \rightarrow a + 2b = 3a $$

$$ \rightarrow 2a + b = 3b $$

$$ \rightarrow b = a \quad , \quad a^2 + b^2 = 1 $$

$$ \rightarrow a = \frac{1}{\sqrt{2}} = b $$

so that

$$ |+3\rangle_C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} $$

and

$$ |-1\rangle_C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} $$

A strange thing has happened. Both operators $\hat{B}$ and $\hat{C}$ have the same eigenvectors (different eigenvalues).

It turns out that the following important property is true:

if two operators commute, then the share a common set of eigenvectors

Special Matrices (operators)

Two types of operators will dominate our discussion of quantum

mechanics.

Hermitian : $\hat{A} = \hat{A}^{+} = \hat{A}^{*Tr}$ or $A_{ij} = A_{ji}^{*}$

Unitary : $\hat{A}^{-1} = \hat{A}^{+} = \hat{A}^{*Tr}$ or $(A^{-1})_{ij} = A_{ji}^{*}$

More Mathematical Details

Definition : A set of vectors is said to be **linearly independent** if a linear relation of the form

$$\sum_{k=1}^{n} c_k |k\rangle = |\mathrm{O}\rangle$$

implies that $c_k = 0$ for **all** k; otherwise the set of vectors is **linearly dependent**.

If a set of vectors is linearly dependent, then we can express a member of the set as a linear combination of the other members of the set.

Examples:

(1) Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = |\mathrm{O}\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the only solution is

$$a_1 = a_2 = a_3 = 0$$

(2) Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly independent since

$$a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 \\ a_1 - a_2 \\ a_3 \end{pmatrix} = |\mathrm{O}\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the solution is

$$a_1 + a_2 = a_1 - a_2 = a_3 = 0 \quad \text{or} \quad a_1 = a_2 = a_3 = 0$$

(3) Consider the set of vectors(3-tuples in this case)

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

This set is linearly dependent since

$$a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix} + \begin{pmatrix} a_2 \\ -a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 2a_3 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 + 2a_3 \\ a_1 - a_2 \\ a_1 + a_3 \end{pmatrix} = |\mathrm{O}\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

implies that the solution is

$$a_1 + a_2 + 2a_3 = a_1 - a_2 = a_1 + a_3 = 0 \quad \text{or} \quad a_3 = -a_1 \ , \ a_2 = a_1$$

and we have (for example)

$$|1\rangle = -|2\rangle + |3\rangle$$

We say that an infinite set of vectors is linearly independent if **every finite subset** is linearly independent.

Definition : The maximum number of linearly independent vectors in a space V is called the **dimension** of the space $\dim(V)$.

Definition : A set of vectors $\{|k\rangle \, , \, k = 1,2,3,.....,n\}$ **spans** the space if **every** vector $|Q\rangle$ in the space can be written as a linear combination of vectors in the set

$$\sum_{k=1}^{n} q_k |k\rangle = |Q\rangle$$

This linear combination, which is given by the coefficients $q_k \, , k = 1,2,.....,n$, is **unique**.

Definition : A set of vectors is a **basis** for the space if it is a linearly independent set **and** spans the space, that is, if $\dim(V) = m$, a set of m linearly independent vectors is called a **basis** on V.

The set of vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is the maximal set of linearly independent vectors since any other vector $|g\rangle$ in the space can always be written as a linear combination of them as

$$|g\rangle = a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

Therefore the dimension of this vector space is 3. This set of vectors is also a basis. The basis is **not unique** since the set of linearly independent vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

also spans the space, i.e.,

$$|g\rangle = c_1|1\rangle + c_2|2\rangle + c_3|3\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 + 2c_3 \\ c_1 - c_2 \\ c_3 \end{pmatrix}$$

implies that

$$c_1 = a_1 + a_2 - 2a_3 \quad , \quad c_2 = a_1 - 2a_3 \quad , \quad c_3 = a_3$$

and, thus, this set is also a basis. Clearly, a basis spans the whole of V.

Example: In the space of 3-tuples, a basis is represented by the three vectors

$$|1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad |2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

so that an arbitrary vector in the space

$$|g\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

can be written

$$|g\rangle = a_1|1\rangle + a_2|2\rangle + a_3|3\rangle = \begin{pmatrix} a_1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

so that $a_1, a_2, and\ a_3$ are the components.

Gram-Schmidt Orthogonalization Process

An orthonormal basis set for an n-dimensional vector space can always be constructed from any set of n linearly independent vectors using the Gram-Schmidt orthogonalization method.

Suppose that we have a set of n linearly independent vectors $|\alpha_i\rangle$, $i = 1,2,.....,n$ that are not a mutually orthonormal set. We can construct a mutually orthonormal set $|\beta_i\rangle$, $i = 1,2,.....,n$ using the following steps:

(1) let $|\beta_1\rangle = |\alpha_1\rangle$

(2) let $|\beta_2\rangle = |\alpha_2\rangle + a_1|\beta_1\rangle$ where we choose a_1 such that $\langle\beta_1|\beta_2\rangle = 0$

(3) this gives

$$\langle\beta_1|\beta_2\rangle = 0 = \langle\beta_1|\alpha_2\rangle + a_1\langle\beta_1|\beta_1\rangle$$

$$a_1 = -\frac{\langle\beta_1|\alpha_2\rangle}{\langle\beta_1|\beta_1\rangle}$$

Now proceed by **induction**.

Suppose we have constructed k mutually orthogonal vectors $|\beta_i\rangle$, $i = 1,2,.....,k$. If we let

$$|\beta_{k+1}\rangle = |\alpha_k\rangle + \sum_{j=1}^{k} a_j|\beta_j\rangle$$

with

$$a_j = -\frac{\langle\beta_j|\alpha_{k+1}\rangle}{\langle\beta_j|\beta_j\rangle}$$

then we have $\langle\beta_j|\beta_{k+1}\rangle = 0$ for $j = 1,2,....,k$. These steps are repeated until we have n mutually orthogonal vectors. We then normalize them to 1 and create an orthonormal set.

For example, suppose we have the set

$$|\alpha_1\rangle = \begin{pmatrix}1\\1\\0\end{pmatrix} \quad , \quad |\alpha_2\rangle = \begin{pmatrix}0\\1\\1\end{pmatrix} \quad , \quad |\alpha_3\rangle = \begin{pmatrix}1\\0\\1\end{pmatrix}$$

These vectors are not orthonormal.

(1) let

$$|\beta_1\rangle = |\alpha_1\rangle = \begin{pmatrix}1\\1\\0\end{pmatrix} \quad , \quad \langle\beta_1|\beta_1\rangle = 2$$

(2) let

$$|\beta_2\rangle = |\alpha_2\rangle + a_1|\beta_1\rangle$$

with

$$a_1 = -\frac{\langle\beta_1|\alpha_2\rangle}{\langle\beta_1|\beta_1\rangle} = -\frac{1}{2}$$

and thus

$$|\beta_2\rangle = |\alpha_2\rangle - \frac{1}{2}|\alpha_1\rangle = \frac{1}{2}\begin{pmatrix}-1\\1\\2\end{pmatrix} \quad , \quad \langle\beta_2|\beta_2\rangle = \frac{3}{2} \quad , \quad \langle\beta_1|\beta_2\rangle = 0$$

(3) let

$$|\beta_3\rangle = |\alpha_3\rangle + a_1|\beta_1\rangle + a_2|\beta_2\rangle$$

with

$$a_1 = -\frac{\langle\beta_1|\alpha_3\rangle}{\langle\beta_1|\beta_1\rangle} = -\frac{1}{2} \quad \text{and} \quad a_2 = -\frac{\langle\beta_2|\alpha_3\rangle}{\langle\beta_2|\beta_2\rangle} = -\frac{1}{3}$$

and thus

$$|\beta_3\rangle = \frac{2}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad , \quad \langle\beta_3|\beta_3\rangle = \frac{4}{3} \quad , \quad \langle\beta_1|\beta_3\rangle = 0 \quad , \quad \langle\beta_2|\beta_3\rangle = 0$$

We normalize the vectors by dividing by their respective **norms**,

$$|\gamma_i\rangle = \frac{|\beta_i\rangle}{\|\beta_i\rangle\|} = \frac{|\beta_i\rangle}{|\langle\beta_i|\beta_i\rangle|^{1/2}}$$

The orthonormal set is then

$$|\gamma_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \; , \; |\gamma_2\rangle = \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \; , \; |\gamma_3\rangle = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

Examples - Functions of Operators

Suppose that we have the eigenvector/eigenvalue equations for a self-adjoint operator

$$\hat{A}|k\rangle = a_k|k\rangle \quad , \quad k = 1,2,....,N$$

We then assume that

$$f(\hat{A})|k\rangle = f(a_k)|k\rangle \quad , \quad k = 1,2,....,N$$

for the eigenvectors.

We can show that this works for polynomials and power series as follows:

$$|\psi\rangle = \sum_{k=1}^{N} |k\rangle\langle k|\psi\rangle$$

$$\hat{A}|\psi\rangle = \hat{A}\sum_{k=1}^{N} |k\rangle\langle k|\psi\rangle = \sum_{k=1}^{N} \hat{A}|k\rangle\langle k|\psi\rangle = \sum_{k=1}^{N} a_k|k\rangle\langle k|\psi\rangle = \left(\sum_{k=1}^{N} a_k|k\rangle\langle k|\right)|\psi\rangle$$

$$\rightarrow \hat{A} = \sum_{k=1}^{N} a_k|k\rangle\langle k| \rightarrow \text{spectral resolution of the operator}$$

Now define the **projection** operator

$$\hat{P}_k = |k\rangle\langle k| \rightarrow \hat{P}_k\hat{P}_j = \hat{P}_k\delta_{kj}$$

We then have

$$\hat{A} = \sum_{k=1}^{N} a_k|k\rangle\langle k| = \sum_{k=1}^{N} a_k\hat{P}_k$$

or any operator is represented by a sum over its eigenvalues and

corresponding projection operators.

We then have

$$\hat{A}^2 = \left(\sum_{k=1}^{N} a_k \hat{P}_k\right)\left(\sum_{j=1}^{N} a_j \hat{P}_j\right) = \sum_{k,j=1}^{N} a_k a_j \hat{P}_k \hat{P}_j = \sum_{k,j=1}^{N} a_k a_j \hat{P}_k \delta_{kj} = \sum_{k=1}^{N} a_k^2 \hat{P}_k$$

$$\rightarrow \hat{A}^n = \sum_{k=1}^{N} a_k^n \hat{P}_k$$

Therefore, for

$$f(x) = \sum_{n=1}^{N} q_n x^n$$

we have

$$f(\hat{A}) = \sum_{n=1}^{N} q_n \hat{A}^n = \sum_{n=1}^{N} q_n \sum_{k=1}^{N} a_k^n \hat{P}_k = \sum_{k=1}^{N}\left(\sum_{n=1}^{N} q_n a_k^n\right)\hat{P}_k = \sum_{k=1}^{N} f(a_k)\hat{P}_k$$

This says that, in general, we have

$$f(\hat{A})|\psi\rangle = f(\hat{A})\sum_{k=1}^{N}|k\rangle\langle k|\psi\rangle = \sum_{k=1}^{N} f(\hat{A})|k\rangle\langle k|\psi\rangle = \sum_{k=1}^{N} f(a_k)|k\rangle\langle k|\psi\rangle = \left(\sum_{k=1}^{N} f(a_k)|k\rangle\langle k|\right)|\psi\rangle$$

$$\rightarrow f(\hat{A}) = \sum_{k=1}^{N} f(a_k)|k\rangle\langle k| \rightarrow \text{spectral resolution of a function of an operator}$$

Numerical example:

$$\hat{A} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

has eigenvalues $7, 1$ with eigenvectors

$$|7\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, |1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

This gives

$$\hat{P}_7 = |7\rangle\langle 7| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \hat{P}_1 = |1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and therefore

$$\hat{A} = 7\hat{P}_7 + \hat{P}_1 = \frac{7}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

$$\hat{A}^2 = 7^2\hat{P}_7 + \hat{P}_1 = \frac{7^2}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 25 & 24 \\ 24 & 25 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$$

$$\log\hat{A} = \log(7)\hat{P}_7 + \log(1)\hat{P}_1 = \frac{\log(7)}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\sqrt{\hat{A}} = \sqrt{7}\hat{P}_7 + \hat{P}_1 = \frac{\sqrt{7}}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \sqrt{7}+1 & \sqrt{7}-1 \\ \sqrt{7}-1 & \sqrt{7}+1 \end{pmatrix}$$

Clearly we then have

$$\log \hat{A}|7\rangle = \frac{1}{\sqrt{2}}\frac{\log(7)}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix} = \log(7)\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix} = \log(7)|7\rangle$$

$$\log \hat{A}|1\rangle = \frac{1}{\sqrt{2}}\frac{\log(7)}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix} = \frac{1}{\sqrt{2}}\frac{\log(7)}{2}\begin{pmatrix}0\\ 0\end{pmatrix} = 0 = \log(1)|1\rangle$$

as expected.

The Dirac δ-Function

History

In the development of quantum mechanics by P. Dirac , the following sequence of ideas occurred (as we shall describe later).

(1) Observable = measurable quantity <--> Hermitian operator
(2) Physical states are linear combinations of eigenvectors
[requires complete orthonormal basis]
(3) Possible measurements are represent by the eigenvalues
[must be real numbers]

$$\hat{O}|\lambda\rangle = \lambda|\lambda\rangle \quad \text{where} \quad |\lambda\rangle = \text{eigenvector} \quad \text{and} \quad \lambda = \text{eigenvalue}$$

(4) Some observables have a discrete spectrum (finite number or denumerably infinite number)

$\langle\lambda'|\lambda\rangle = \delta_{\lambda'\lambda}$ for different eigenvalues this is a well-defined statement

(5) Other observables have a continuous spectrum (non-denumerably infinite number of eigenvalues).

For example , the position operator $\hat{X}$, which we will talk about later, is such that

$$\hat{X}|x\rangle = x|x\rangle \quad \text{where} \quad |x\rangle = \text{eigenvector} \quad \text{and} \quad x = \text{eigenvalue}$$

We now ask the question, what is $\langle x'|x\rangle$? Dirac assumed that

$$\langle x'|x\rangle \equiv \delta_{x'x} \equiv \delta(x - x')$$

where

$$\int_{-\infty}^{\infty}\delta(x - x')dx' = 1$$

$$\delta(x - x') = 0 \quad \text{if} \quad x' \neq x$$

$$\int_{-\infty}^{\infty} f(x')\delta(x - x')dx' = f(x)$$

This gives the correct physical theory in the sense that all predictions agree with experiment (1929).

Eventually (1960) the mathematicians, who initially vehemently disputed Dirac's assumption of this new "function", caught up to the physicists and proved all of its properties in the Theory of Distributions.

Here is how Dirac(in his own words) introduced this function.

Our work in Section 10 led us to consider quantities involving a certain kind of infinity. To get a precise notation for dealing with these infinities, we introduce a quantity $\delta(x)$ *depending on a parameter x satisfying the conditions*

$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

$$\delta(x) = 0 \text{ for } x \neq 0$$

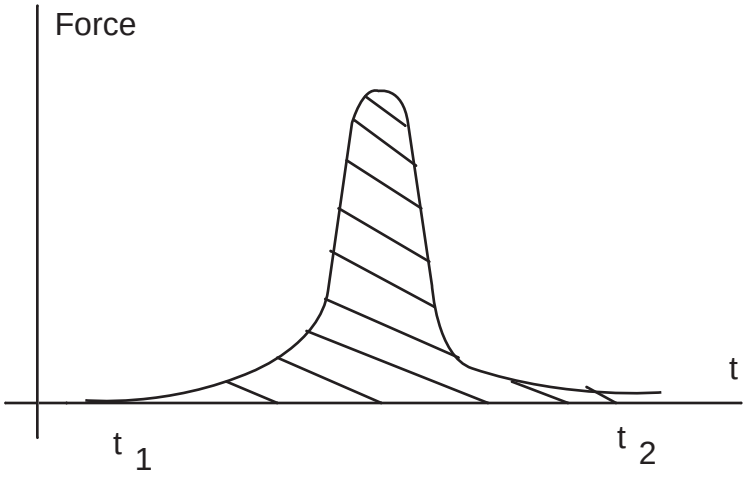
To get a picture of $\delta(x)$*, take a function of the real variable x which vanishes everywhere except inside a small domain, of length* ε *say, surrounding the origin x = 0, and which is so large inside this domain that its integral over the domain is unity. The exact shape of the function inside this domain does not matter, provided there are no unnecessary wild variations (for example provided the function is always of order* ε^{-1}*). Then in the limit* $\varepsilon \to 0$ *this function will go over into* $\delta(x)$*.*

$\delta(x)$ *is not a function of x according to the usual mathematical definition of a function, which requires a function to have a definite value for each point in its domain, but is something more general, which we may call an "improper function" to show up the difference from a function defined by the usual definition. Thus* $\delta(x)$ *is not a quantity which can be generally used in mathematical analysis like an ordinary function, but its use must be confined to certain simple types of expression for which it is obvious that no inconsistency can arise.*

Do such functions naturally appear in theoretical physics? These functions are infinitely discontinuous. They are **singular functions**.

We can get a physical feel for their properties by considering the concept of an **impulse**.

A standard impulse is described by a function of the form

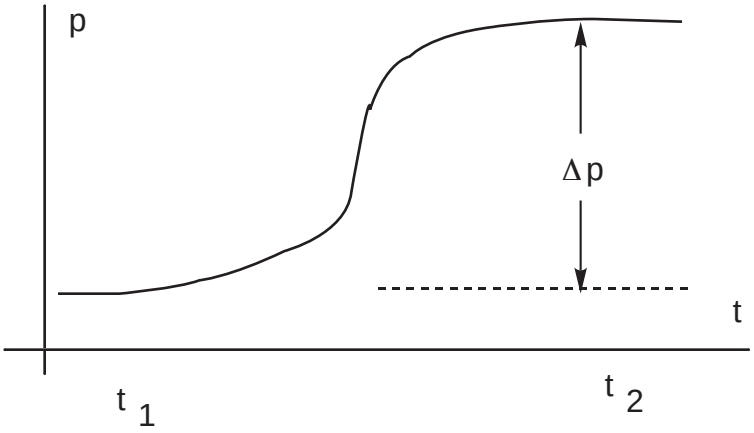


where

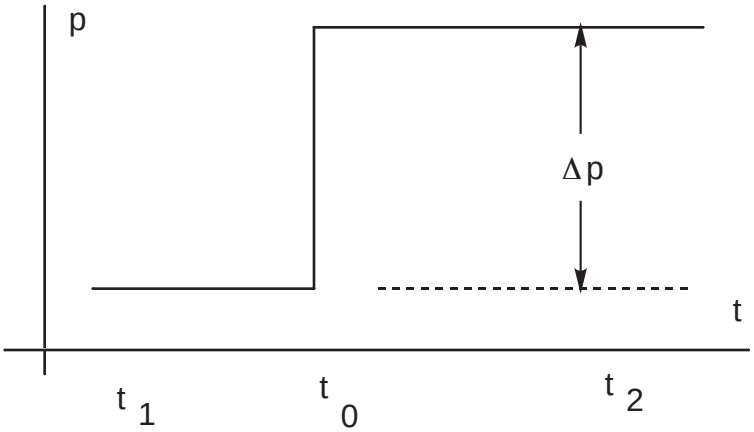
$$\text{Impulse} = \int_{-\infty}^{\infty} F(t)dt = \int_{t_1}^{t_2} F(t)dt = \text{shaded area}$$

$$= \int_{t_1}^{t_2} m_0 \frac{dv}{dt} dt = \int_{t_1}^{t_2} m_0 dv = m_0(v(t_2) - v(t_1)) = \Delta(m_0 v) = \Delta p$$

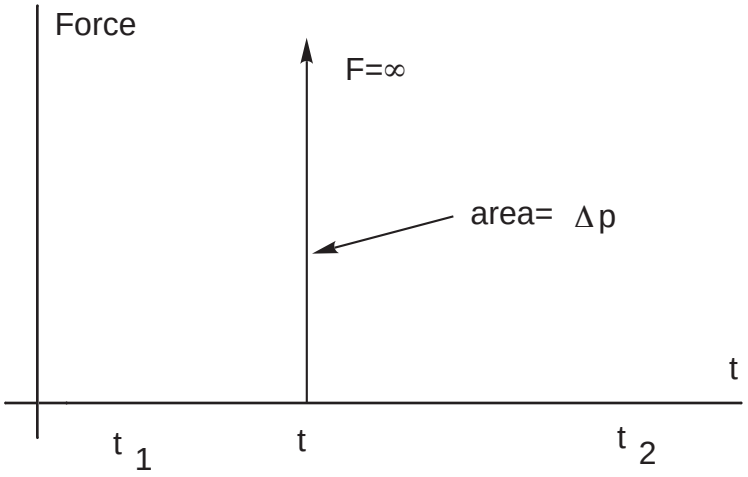
The actual transfer(gradual) of momentum with respect to time looks like



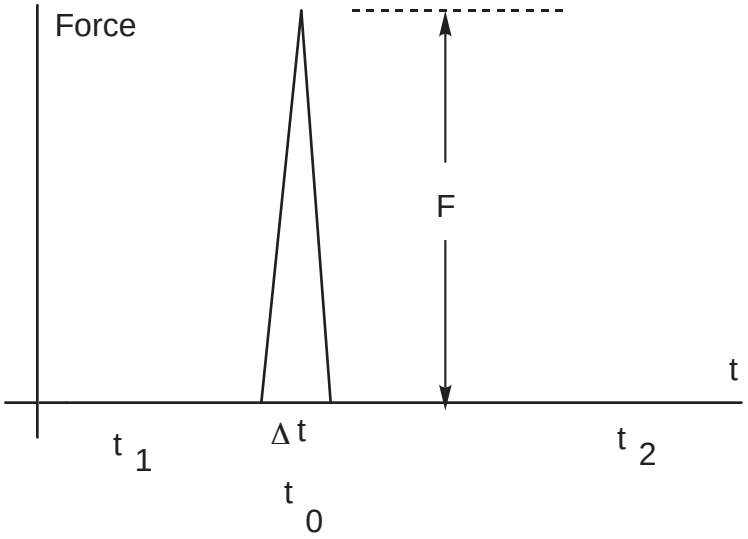
An idealized impulse, on the other hand, transfers Δp instantaneously.



If we regard this as a regular function, then we would need to have



In the real world we have



where in the limit $\Delta t \to 0, F \to \infty$

$$\int_{t_0^-}^{t_0^+} F(t)dt = \Delta p$$

For the ideal impulse we write the forcing function as

$$F(t) = \Delta p \delta(t - t_0)$$

where

$$\int_{t_0^-}^{t_0^+} \delta(t - t_0)dt = 1$$

where, in general, we define

$$\delta(t) = 0 \quad \text{for} \quad t \neq 0$$

$$\int_{t-}^{t+} \delta(t)dt = \begin{cases} 1 & t- < 0 < t+ \\ 0 & \text{otherwise} \end{cases}$$

Examples from Physics

The standard idealizations of point masses and point charges are describable by Dirac delta-functions.

For mass in physical equations we define $\rho_m(\vec{r})$ = mass per unit volume. We assume that this is a continuous function of the position vector.

Similarly, for charge we define $\rho_Q(\vec{r})$ = charge per unit volume.

For a point mass located at (x_0, y_0, z_0) we write $\rho_m(\vec{r}) = m_0\delta(x - x_0)\delta(y - y_0)\delta(z - z_0)$ and

$$\int_V d\tau \rho_m(\vec{r}) = \begin{cases} m_0 & \text{if } (x_0, y_0, z_0) \text{ in V} \\ 0 & \text{otherwise} \end{cases}$$

and similarly for point charges.

Definition of $\delta(t)$

There are two ways to define delta functions:

[1] Use generalized function theory. This corresponds to defining the delta function by its behavior inside integrals.

It is important to realize the delta function is meaningless outside of an integral. When we state properties, we will always be indicating behavior that corresponds to using the delta function inside an integral.

[2] Define the delta function as the limit of an infinite sequence of continuous functions (can be very tricky).

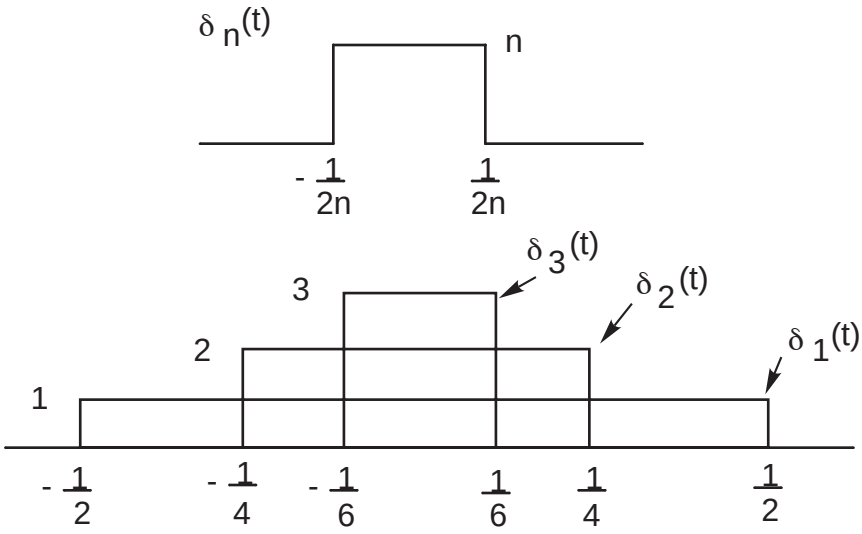
$$\delta(t) = \lim_{n\to\infty} \delta_n(t)$$

where $\delta_n(t)$ is finite for **all** t .

Examples:

(1)

$$\delta_n(t) = \begin{cases} n & -\frac{1}{2n} < t < \frac{1}{2n} \\ 0 & \text{otherwise} \end{cases}$$



For any n

$$\int_{-\infty}^{\infty} dt\delta_n(t) = 1$$

In the limit $n \to \infty$, $\delta_n(t) = 0$ for all t, except $t = 0$.

Let us look at this one a different way. The simplest such object is a rectangular spike,

$$D_\varepsilon(x - x') = \begin{cases} 1/\varepsilon & |x - x'| \le \varepsilon/2 \\ 0 & otherwise \end{cases}$$

Note that the area under the spike is = 1 independent of ε. We then have

$$\int_{-L}^{L} f(x')D_\varepsilon(x - x')dx' = \int_{x-\varepsilon/2}^{x+\varepsilon/2} f(x')\frac{1}{\varepsilon}dx'$$

We now substitute a Taylor expansion for $f(x')$ about the point $x' = x$.

$$f(x') = f(x) + (x' - x)f'(x) + \frac{1}{2!}(x' - x)^2 f''(x) + ...$$

to get

$$\int_{-L}^{L} f(x')D_\varepsilon(x - x')dx' = \frac{1}{\varepsilon}[f(x)\int_{x-\varepsilon/2}^{x+\varepsilon/2} dx' + f'(x)\int_{x-\varepsilon/2}^{x+\varepsilon/2}(x' - x)dx' + \frac{1}{2!}f''(x)\int_{x-\varepsilon/2}^{x+\varepsilon/2}(x' - x)^2 dx' + ...]$$

which gives

$$\int_{-L}^{L} f(x')D_\varepsilon(x - x')dx' = \frac{1}{\varepsilon}[f(x)\varepsilon + f'(x)(0) + \frac{1}{2!}f''(x)\frac{\varepsilon^3}{4} + ...] = f(x) + \frac{1}{8}f''(x)\varepsilon^2 + ...$$

Taking the limit as $\varepsilon \to 0$, we get $f(x)$. This implies that

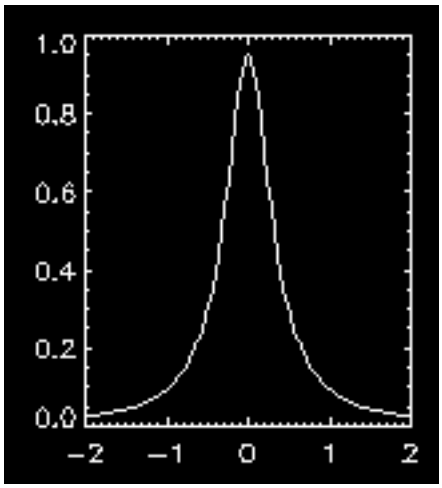
$$\lim_{\varepsilon \to 0} D_\varepsilon(x' - x) = \delta(x' - x)$$

This object is not a well-defined mathematical function, which must

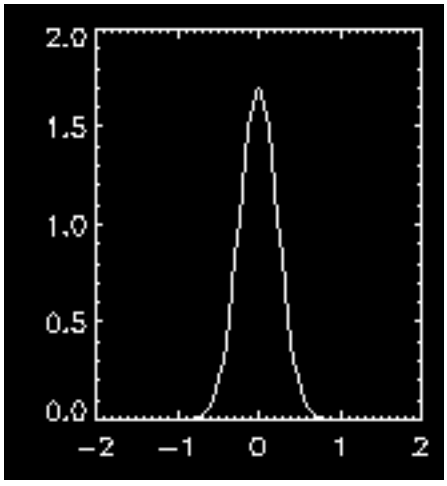
have a definite value at every point where it is defined. Dirac called it an "improper" function. This means that it only makes sense when it appears within an integral and then it has a well-defined effect. It is now known to be a "distribution".

Other Sequences

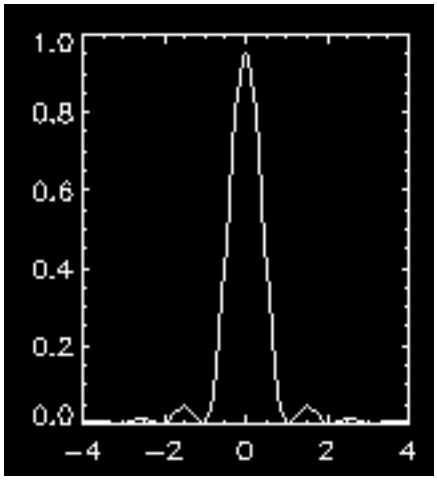
[1] $\delta_n(t) = \dfrac{\dfrac{n}{\pi}}{1 + n^2 t^2}$



[2] $\delta_n(t) = \dfrac{n}{\sqrt{\pi}} e^{-n^2 t^2}$ Gaussian function



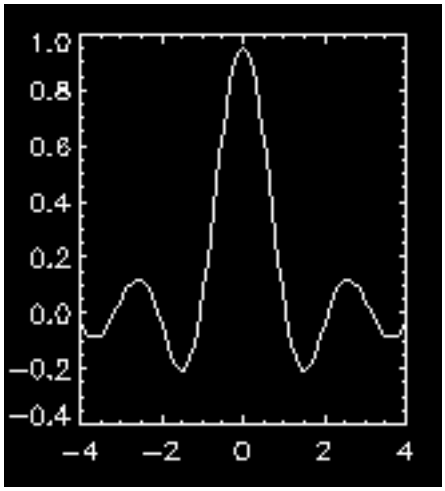
[3] $\delta_n(t) = \dfrac{\sin^2 nt}{n\pi t^2}$ Sinc Squared function



Using Integrals to define the Delta Function

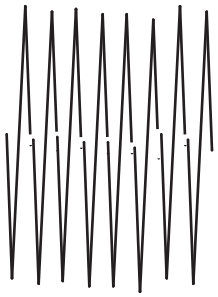
We use the definition $\int_{t_0^-}^{t_0^+} dt\delta(t-t_0)f(t) = f(t_0)$, which is called the **sifting** or **filter** integral. It filters out a single value!!

Example: The following limit sequence can only be understood via the integral definition. Consider the **Sinc** sequence $\delta_n(t) = \frac{\sin nt}{\pi t}$, Sinc function



As $n \to \infty$, $\delta_n(t)$ does not $\to 0$ for all $t \neq 0$. It does not become a function with an infinitely narrow and infinitely tall peak.

As $n \to \infty$, $\delta_n(t)$ oscillates infinitely fast except at $t = 0$. When multiplied by a continuous function in an integral, the contribution to the integral is zero everywhere except in the region near $t = 0$.



x

continuous function

fast oscillations

that is,

$$\int_{0-}^{0+} dt \left[\lim_{n\to\infty} \frac{\sin nt}{\pi t} \right] f(t) = f(0) \quad (\text{as} \quad n \to \infty)$$

and therefore

$$\lim_{n\to\infty} \frac{\sin nt}{\pi t} = \delta(t)$$

Complicated Arguments

(1) Consider

$$\int_{0-}^{0+} dt'\, \delta(-t')f(t') = -\int_{-0-}^{-0+} dt'\, \delta(t')f(-t') = -\int_{0+}^{0-} dt'\, \delta(t')f(-t') = \int_{0-}^{0+} dt'\, \delta(t')f(-t') = f(0)$$

which implies that $\delta(-t) = \delta(t)$ in the sense of using integrals, i.e.,

$$f(0) = \int \delta(-t) f(t) dt = \int \delta(t) f(t) dt$$

(2) Consider

$$\int_{0-}^{0+} dt'\, \delta(at')f(t') = \frac{1}{|a|} \int_{0-/a}^{0+/a} dt''\, \delta(t'')f\left(\frac{t''}{a}\right) = \frac{1}{|a|} f(0)$$

which implies that $\delta(at) = \frac{1}{|a|}\delta(t)$ in the sense of using integrals, i.e.,

$$\frac{1}{|a|} f(0) = \int \delta(at) f(t) dt = \frac{1}{|a|} \int \delta(t) f(t) dt$$

(3) Consider

$$\int_{0-}^{0+} dt'\, t'\, \delta(t')f(t') = f(0)(0) = 0$$

If $f(0) \neq 0$, then this implies that

$$t\delta(t) = 0$$

(4) Consider

$$\int_{-L}^{L} f(x)\delta(x^2-b^2)dx = \int_{-L}^{0} f(x)\delta(x^2-b^2)dx + \int_{0}^{L} f(x)\delta(x^2-b^2)dx$$

Now near $x=-b$, we can write

$$x^2-b^2 = -2b(x+b)$$

and near $x=b$, we can write

$$x^2-b^2 = 2b(x-b)$$

therefore we have

$$\int_{-L}^{L} f(x)\delta(x^2-b^2)dx = \int_{-L}^{0} f(x)\delta(-2b(x+b))dx + \int_{0}^{L} f(x)\delta(2b(x-b))dx$$

$$= -\frac{1}{2|b|} f(-b) + \frac{1}{2|b|} f(b)$$

or

$$\delta(x^2-b^2) = \frac{[\delta(x-b)+\delta(x+b)]}{2|b|}$$

(5) Now consider a function $g(x)$ which has a single zero at $x=x_0$. This implies that near x_0 we have

$$g(x) = (x-x_0)g'(x_0)$$

In this case we can write

$$\int_{-\infty}^{\infty} f(x)\delta(g(x))dx = \int_{-\infty}^{\infty} f(x)\frac{\delta(x-x_0)}{|g'(x_0)|}dx = \frac{f(x_0)}{|g'(x_0)|}$$

and generalizing, if $g(x)$ has N zeroes at $x=x_i$, $i=1,2,3,.....,N$, then near each zero we can write

$$g(x) = (x-x_i)g'(x_i)$$

and we have

$$\delta(g(x)) = \sum_{i=1}^{N} \frac{\delta(x-x_i)}{|g'(x_i)|}$$

Examples :

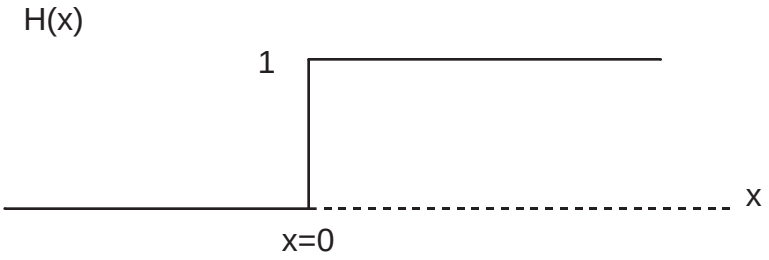
$$\int_{-\infty}^{\infty} e^{-x}\delta(x^2-a^2)dx = \int_{-\infty}^{\infty} e^{-x}\frac{\delta(x-a)+\delta(x+a)}{2a}dx = \frac{e^{-a}+e^{a}}{2a} = \frac{\cosh(a)}{a}$$

$$\int_{-\infty}^{\infty} e^{-x^2}\delta(\sin x)dx = \int_{-\infty}^{\infty} e^{-x^2}\sum_{n=-\infty}^{\infty}\frac{\delta(x-n\pi)}{|\cos n\pi|}dx = \sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2}\delta(x-n\pi)dx = \sum_{n=-\infty}^{\infty} e^{-(n\pi)^2}$$

Integrals and Derivatives of $\delta(t)$

We first define a new function as an integral of the delta function, namely, the **Heaviside Unit Step Function**. Let

$$H(x) = \int_{-\infty}^{x} dt\,\delta(t) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$

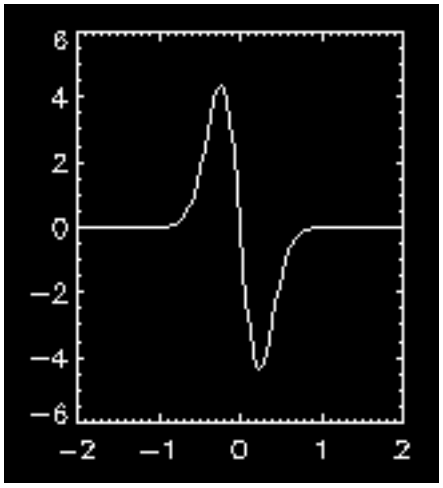


$H(0)$ is not well-defined. However, since $\delta(t) = \delta(-t)$ (an even function), by convention we define $H(0) = \frac{1}{2}$.

Derivative of the delta function can be defined in terms of the limit sequences. Consider

$$\frac{d\delta(t)}{dt} = \lim_{n\to\infty} \frac{d\delta_n(t)}{dt} = \lim_{n\to\infty} \frac{d}{dt}\left[\frac{n}{\sqrt{\pi}} e^{-n^2t^2}\right] = \lim_{n\to\infty}\left[-\frac{2n^3t}{\sqrt{\pi}} e^{-n^2t^2}\right] = \delta'(t)$$

It is called a **doublet** function because it looks like ($n = 3$)



where the peaks are infinitely high, infinitely narrow and infinitely close together in the limit.

We then have

$$\int_{0-}^{0+} dt f(t)\delta'(t) = f(t)\delta(t)\Big|_{t=0-}^{t=0+} - \int_{0-}^{0+} dt \frac{df(t)}{dt}\delta(t) = -\int_{0-}^{0+} dt \frac{df(t)}{dt}\delta(t) = -f'(0)$$

where we have used integration by parts. Thus, the doublet is the sifting function for the negative of the derivative.

We will use this formalism to describe the theory/experiments in quantum mechanics. These ideas about vectors/operators will be sufficient for us to develop a quantum theory. We will extend these mathematical ideas as needed.