The Mathematical Language of Quantum Mechanics

I. Probability and Statistics

We first 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

$$p_k = \frac{n_k}{N} \qquad (01)$$

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

$$0 \le p_k \le 1 \quad \text{for all } k \qquad (02a)$$

and

$$\sum_k p_k = 1 \qquad (02b)$$

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 will be able to, after making a very **large** number of measurements on these "**identical**" systems, calculate 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 carried out.

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(recall from earlier discussion):

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.

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

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

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

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

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

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

Arranging and Choosing

Suppose that you have 5 books on physics. In howmany ways can you arrange them on your bookshelf?

Any of the 5 can go on the left. This leaves 4 possibilities for the 2nd book and therefore by the multiplication rule there are 5 x 4 = 20 ways to put the 1st 2 books on your shelf. That leaves 3 choices for the 3rd book, yielding 5 x 4 x 3 = 60 ways of shelving the 1st 3 books. Then there are 2 possibilities for the penultimate book, and only 1 choice for the last book, so there are altogether

$$5 \text{ x } 4 \text{ x } 3 \text{ x } 2 \text{ x } 1 = 120 = 5!$$

ways of arranging them.

Incidently, in the course of showing this we have shown that the number of ways of arranging a selection of r books, $0 \le r \le 5$, is

$$5 \times ..... \times (5-r+1) = \frac{5!}{(5-r)!}$$

that is, we can choose the 1st in 5 ways, the 2nd in 5-1 ways, and so on, with the last chosen in 5-r+1 ways and then by the product rule, this gives $5 \times ..... \times (5-r+1)$ ways in total as stated above.

The conventional name for an ordering or arrangement is a **permutation.**

The number of permutations of r things from n things is then

$$n(n-1).....(n-r+1) = \frac{n!}{(n-r)!} \quad , \quad 0 \le r \le n$$

In this example involving books we naturally assumed that the books were all different. But, suppose that, for whatever strange reason, you happen to have 2 copies of the same book. They are unmarked and hence indistingusihable. How many different permutations of all 5 books are possible now? In the 120 arrangements calculated above there are 60 pairs in which each member of the pair is obtained by exchanging the positions of the 2 identical books. But these pairs are indistinguishable and therefore the same. So there are just 60 different permutations. Mathematically this corresponds to the formula

$$\frac{5!}{2!} = \frac{120}{2} = 60$$

We can generalize this result as follows. If there are n objects of which n_1 form one indistinguishable group, n_2 another, and so on up to n_r where

$$n_1 + n_2 + ... + n_r = n$$

then there are

$$M(n_1, n_2, ..., n_r) = \frac{n!}{n_1! n_2! ..... n_r!}$$

distinct permutations of these n objects.

In the example above we have

$$n_1 = 2, n_2 = n_3 = n_4 = 1$$

$$n_1 + n_2 + n_3 + n_4 = 2 + 1 + 1 = 1 = 5$$

and

$$M(2,1,1,1) = \frac{5!}{2!1!1!1!} = 60$$

Finally we then have the number of permutations of r indistinguishable things from n things is then

$$M(r, n-r) = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad , \quad 0 \le r \le n$$

This is called the **binomial coefficient.**

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

In fact, for a lottery where we must pick r different numbers from n possibilities (no repeats) there are

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

ways of choosing r different numbers from n possibilities (all equally likely). The probability that your single selection of r different numbers wins is therefore

$$p(winning) = \frac{1}{\binom{n}{r}} = \frac{1}{\#\ combinations(r,n)}$$

Suppose we are to pick the correct 6 numbers out of a selection of 49 numbers(no repeats). We buy 1 ticket. The number of possible combinations, however, is

$$\binom{49}{6} = \frac{49!}{6!(43)!} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{1 \times 2 \times 3 \times 4 \times 5 \times 6} = 13983816$$

so that the probability of winning is

$$p(winning) = \frac{1}{\binom{49}{6}} = 7.1511 \times 10^{-8}$$

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

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 v-number and return it to the box a total of M times. We denote by $v^{(i)}$ the v-value recorded on the i^{th} sampling, and we make the following two definitions:

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

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

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

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

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

To calculate this quantity, we must first calculate the deviation from the mean $v^{(i)} - \langle v \rangle$, of each v-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 Δv is the **square root** of the **mean** of the **squares** of the **deviations** of the $v^{(i)}$ values from $\langle v \rangle$. This quantity is also called the **rms dispersion**, since it clearly measures the extent to which the $v^{(i)}$ values are dispersed about $\langle v \rangle$.

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

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

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

or

$$\Delta v = \sqrt{\left\langle v^2 \right\rangle - \langle v \rangle^2} \tag{06}$$

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 $\left\langle v^2 \right\rangle = \langle v \rangle^2$ **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 $v^{(1)}, v^{(2)}, ..., v^{(M)}$ 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

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

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

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

Finally,

$$(\Delta h)^2 = \langle h^2 \rangle - \langle h \rangle^2 = 271.2$$
$$\Delta h = 16.5$$

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 $\langle v \rangle$ and Δv. Certainly a **complete** description of the expected results of a "multiple sampling" experiment requires the specification of **all** the numbers $(v_1, p_1), (v_2, p_2), (v_3, p_3), \ldots\ldots\ldots$. However, if we are asked to describe the results with **only two** numbers, we would evidently do well to state the values $\langle v \rangle$ and Δv. $\langle v \rangle$ is essentially a "**collective value**" for the set of v-numbers, while Δv (or the smallness thereof) provides a quantitative measure of the degree to which it is actually meaningful to so characterize the set of v-values by the **single** value $\langle v \rangle$.

Probability Concepts (more mathematical 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

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

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

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

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 $\{A_i\ ,\ \mathrm{i=1,2,....,n}\}$. 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

$$P(A|M)=p$$

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 $0\leq n_A\leq n$.

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

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

The sum $\sum_{\substack{\text{all possible} \\ \text{orderings}}}$ 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(as we saw earlier) 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

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

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

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

and

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

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

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

This change of variables is always possible.

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 More Useful Mathematics

· Taylor Series

Functions can be expanded in power series. Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ (a power series). 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(a **Taylor series** for $f(x)$ about the point $x = a$). 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

$$f(a) = a_0 \quad , \quad f'(a) = a_1 \quad , \quad \frac{1}{2!} f''(a) = a_2 \text{ and so on.....}$$

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.**

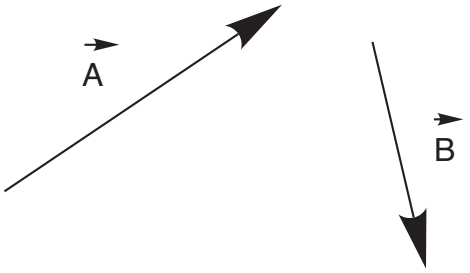
A vector has many levels of complexity and is a very abstract mathematical object. A vector is a mathematical(geometrical) object that is representable by two numbers in two dimensions, three numbers in three dimensions, and so on. One characterization is to specify its magnitude or length and orientation 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

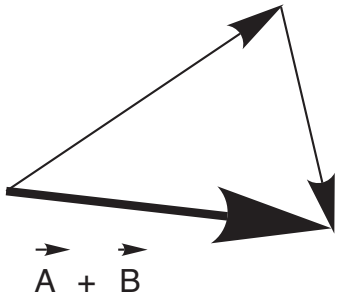
As we said, in ordinary space, we can represent a vector by a directed line segment(an arrow).

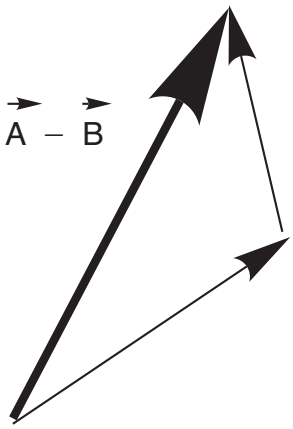
A straightforward property of a vector is multiplication of the vector by a scalar (a real number) $\vec{C} = \alpha \vec{A}$. In this case the magnitude of the vector changes and the direction stays the same (it might reverse if $\alpha < 0$).

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

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

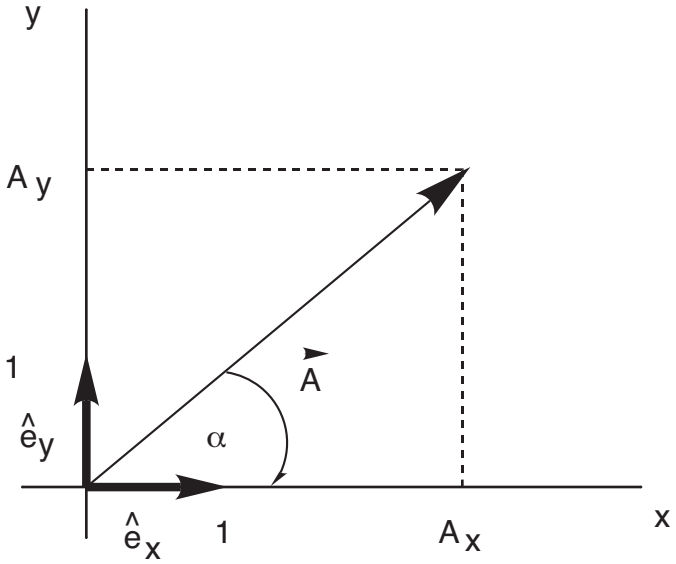
which is also a well-defined vector.

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 vectors shown below:



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

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

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

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

where

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

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

We now define

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

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:

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

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:

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

The inner product satisfies the following relations:

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

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

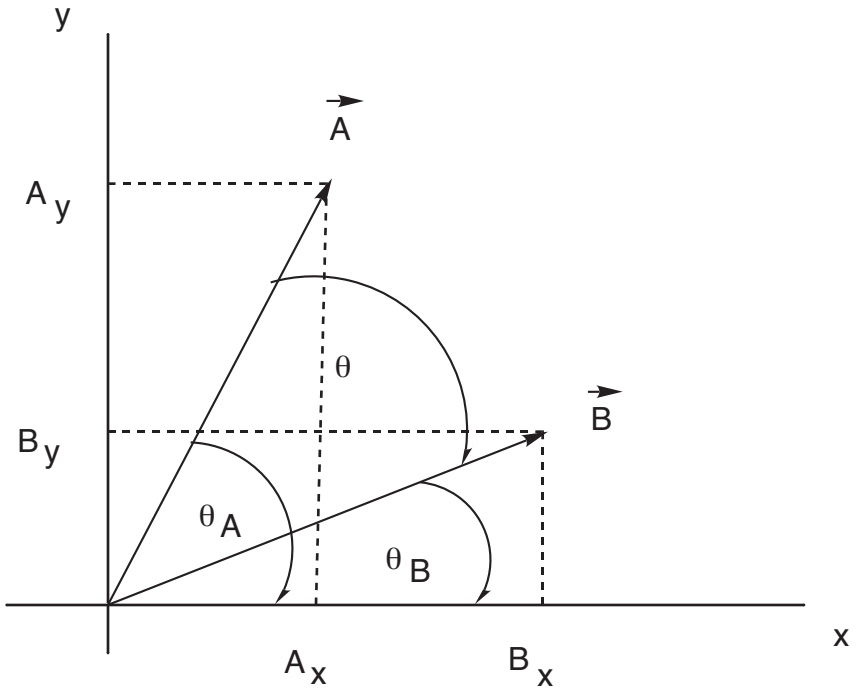
$$\vec{A} = A_x\hat{e}_x + A_y\hat{e}_y$$
$$\vec{B} = B_x\hat{e}_x + B_y\hat{e}_y$$
$$\vec{A}\cdot\vec{B} = (A_x\hat{e}_x + A_y\hat{e}_y)\cdot(B_x\hat{e}_x + B_y\hat{e}_y) \qquad (21)$$
$$= A_xB_x\hat{e}_x \cdot \hat{e}_x + A_xB_y\hat{e}_x \cdot \hat{e}_y + A_yB_x\hat{e}_y \cdot \hat{e}_x + A_yB_y\hat{e}_y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

We note that

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

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

Although we are deriving this result using two vectors in a plane, the result is completely general since two vectors always define a plane.



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 unit 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

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

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 given 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

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

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

so that

$$\hat{e}_1 \cdot \vec{A} = A_1\delta_{11} + A_2\delta_{12} + A_3\delta_{13} = A_1(1) + A_2(0) + A_3(0) = A_1 = 1 - component$$
$$\hat{e}_2 \cdot \vec{A} = A_1\delta_{21} + A_2\delta_{22} + A_3\delta_{23} = A_1(0) + A_2(1) + A_3(0) = A_2 = 2 - component$$
$$\hat{e}_3 \cdot \vec{A} = A_1\delta_{31} + A_2\delta_{32} + A_3\delta_{33} = A_1(0) + A_2(0) + A_3(1) = A_3 = 3 - component$$

or

$$\hat{e}_k \cdot \vec{A} = A_k = k^{th} - component$$

Therefore,

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

Thus, any vector can be written in terms of its components and the unit vectors.