4 Motion

Without exception, everything that happens involves motion. No natural phenomenon is more fundamental than motion. Greek philosophers in the fifth century BCE dubbed the study of motion "physics."

Modern physics theories model an object's motion by describing how it is affected by *interactions* with other objects. These theories are all based on the following principles:

1. A physical object is an ensemble of interacting point-like particles.
2. Particles that do not interact move in straight lines at a constant rate (that is to say, at constant velocity).
3. An interaction is a physical relationship between two (and only two) particles.
4. A single interaction between two particles changes the motion of both particles.
5. Such an interaction transfers some substance-like (extensive) quantity from one particle to the other.
6. A macroscopic object, consisting as it does of point-like particles, interacts with another macroscopic object as if both were point-like particles located at their respective centers-of-mass.

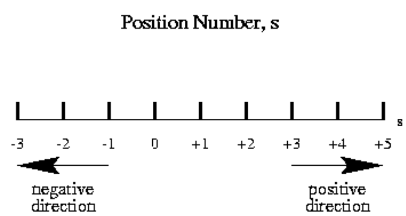
4.1 Rectilinear motion[1]: position and displacement

Ordinary motion is incredibly complex. Consider the three-dimensional motion of changing cloud shapes in the sky, or the bobbing motion of grass in the wind, or the turbulent fluid motions of air or water, or the spinning, curving motion of a thrown object, or the motion of a car on a winding road, body vibrating and wheels rotating. Just what is meant when a quantity is identified as one of these object's rate of motion (velocity)? Whatever the meaning, doing so focuses attention on perhaps the most interesting aspect of the motion by neglecting the infinity of complex details that, at the moment, is irrelevant. That is, fuzzy shapes and erratic swirls, bobbings, and vibrations are ignored while some quantity is abstracted from the complexity. It may even be (as in the case of a car) that the geometrical complexities of the object's path are ignored, and the motion is viewed as if it were along a straight line. Such abstractions from physical reality lie at the very core of science. Successful scientific inquiry almost invariably starts by examining the very simplest abstractions, building up to more complex situations as insights are deepened through further examination.

[1] Also referred to as "linear motion" or "one-dimensional motion," rectilinear motion is motion along a straight line that can therefore be described mathematically using only one spatial dimension.

To measure velocity, for example, requires reference to position. A numerical description of an object's position along a straight line is usually made with reference to a series of numerical markers. Denoting an *arbitrarily chosen* "origin" point by 0, laying out other points, $+1$, $+2$, -1, -2, etc., creates a coordinate scale. The size of the spacing between integer numbers along the line is also arbitrary; any standard unit of length available through national or international definition may be adopted. Along a road, the spacing is typically measured in kilometers, or, in the U.S., in miles. On a laboratory table, the unit is likely to be centimeters or millimeters. In a precision experiment, the units could be microns. Positions between those marked by integers are assigned appropriate fractional values. For example, at some instant of time a car on the road may be at position 7.49 on a scale determined by spacing in kilometers. Note that the integer spacing need *not* be uniform: position could still be described with numerical values at points on an irregular scale. Because of the simplicity desired in science and generally, a uniform scale is invariably chosen, one which others can easily reproduce, which is readily comparable with other physical situations, and which reveals the orderliness of what the scale is used to describe.

This kind of reference line or coordinate axis is sometimes called a "reference frame." To describe positions in two- or three-dimensional space, two or three such axes are erected, usually perpendicular to each other, to obtain two- or three-dimensional reference frames.



The choice of a reference frame's origin, axis orientation, and unit spacing is arbitrary, but not without consequences. A wise choice can simplify the analysis considerably. But sometimes, the choice has philosophical as well as physical implications. Does Earth move around the sun or does the sun move around Earth? Is it more fruitful to describe the motion of the moon as tracing a nearly circular path around Earth or a more complicated path around the sun?

The mathematical description of motion and other physical quantities requires the use of symbols to denote those quantities. Here, the symbol s denotes a generic position. Symbols with subscripts, for example, s_0, s_1, s_2, etc., or s_A, s_B, etc., or s_i and s_f, denote particular positions.[2] The subscripts allow one position to be distinguished from the others and to associate the position reading with a clock reading.

The symbol, s, represents position, *not* displacement (distance traversed). Thus, numbers such as $s_1 = +8.63$ km or $s_2 = -3.40$ km indicate locations of a car on the road, not distances traveled during an interval of time.

[2]The subscripts 1, 2, etc., have no intrisic association with specific position values; they refer to the first and second positions observed. Position subscripts correspond to the subscripts of the associated clock readings.

Motion is recognized when an object's position changes from instant to instant. Denoting a first, or initial, position by s_1 and a position at a later instant by s_2, the number $s_2 - s_1$ gives information about what is called "displacement" or "change of position." For example, with $s_2 = -3.40$ km and $s_1 = +8.63$ km, the car changes it position, or is displaced $s_2 - s_1 = -3.40 - 8.63 = -12.03$ km between instant 1 and instant 2, assuming the spacing between integer markers is in kilometers. The symbol Δs is shorthand for the number $s_2 - s_1$. That is,

$$\Delta s \equiv s_2 - s_1, \tag{1}$$

where Δs is called "displacement" or "change in position." The $\equiv$ symbol is read as "is defined by" or "is identical with" or "is equivalent to." This is not an equation in the sense that $y = 3x^2 - 2x + 1$ is an equation. Here, the relationship signifies that Δs is a name for $s_2 - s_1$. It has dimension $[\Delta s] = [L]^1$.

Δs has definite algebraic properties: it may be positive, negative, or zero depending on the numerical values and algebraic signs of s_2 and s_1.

1. **What does the algebraic sign of a given s mean, for example, what do the + and − signs in $s_1 = +8.63$ km or $s_2 = -3.40$ km mean?**

2. **What does a positive Δs mean? What does a negative Δs mean?**

3. **Describe or interpret the physical situation in which**

   (a) **s_2 and s_1 are both positive and Δs is positive;**

   (b) **s_2 and s_1 are both negative and Δs positive;**

   (c) **s_2 and s_1 are both negative and Δs negative; and**

   (d) **s_2 and s_1 are both positive but Δs negative.**

4. **Δs is a name for a difference between two positions, a displacement, or a change of position. Is this the same as the distance traveled between these two positions? Consider an object which moves back and forth along a line in either a regular or erratic manner.**

5. **Describe a case in which $\Delta s = 0$, but the distance traveled is not zero.**

Mathematical notes:

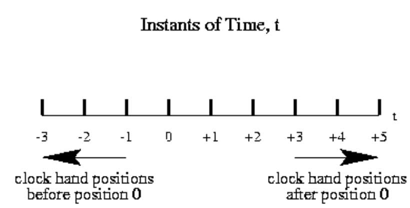
1. It is conventional to speak of the numbers increasing toward the right and decreasing toward the left regardless of whether the numbers are positive or negative. That is, the numbers increase from -3 to -1, for example, because $-1 > -3$. Consistent with this convention, the "positive direction" is identified as being toward the right, even if the positions are associated with negative values, and the "negative direction" as toward the left.

2. The symbol $|x|$, read "absolute value of x" or "magnitude of x," stands for the positive number associated with x. Thus, if x is positive, $|x| = x$, but $|x| = -x$ if x is negative. For example, $|+4| = 4$ and $|-4| = -(-4) = 4$.

4.2 Instants and intervals of time

Time interval is measured as arbitrarily as displacement (change of position). Fortunately there are fewer conventional units. The most widely used unit, the "second,"was originally defined as 1/86,400th part of a mean solar day, where a mean solar day is the average time between successive crossings by the sun of the meridian plane at a point on Earth. The average must be taken because this interval varies slightly but systematically during the course of the year, with the systematic variation of Earth's orbital velocity around the sun. Presently, it is defined as the time duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground-state of the caesium-133 atom.

Measuring intervals of time, then, amounts to counting a regular variation or oscillation. Clocks count oscillations of a balance wheel or a pendulum or a vibrating quartz crystal. Instants of time can therefore be numbered in sequence and produce a number line as for position.



By a convention that accords with the intuitive sense that time flows forward, succeeding instants are labeled with increasingly positive numbers. As is the case when denoting the position origin $s = 0$, the instant denoted $t = 0$ is an arbitrary choice. Instants labeled with negative numbers simply refer to instants before the arbitrarily chosen instant $t = 0$. They do not imply that time runs backward. The instant $t = -3$ s refers to an instant 3 seconds before the instant labeled $t = 0$; Time proceeded in the sequence -3 s, -2 s, -1 s, 0, $+1$ s, etc.

Analogously with the position symbols s, the symbols t, t_0, t_1, t_2, etc., or t_A, t_B, etc., or t_i and t_f,[3] are used for "instants of time," *not intervals* of time. An interval of time, like a displacement, is a difference between two instants:

$$\Delta t \equiv t_2 - t_1 \tag{2}$$

where $t_2 > t_1$, so that Δt is positive, the "natural" ordering of clock readings. The difference between two clock readings can be negative, but only if the later

[3]The subscripts 1, 2, etc., have no intrisic association with specific clock readings; they refer to the first and the second instants observed. By the convention that time flows forward, $t_2 > t_1$; thus, because a final instant has to come after an initial instant, $t_f > t_i$. On the other hand, t_0 is usually reserved for the arbitrarily chosen $t = 0$. Position subscripts correspond to the subscripts of the associated clock readings. In particular, s_0 is the position of the object at the arbitrarily chosen $t_0 = 0$.

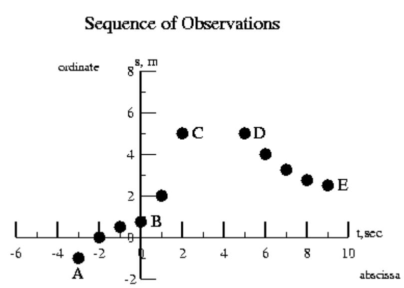
time is subtracted from an earlier time. The difference refers to how long before the second instant (in the past) something happened. There's no implication of time flowing backwards. The dimension of a time interval is $[\Delta t] = [T]^1$.

4.3 Displacement and corresponding time interval

Observing that, at instant t_1, an object is located at position s_1, and, at instant t_2, the same object is located at position s_2 suggests that, by associating each position measurement with the simultaneous clock reading, $(t_1,\ s_1)$ and $(t_2,\ s_2)$, the consequent displaceiment, $\Delta s = s_2 - s_1$, corresponds to the time interval, $\Delta t = t_2 - t_1$. Furthermore, it can be concluded that, if $\Delta s \neq 0$ and $\Delta t \neq 0$, the object has moved, while, if $\Delta s = 0$ when $\Delta t \neq 0$, no conclusion as to the object's motion may be made, because, after time interval Δt, its displacement is zero.

4.4 s-versus-t "histories" of motion

Associating positions with clock readings offers a particularly simple way of presenting information about the motion of an object: make a graph of position as a function of time. A pair of coordinate axes are set up and labeled s and t, instead of y and x as in an elementary algebra class. In this representation, each "event" in the rectilinear motion of an object (position s occupied at instant t) becomes associated with a point in the s-t plane, defined by the two coordinate axes. An accumulation of data–corresponding values of t and s–can then be "plotted" in such a diagram to give a visual picture of the history of an observed motion. Alternatively, a line or curve drawn on an s-t diagram can be interpreted as the history of a hypothetical motion. In coordinate representations of this kind, the horizontal coordinate, here t, is called the "abscissa," and the vertical coordinate, here s, is called the "ordinate," and the graph is referred to as an ordinate-name versus abscissa-name plot, here, an ***s* versus *t*** or **position versus time** plot. When naming these sorts of "scatter plots," the ordinate always precedes the abscissa.



6. **In the figure above, a sequence of observations of an object's rectilinear motion is plotted, starting at instant $t = -3$ s and ending at $t = +9$ s.**

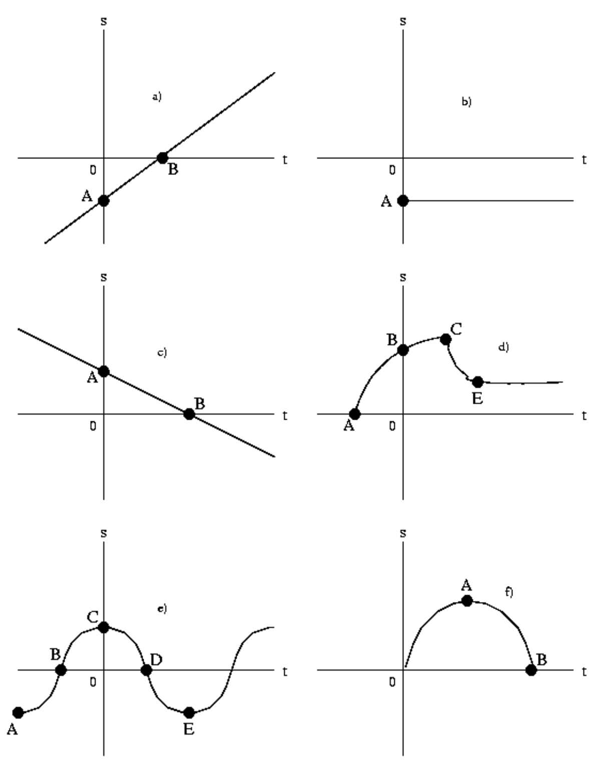
(a) Under an assumption that the motion was observed to be smooth and regular rather than jerky, how would the history between the plotted points be filled in (where might the object be at intermediate clock readings)? Would this filling-in be entirely reliable?

(b) Without making an assumption of smooth and regular motion, what alternative histories are possible? Where would intermediate events be plotted?

(c) What knowledge or suppositions must supplement the numerical data in order to draw continuous lines or curves through a graph of experimental points?

(d) Create a table with the numerical values of time intervals, Δt, in one column, and the corresponding displacements, Δs, in another column for the following pairs of points in the figure: A and C, B and D, C and D, and C and E.

Hypothetical s–Versus–t Diagrams



7. The figure above shows a number of s-versus-t diagrams representing hypothetical histories of straight-line motion of an object. Describe each history in words, noting, in particular, what

happens at the events identified with letters. Prepare to execute each motion with a hand.

4.5 The number $\frac{\Delta s}{\Delta t}$

A new concept intended to give information concerning the rate of motion can be formed from the concepts of displacement, Δs, and corresponding time interval, Δt: a number denoted by $\frac{\Delta s}{\Delta t}$ or $\frac{s_2 - s_1}{t_2 - t_1}$.

What can be said about this number?

1. If the motion is monotonic (that is, if the motion is always and only in one direction), then the magnitude of the number, regardless of algebraic sign, gives some sense of the rate at which distance is traversed in terms of rapidity–how much an object is displaced in some interval of time. Understand that looking at motion this way is a matter choice. If rate of motion were calculated as $\frac{\Delta t}{\Delta s}$ instead of $\frac{\Delta s}{\Delta t}$, then the size of the number would have indicated the object's slowness rather than its rapidity, a measure used in seismology to analyze wave motion through Earth layers.

2. The interval $\Delta t = t_2 - t_1$ has been defined to be a positive number. The *direction* of the displacement therefore determines the algebraic sign of $\frac{\Delta s}{\Delta t}$: positive values indicate motion in the direction of increasing position numbers; negative values indicate motion in the direction of decreasing position numbers (recall that, for example, $-1 > -3$).

3. For monotonic motion, the magitude of $\frac{\Delta s}{\Delta t}$ (that is, ignoring its algebraic sign), is the number used to describe the average rate at which an object moves.

4. If the motion is not monotonic, but reverses itself one or more times during the total time interval Δt observed, the displacement Δs no longer represents the distance traversed by the object. In such case, $\frac{\Delta s}{\Delta t}$ is to be interpreted as the rate which would have been associated with a direct displacement from the initial to the final position, regardless of the complexity of the actual motion. If, at t_2, the object has returned to its initial position, $\frac{\Delta s}{\Delta t} = 0$ regardless of the rate at which the object moved in executing its trip back and forth.

5. Since the object's rate of motion may well be different for different portions of the motion, the number $\frac{\Delta s}{\Delta t}$ seems to be a sort of "lumped" or "smeared out" measure of rapidity. There seems to be a need for a more refined calculation and a more detailed description of the variations in rapidity during a particular displacement. When the rate of motion changes with time, it is called "nonuniform." "Uniform" motion is motion in which equal displacements occur in successive equal time intervals. Note that this is the character of the motions illustrated in histories a), b), and c) of the hypothetical s-t plots in the figure on page 6.

8. Relate each of the five assertions regarding the number $\frac{\Delta s}{\Delta t}$ to relevant diagrams on page 6, describing what happens to the average velocity in each case.

The number $\frac{\Delta s}{\Delta t}$ has been nameless throughout this discussion. Naming a concept is not essential. The important concerns of a definition of a concept reside in how the number is to be calculated and how it is to be interpreted. But the absence of a name for a frequently-used concept becomes cumbersome. The standard name of $\frac{\Delta s}{\Delta t}$ is "average velocity," and it is often symbolized by $\bar{v}$, read read "v-bar."

$$\bar{v} \equiv \frac{\Delta s}{\Delta t} = \frac{s_2 - s_1}{t_2 - t_1} \tag{3}$$

This number is a derived quantity, calculated by relating base quantities of two different dimensions, $[L]^1[T]^{-1}$ (length/time). The corresponding derived unit is the ratio of a length unit to a time unit, such as meters/second, miles/hour, etc. If two average velocities are compared by forming a ratio, such as $\frac{\bar{v}_A}{\bar{v}_B}$, the result is interpreted as $\bar{v}_A$ being so many times $\bar{v}_B$. This result is a "pure" or "dimensionless" number.

The name "average velocity" and the symbol $\bar{v}$ have been given to $\frac{\Delta s}{\Delta t}$ without associating this number with the familiar arithmetic mean, found by summing a group of values and dividing by the number of values. Such a connection does exist, but it involves rather subtle ideas that will be addressed later.

For the special case of uniform motion, in which equal distances are traveled in successive equal intervals of time, the same number for $\frac{\Delta s}{\Delta t}$ is found regardless of time interval and corresponding displacement. It is convenient in this case to drop the bar notation and to write (for uniform velocity only):

$$v \equiv \frac{\Delta s}{\Delta t} = \frac{s_2 - s_1}{t_2 - t_1} \tag{4}$$

9. Suppose the definition of $\frac{\Delta s}{\Delta t}$ were

$$\frac{\Delta s}{\Delta t} \equiv \frac{s_1 - s_2}{t_1 - t_2}$$

How would the interpretation of $\frac{\Delta s}{\Delta t}$ be different? [Note that the Δ symbol is reserved for final minus initial values, not initial minus final values.]

4.6 Graph of s-versus-t for uniform motion

An initial clock reading is perhaps more often designated t_i than t_1, and the corresponding initial position is then designated s_i rather than s_1. Further, an arbitrary later time and corresponding position may be designated t and s rather than t_2 and s_2. The definition of uniform velocity, $v \equiv \frac{\Delta s}{\Delta t}$ would then be expanded as:

$$v = \frac{s - s_i}{t - t_i}$$

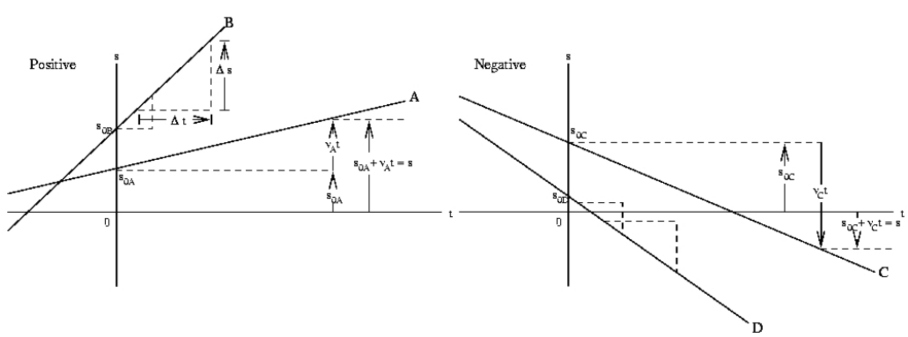
Terms, then, could be rearranged to give position as a function of clock reading:

$$s(t) = v(t - t_i) + s_i = v\Delta t + s_i \tag{5}$$

This statement and the definition from which it was derived are very different logically. This statement is an *equation*, and so specifies a way of calculating the position s associated with some instant of time t provided the values v (the uniform velocity) and s_i (the initial position, that is, the position at clock reading t_i) are known.

To be clear: this is for uniform motion. Therefore, v stands for a particular number, such as 3 m/s. s_i also stands for a particular number: the position of the object at t_i, say at the two-meter mark on the positive side of the origin. The equation has a general form corresponding, with the example values, to $y = 3x + 2$, or, more generally, to an equation of the form $y = mx + b$, where m and b are constants.

This form is familiar from elementary algebra: its graph is a straight line, where m is the slope of the line and b is the y-intercept. To repeat, this straight line is a graph, a picture in an s-t system of coordinates or in the "s-t plane." It is *not* the straight-line path along which the object moves and along which the position numbers s are marked.



The figures above allows additional interpretations of Equation 5:

1. The straight-line s-t history indicates that equal displacements occur in equal intervals of time, precisely what is meant by "uniform velocity": the value of $\frac{\Delta s}{\Delta t}$ will be the same regardles of the time interval chosen (notice the similar triangles).

2. The ratio $\frac{\Delta s}{\Delta t}$ is called the "slope". For a straight line, the slope, and, therefore, $\frac{\Delta s}{\Delta t}$ is constant. In the top plot, Δs is always positive; so too are the velocities and the slopes: the objects move at uniform velocity in what has been (arbitrarily) designated the positive direction. The bottom

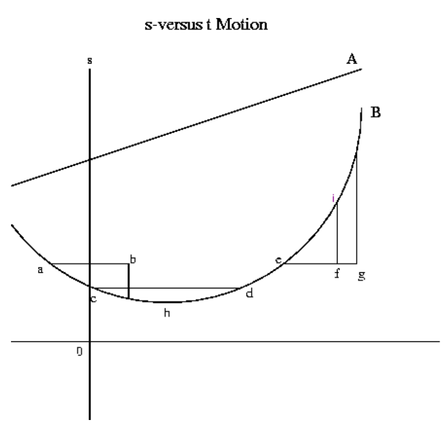
plots show graphs of negative uniform velocities. A negative slope (negative velocity) indicates that the motion is in the direction (arbitrarily) designated negative. The steepness of the slope represents the magnitude of the velocity: the steeper the slope, the larger the magnitude of the velocity.

3. In standard mathematical terminology, $v_B > v_A$, $v_C > v_D$, $|v_C| < |v_D|$[4], where the subscripts refer to the lines in the figure.

4. The figures give geometrical interpretations to the terms $s_i = s_0$ and $v\Delta t = vt$ that appear in the equation.

10. Plot the graph of $s(t) = v\Delta t + s_i = v(t - t_i) + s_i$ for the case $v = +3$ m/s and $s_i = s_0 = +2$ m when $t_i = t_0 = 0$. What happens to the graph if v remains unchanged, but $s_0 = 5$ m, $s_0 = -2$ m, or $s_0 = 0$ m. What happens to the graph if s_0 remains unchanged, but $v = +5$ m/s, $v = -2$ m/s or $v = 0$ m/s? Which of s_0 and v corresponds to the slope and which to the y-intercept of the line? Generalize in your own words the regularities you perceive in the results: What are the effects of different numerical values of s_0 and v on the location and character of the straight line in the s-t plane?

4.7 Refining and extending the concept of velocity



Consider these two s-t histories, A and B. Case A should be recognized as a graph of uniform motion. Case B, however, is a graph of nonuniform motion: each different Δt and corresponding Δs give a different average velocity: between points a and b, $\bar{v}$ is negative, between points c and d, $\bar{v} = 0$, between points e and f, $\bar{v}$ is positive, and, while $\bar{v}$ is also positive between points e and g, its magnitude is greater than its magnitude between e and f.

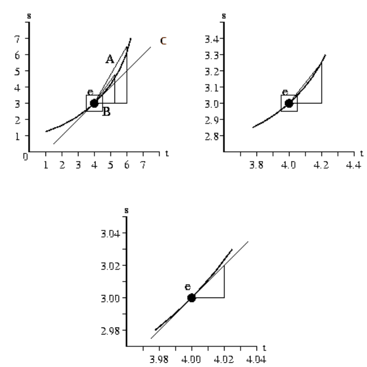
No single number can characterize the velocity throughout the motion plotted in graph B. The object's mostion is changing. It's velocity seems to be systematically increasing: the

[4]The number $|v|$ is frequently referred to as "speed," employed in everyday speech to describe the rapidity of motion. It will not be used much in this course, because velocity, as defined here, is a more sophisticated concept which can be generalized to more complicated motions, such as nonuniform and multidimensional motions.

"average slope" of the curve begins negative and becomes increasingly positive after passing through zero.[5]

Imagine magnifying the portion of the curve in the neighborhood of point e, which, for the sake of illustration, is said to be at $t = 4$ s [see Figure]. Let $t_1 = 4$ s (the time associated with point e) in each plot, and compute $\frac{\Delta s}{\Delta t}$ with t_2 progressively closer to t_1, as suggested by the triangles in the three plots. The value of t_2 (the base of each triangle) decreases from $t_2 = 6$ s for triangle A, to $t_2 = 5$ s for triangle B. That is, Δt decreases from 2 s to 1 s from triangle A to triangle B. Δs decreases even more, from 3.5 m ($s_2 - s_1 = 6.5 - 3.0$ m) to 1.5 m ($s_2 - s_1 = 4.5 - 3.0$ m). Thus $\frac{\Delta s}{\Delta t}$ decreases from $3.5/2 = 1.75$ m/s to $1.5/1 = 1.5$ m/s. Notice, though, that as t_2 approaches the value of t_1, a change of that magnitude in $\frac{\Delta s}{\Delta t}$ requires progressively smaller values of Δt. In the figure on the top right, $\Delta t = 4.2 - 4.0 = 0.2$ s, while $\Delta s = 3.25 - 3.0 = 0.25$ m, so $\frac{\Delta s}{\Delta t} = 1.25$ m/s, another drop of 0.25 m/s. In the bottom figure, $\Delta t = 4.02 - 4.0 = 0.02$ s while $\Delta s = 3.02 - 3.0 = 0.02$ m, so $\frac{\Delta s}{\Delta t} = 1$ m/s, again, a drop of 0.25 m/s. In the upper, left plot, a reduction of Δt by a factor of 2 changed the slope by 0.25 m/s from triangle A to triangle B; a further reduction of Δt by a factor of 5 changed the slope associated with triangle B to the slope in the upper, right plot by 0.25 m/s, and a reduction by a factor of 10 changed the slope of the upper, right plot to the slope of the bottom plot by 0.25 m/s.

Successively Expanded Plots Near Point e



In the bottom figure, the curve is so nearly straight that, to a few decimal places, the same value of $\frac{\Delta s}{\Delta t}$ would be found regardless of the value of t_2. This "final" value of $\Delta s/\Delta t$ is the slope of the straight line which is a very good representation of the expanded section of the curve in the bottom plot. These geometrical relationships thus suggest that this straight line is what is called the "tangent" to the curve at point e. Note that the tangent line C in the top left plot is parallel to the segment in the third.

The same procedure could be applied to every point in case B of the upper figure on page 10. The slope of the tangent at point i is a larger positive value than the slope of the tangent at point e. The slope of the tangent at point h should be very close to zero. The slope of the tangent at point a is negative.

11. Take, as the equation of an s-t curve $s = 3 - t^2$ (with s in meters and t in seconds). Examine the curve in the neighborhood of $t = 1$ by making successively expanded plots. Verify that the

[5]The *magnitude* of the slope first decreases to zero and then increases, but when a signed number (not just its magnitude) becomes more positive, its value increases. Recall, again, $-1 > -3$.

slope of the final, expanded, nearly straight segment is about −2 m/sec.

This procedure of finding a "final" value of $\frac{\Delta s}{\Delta t}$ is so common in physics that it's given a short expression:

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$$

which is read: "the limit as delta t approaches zero of the ratio delta s over delta t," which is conventionally summarized as

$$\frac{ds}{dt}$$

The "d"s are part of the notation, not arithmetical symbols which may be canceled out.

1. The number $\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$ has the character of velocity as previously defined. For each point in an s-versus-t history of motion, there seems to be a single, unique value obtained from this calculation. Obtaining a unique value for a given instant of time suggests the interpretation of this number as a velocity associated with the particular instant. This number is therefore given the name "instantaneous velocity," and given the symbol $v(t)$:

$$v(t) \equiv \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} \equiv \frac{ds}{dt} \tag{6}$$

   Note that this symbol is still applicable to uniform motion, since in this case the instantaneous velocity is always the same.

2. The result of the limit calculation can always be thought of as a velocity associated with some uniform motion, represented by a straight line on the diagram, the velocity being the slope of the line. Therefore, it might be stated that the instantaneous velocity is the slope of the line tangent to the s-t curve at the point in question.

3. Both instantaneous and average velocities can be viewed as slopes of straight lines on an s-t plot: instantaneous velocity as the slope of the tangent at a point and average velocity as the slope of the chord connecting two points between which the average velocity is calculated. This view suggests a simple additional interpretation for each number: Average velocity may be described as that *uniform* velocity at which an object would move from one point to another during a time interval of the same magnitude as that which transpires when the particle moves with the actual, varying velocity. Instantaneous velocity might be interpreted as that uniform velocity at which the motion would continue if all velocity *change* ceased at the instant the instantaneous velocity is being determined. It should be carefully noted that these statements are *interpretations* of the

numbers instantaneous and average velocities, not definitions of them. The definitions are the previously given formulae, which prescribed the specific arithmetical operations which must be carried out to obtain each number.

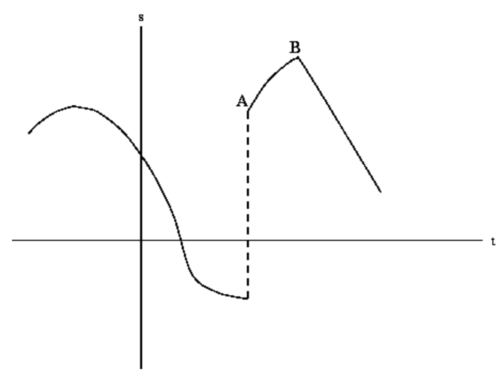
But note:

1. These are very subtle, and perhaps not intuitive, concepts. They couldn't be formulated until analytical geometry and calculus were developed (in the seventeenth century).

2. In particular, the notion of a limit is tricky: what does it mean for a difference to "approach zero?" Dividing by zero makes no sense; how small is "small enough?" Does a limit always, or even ever, exist, and, if so, how can this be predicted?

3. In fact, taking a limit can fail.

s-Versus-t History



The hypothetical s-t history in the figure shows instantaneous position changes at A and B. Such instantaneous changes of position are of course not physical, but can appear to be so if the time resolution of the measuring instrument is coarse enough. Suppose, for example, that the motion of a ball before, during, and after it is struck by a racket, club, or bat were observed with a timing device whose smallest intervals are 0.2 s. An s-t history of the interaction, showing a few seconds on each side of the contact, might look similar to plot in the figure: the interaction would appear instantaneous within the basic coarseness or "graininess" of the time measurements.

In any case, what is the value of $\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$ at the clock reading associated with point A or at the clock reading associated with point B. What value of s should be taken at A? Which curve should be considered at B?

Although the utility of the notion of instantaneous velocity (and other concepts based on similar limit calculations) came to be appreciated during the seventeenth century, questions concerning the logic and validity–the mathematical rigor and meaning–of these calculations troubled even the greatest mathematicians for the next 250 years. Until their logic and validity are confirmed, concepts are used on faith and with the appreciation that they "work."

12. The instantaneous velocity $v(t)$ is the slope of the straight line tangent to an *s*-*t* curve at the instant in question. Sketch *v*-*t* curves for the motions shown in the *s*-*t* plots of problem 7. [It is good practice to draw *v*-*t* curves directly below the corresponding *s*-*t* curves such that the instants of time match up along a vertical line; this gives a clearer sense of the relationship between the two curves.] In each case, describe what a driver of a car would need to do with the brake and accelerator pedals in order to produce the motion plotted.

Notice how the idea of velocity has evolved. Starting with a simple notion of uniform velocity, ignoring the complexity associated with variation from instant to instant, the concept was redefined to clarify a distinction between average and instantaneous velocities. The process of redefining terms, ideas, and concepts lies at the heart of all inquiry, wherever careful analytical thinking is done. The concept of velocity is still far from complete: it remains to be extended to two and three dimensions and to extreme values.

4.8 v-versus-t "histories" of motion

Instantaneous velocity is, by definition, the velocity at an instant. A pair of numbers, $(t,\ v)$, can, as in the case of $(t,\ s)$ be associated and plotted to provide another sort of history of motion.

13. Consider again the plots of problem 7, but change the symbol of the ordinate from s to v, and so regard them as *v*-*t* histories. Describe (in words) the motion represented in each *v*-*t* plot. State carefully the direction of motion, whether or not the direction changes during the time sketched, whether velocity is increasing or decreasing. Then, sketch *s*-*t* curves which would correspond to these *v*-*t* curves. What information is missing as far as positioning the *s*-*t* curves is concerned? [Sketch the *s*-*t* histories directly above or below their associated *v*-*t* plots, so that corresponding times match up on parallel abscissas (time axes).]

If instantaneous velocities, v_1 and v_2, of a particle at any two instants or positions are known, then the corresponding change of velocity is $\Delta v \equiv v_2 - v_1$. For example:

1. If the velocity is -3.2 m/s initially and $+1.3$ m/s finally, then $\Delta v = +1.3 - (-3.2) = +4.5$ m/s.

2. If the velocity is $+2.1$ m/s initially and -4.2 m/s finally, $\Delta v = -4.2 - (+2.1) = -6.3$ m/s.

The algebraic sign associated with Δv indicates the direction in which the change occurred, i.e., whether the velocity has become more positive or more negative. A positive Δv (velocity change in the positive direction), for example, may actually be associated with a decrease in the magnitude of the velocity (speed), as illustrated in the first example.

4.9 Describing rate of change of velocity

Calculating an object's velocity at an instant of time is effectively calculating its velocity at a position in space. How might the rate at which the velocity is changing be calculated? There are at least two possible ways to calculate a relevant measure:

1. The velocity change with respect to displacement, $\frac{\Delta v}{\Delta s}$, i.e, the velocity change in one unit of displacement ($[L]^1[T]^{-1}/[L]^1 = [T]^{-1}$); or

2. The velocity change with respect to time interval, $\frac{\Delta v}{\Delta t}$, i.e. the velocity change in a unit time interval ($[L]^1[T]^{-1}/[T]^1 = [L]^1[T]^{-2}$).

While physics has settled on the second alternative, the first is not "wrong" in any absolute sense. Both ways describe how velocity is changing, but the second provides a simpler way to describe free fall motion, because $\frac{\Delta v}{\Delta t}$ is uniform (at least close to the surface of the Earth), while $\frac{\Delta v}{\Delta s}$ is not. $\frac{\Delta v}{\Delta t}$ also proves to be central to the science of dynamics, which *explains* how motion changes.

The concept represented by the calculation $\frac{\Delta v}{\Delta t}$ is named "acceleration" and assigned the symbol a:

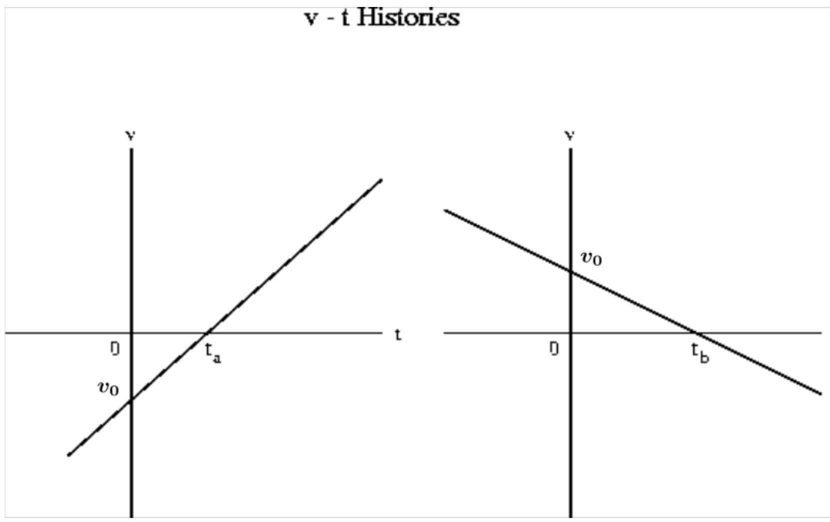
$$\bar{a} \equiv \frac{\Delta v}{\Delta t} \qquad \text{(average acceleration)} \qquad (7)$$

$$a(t) \equiv \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} \equiv \frac{dv}{dt} \qquad \text{(instantaneous acceleration)} \qquad (8)$$

The dimensions of acceleration are $[L]^1[T]^{-1}/[T]^1$ or, shorter, $[L]^1[T]^{-2}$, and common units are m/s^2, (mi/hr)/s, and (km/hr)/s.

Note the analogy with average and instantaneous velocity. In particular, acceleration, like velocity, is an algebraic quantity with plus and minus signs which arise in the calculation and which must be given a physical interpretation.

14. Analyze the following *v*-*t* histories.



Apply the definitions of acceleration to these situations. Sketch *s-t* diagrams for each. Describe at least two simple physical situations in which each sort of motion might be observed.

Notice that there is only one sort of acceleration, not separate ones for speeding up and slowing down. The definitions are algebraic, as they were for velocity, so there are plus and minus signs that require interpretation. Understand that a will be a positive number if $v_2 > v_1$, whether the velocities are positive, negative, or change signs in the interval. Similarly, a will be a negative number if $v_2 < v_1$ algebraically, regardless of the absolute values of v_2 and v_1. In the left-hand diagram of the previous problem, the acceleration is uniform and positive: the velocity is increasing. The result is that the magnitude of the negative velocity decreases up to t_a, when the motion reverses direction. Thereafter, the magnitude of the positive velocity increases.

15. What's going on in the right-hand plot?

4.10 The "kinematic relations" for uniformly accelerated motion: algebraic relations among s, v, a, and t

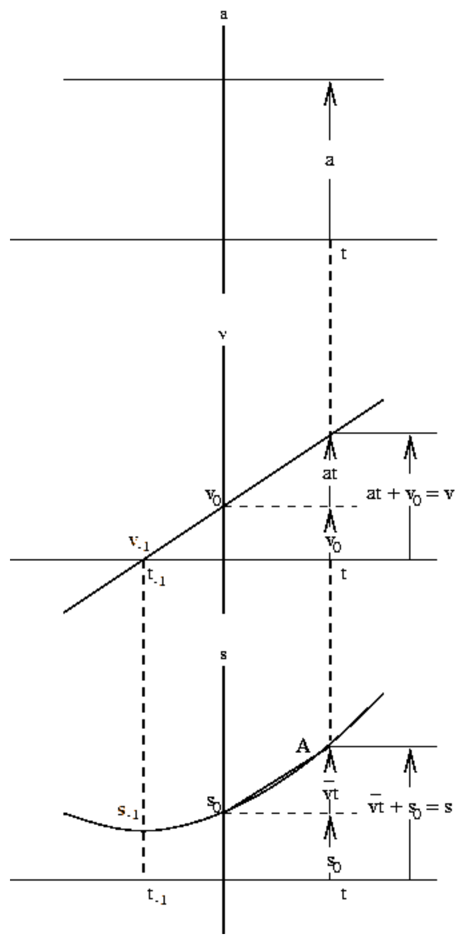
The equations

$$\bar{v} \equiv \frac{\Delta s}{\Delta t} \qquad v(t) \equiv \frac{ds}{dt}$$
$$\bar{a} \equiv \frac{\Delta v}{\Delta t} \qquad a(t) \equiv \frac{dv}{dt}$$

are algebraic statements that create meaning for the symbols v and a. They are definitions, not derived from mathematical manipulations of other relationships. But once the symbols are defined, the definitions can be manipulated algebraically to express other relationships among the symbols, as was done to derive $s(t) = v\Delta t + s_i = v(t - t_i) + s_i$.

Consider uniformly accelerated, rectilinear motion [see Figure].

Uniformly Accelerated, Rectilinear Motion



The top plot is an a-t diagram for such a motion, with constant positive acceleration. The middle plot is a corresponding v-t diagram. The straight line equation of this graph stems from the definition of acceleration, just like the previous derivation followed from the definition of velocity:

$$a \equiv \frac{\Delta v}{\Delta t} = \frac{v - v_i}{t - t_i}$$
$$v(t) = a\Delta t + v_i = a(t - t_i) + v_i$$

As the figure demonstrates, it is often convenient to reference $t_0 = 0$ and, thus, v_0 rather than t_i and v_i, but keep in mind that that the t of the product at in the figure is actually an interval, $t - t_0 = t - 0 = t$, not an instant. The product at would make no sense if t were an instant.

In more mathematical terminology, this equation is an expression for the "dependent variable" v in terms of the "independent variable" t and the constants or "parameters" a, v_i, and t_i.

The bottom figure is an s-t plot corresponding to the v-t plot above it.

$$\bar{v} \equiv \frac{\Delta s}{\Delta t} = \frac{s - s_i}{t - t_i}$$
$$s(t) = \bar{v}\Delta t + s_i = \bar{v}(t - t_i) + s_i$$

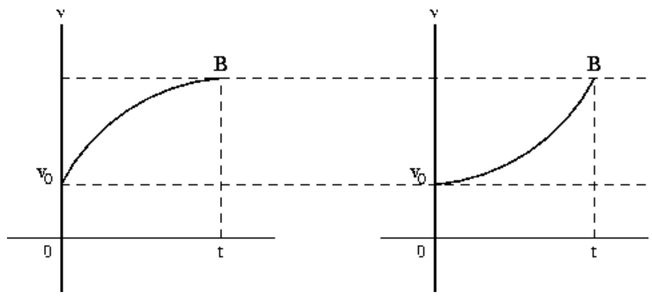
where the value of $\bar{v}$ will be different for every different time interval Δt. An equation of the chord s_0A can be obtained by referencing $t_0 = 0$ instead of t_i as the earlier time, and, thus, s_0 instead of s_i as the corresponding reference position. Then, the equation of the chord is $s(t) = \bar{v}\Delta t + s_0 = \bar{v}(t - t_0) + s_0 = \bar{v}t + s_0$.

Since a, v_i, and t_i are considered to be known numbers, the equation $v(t) = a(t - t_i) + v_i$ is a perfectly satisfactory description of the middle plot. This equation gives a value of instantaneous velocity v at any instant of time t. This is not the case for the equation $s(t) = \bar{v}(t - t_i) + s_i$: $\bar{v}$ is not a single, constant value, and therefore s is not uniquely established in terms of instantaneous values of t. In particular, the equation of the s-t curve must depend somehow on the value of a, but a does not appear explicitly in the equation. How is $\bar{v}$ related to the instantaneous velocities v_i and v?

Average velocity $\bar{v}$ was interpreted previously as the uniform velocity at which a given displacement (occurring with varying velocity) would have taken place in the same time interval. Consider the two v-t histories of varying velocities in the following figures.

The instantaneous velocity in both graphs changes from the same initial value $v_i = v_0$ at $t_i = t_0 = 0$ to the same final value v at instant t. In the left graph, the velocity increases rapidly at first, and then changes slowly: most of the plotted interval was spent at velocities near v. In the right graph, the velocity changes slowly at first, most of the interval being spent at velocities near v_0, and then increases rapidly to the final value v. The average velocity $\bar{v}$ should be different in the two cases, closer in numerical value to v than to v_0 in the left history and vice versa in the right. Since the motion is in the same direction in both cases, Δs should be larger in the left graph than in the right, consistent with the sense, based on interpreting $\bar{v}$ as that uniform velocity at which a given displacement occurs in the same time interval, that $\bar{v}$ is bigger in the left history than in the right.

Varying Velocity



Imagine, then, deforming the two histories so that each curve becomes straighter and both approach a straight line connecting the points marked v_0 and B. Thus, $\bar{v}$ would decrease in the left history and increase in the right one until the two values become identical for the straight-line history. The numerical value of $\bar{v}$ therefore must be exactly halfway between v_0 and v. For a straight-line v-t history, the value of $\bar{v}$ must be between those of v_0 and v and not closer to one or the other, as the values of $\bar{v}$ are in the varying velocity examples. From this argument, and assuming an arithmetical average,[6] it follows that for uniformly accelerated motion (i.e., a straight-line v-t history):

$$\bar{v} = \frac{v + v_i}{2} \tag{9}$$

or, after substituting $v(t) = a\Delta t + v_i = a(t - t_i) + v_i$,

$$\bar{v} = \frac{1}{2}a\Delta t + v_i = \frac{1}{2}a(t - t_i) + v_i \tag{10}$$

Note carefully: these statements are valid *only* for uniformly accelerated motion.

Now, substitute the last equation for $\bar{v}$ into $s(t) = \bar{v}\Delta t + s_i = \bar{v}(t - t_i) + s_i$:

$$\begin{aligned} s(t) &= \left(\frac{1}{2}a\Delta t + v_i\right)\Delta t + s_i = \left[\frac{1}{2}a(t - t_i) + v_i\right](t - t_i) + s_i \\ &= \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i = \frac{1}{2}a(t - t_i)^2 + v_i(t - t_i) + s_i \end{aligned} \tag{11}$$

Since parameter a is assumed to be constant, and parameters v_i and s_i are initial conditions of the motion, this equation can now be identified as the equation of the s-t graph in the figure on page 17, taking $t_i = t_0$, $s_i = s_0$, and $v_i = v_0$. There's a direct connection between position values s and instants of time t in terms of fixed numbers a, v_i, and s_i. The last equation gives the dependent variable s in terms of the independent variable t and the constant parameters a, v_i, s_i, and t_i.

[6] There is no logical justification at this point for this assumption, except that it works. Other arguments, including more mathematically rigorous arguments, lead to the same result.

Furthermore, v and s can be related by eliminating $\Delta t = t - t_i$ from $v(t) = a\Delta t + v_i = a(t - t_i) + v_i$ and $s(t) = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i = \frac{1}{2}a(t - t_i)^2 + v_i(t - t_i) + s_i$:

$$v(t)^2 - v_i^2 = 2a[s(t) - s_i] = 2a\Delta s. \tag{12}$$

16. Derive $v(t)^2 - v_i^2 = 2a[s(t) - s_i]$ by solving for $\Delta t = (t - t_i)$ in $v(t) = a\Delta t + v_i = a(t - t_i) + v_i$ and substituting the result into $s(t) = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i = \frac{1}{2}a(t - t_i)^2 + v_i(t - t_i) + s_i$.

4.11 Interpretation and use of the kinematic equations

In sum, uniformly accelerated rectilinear motion may described mathematically with the following equations:

$$v(t) = a\Delta t + v_i = a(t - t_i) + v_i \tag{13}$$

$$s(t) = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i = \frac{1}{2}a(t - t_i)^2 + v_i(t - t_i) + s_i \tag{14}$$

$$v(t)^2 - v_i^2 = 2a\Delta s = 2a(s - s_i) \tag{15}$$

These equations are frequently referred to as "kinematic relations" for *uniformly accelerated rectilinear motion.* The term "kinematic," as used in this context, refers to the *description* of motion, without reference to the interactions that induce or alter the motion. The science of interactions is referred to as "dynamics" or "kinetics," and describes *how* motion changes through momentum currents (force) and other substance-like currents that carry energy. The discussion of these ideas is postponed until next semester.

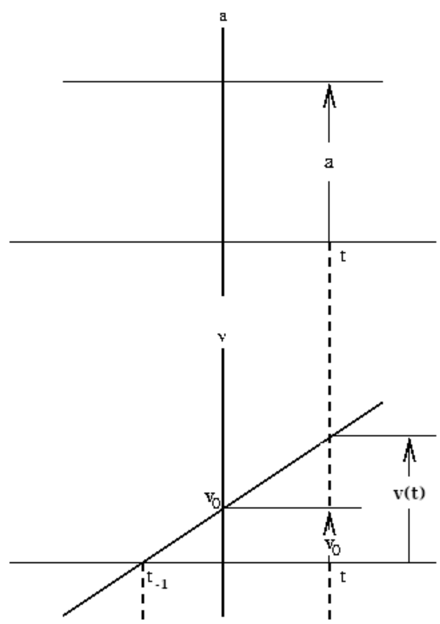
The kinematic relations provide a complete description of uniformly accelerated rectilinear motion. For example, if the motion reverses directions at some instant, the equations say so directly. The same equations describe the motion on either side of the direction reversal–there is no need for one set of equations for one direction of motion and a second set for the other direction.

17. Refer to the two plots in the figure below. They show the acceleration and velocity histories of an object moving in uniformly accelerated rectilinear motion. $v_0 > 0$ is the particle's velocity measured at $t = t_0 = 0$.

(a) Show that $v(t) = at + v_0 = a(t - t_{-1})$ [Note that $t_0 = 0$ and $v_{-1} = 0$, and use the point-slope formula.]

(b) Describe what is happening to the object's motion at instant $t = t_{-1}$. [Note that the characterization "stopped" implies an *interval* during which the particle does not move. Which direction was the motion the instant before t_{-1}, and which direction was the motion the instant after? What is the object's acceleration at $t = t_{-1}$? Would the corresponding s-t plot show a horizontal line at or arround $t = t_{-1}$ Is there any indication in the histories that the object stopped?]

(c) What is the value of t_{-1} in terms of v_0 and a?

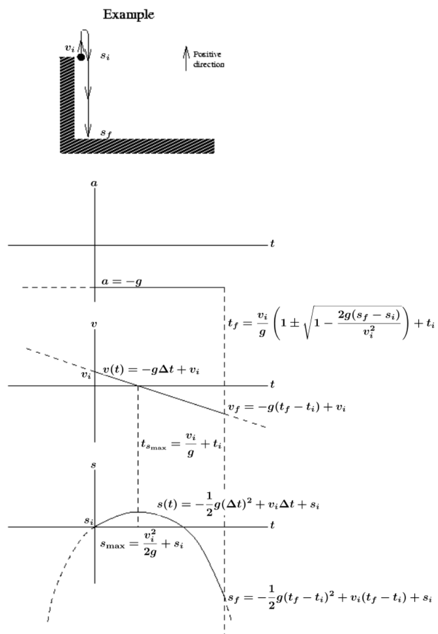


A careful, systematic approach aids in solving problems involving the kinematic relations. Consider these guidelines followed in the example below.

1. Choose a positive direction and the origins of time and position;
2. Write down algebraic statements of the known values;
3. Translate the verbal requirements into symbols;
4. Select the equation which connects the known and unknown quantities in the most direct way; and
5. Calculate the unknown quantity or quantities.

Example A stone is thrown vertically upward with an initial velocity v_i from a cliff edge in such a way that it rises and falls along a straight line which misses the edge of the cliff. In order to apply the kinematic relations here, it is necessary to assume not only that the motion is rectilinear, but that it is also uniformly accelerated. Rough sketches of a-t, v-t, and s-t histories of the motion are illustrated in the figure. The kinematic relations and interesting quantities derived from them are also written out in the figure.

Choosing the position axis orientation and the time and position origins is arbitrary, but some choices prove to be more effective than others. Here, a good (but not the only good) choice is suggested in the figure: The positive position direction is assigned to be upward, so that a is a constant negative number of magnitude g, $a = -g$, and the position and time origins are taken to be the initial position at the time of release, $s_i = s_0 = 0$ and $v_i = v_0 > 0$ at $t_i = t_0 = 0$. One implication of these choices is that the stone's subsequent velocities are always less (more negative, not necessarily slower) than the initial velocity: $v(t) = a\Delta t + v_i = -g(t - t_0) + v_0 = -gt + v_0$.



At the stone's highest position, the motion reverses direction, so the sign of the velocity switches from positive to negative and therefore passes instantaneously through 0. The acceleration, of course, remains $-g$ throughout the stone's motion going up, coming down, and at the instant the motion reverses direction.

What then is the stone's maximum height, $s_{\max} = ??$ when, instantaneously, $v_{s_{\max}} = 0$? This is a relationship between position and velocity only (not time), so the relevant kinematic relation is

$$\begin{aligned} v(t)^2 - v_i^2 &= 2a\Delta s = 2a[s(t) - s_i] \Rightarrow \\ -v_0^2 &= -2gs_{\max} \Rightarrow \\ s_{\max} &= \frac{v_0^2}{2g}, \end{aligned}$$

since $s_i = s_0 = 0$ and $v_i = v_0$. Because g is constant, the maximum height depends entirely on the initial velocity (squared). If the initial velocity were doubled, the maximum height would be four times greater; if the initial velocity were halved, the maximum height would be a quarter as large.

At what clock reading does the stone reach it maximum height, $t_{s_{\max}} = ??$ when, instantaneously, $v_{s_{\max}} = 0$? Because velocity and time are most directly related by the relation:

$$\begin{aligned} v(t) &= a\Delta t + v_i \Rightarrow \\ v_{s_{\max}} &= -g(t_{s_{\max}} - t_i) + v_i \Rightarrow \\ t_{s_{\max}} &= \frac{v_0}{g} \end{aligned}$$

since $t_i = t_0 = 0$ and $v_i = v_0$. As g is a constant (the stone is accelerating downward at g at this instant–even though $v = 0$–and at all other instants during its trajectory), the time at which the stone reaches its highest position, and therefore how long it take to get there, depends only on the initial velocity (to the first power): if the initial velocity were doubled, it would take twice as long for the stone to reach its highest position (despite that position being four times as far), and if the initial velocity were halved, it would take half the time to reach its highest position (despite that position being only a quarter as high).

What is the stone's velocity when, on its way down after reaching its maximum height, it passes the edge of the cliff from which it was thrown, $v_{s_i} = ??$ when $s = s_i$? Again, the most direct relation between s and v is given by

$$v(t)^2 - v_i^2 = 2a\Delta s = -2g[s(t) - s_i].$$

Here, $s = s_i = s_0 = 0$, so the right hand side is zero. Therefore,

$$v_{s_i}^2 = v_i^2.$$

This gives two solutions,

$$v_{s_i} = \pm v_i.$$

Clearly, $+v_i$ is the stone's initial velocity when leaving s_i on the upward path, so

$$v_{s_i\downarrow} = -v_0$$

must be the stone's velocity when returning to s_i on the way down: under these conditions, the stone returns to s_i with a velocity that is equal in magnitude but opposite in direction to v_i, which here is v_0

At what instant does the stone hit the ground at the foot of the cliff, $t_f = ??$ when $s(t_f) = s_f$? The connection between s and t is given by

$$s(t) = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i \Rightarrow$$
$$0 = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t + s_i - s(t) = \frac{1}{2}a(\Delta t)^2 + v_i\Delta t - \Delta s$$

As the time difference is squared, the solution requires the quadratic formula:

$$\Delta t = t_f - t_i = \frac{-v_i \pm \sqrt{v_i^2 + 2a\Delta s}}{a} = \frac{-v_i \pm v_i\sqrt{1 + \frac{2a(s_f - s_i)}{v_i^2}}}{a} \Rightarrow$$
$$t_f = \frac{v_0}{g}\left(1 \mp \sqrt{1 - \frac{2gs_f}{v_0^2}}\right)$$

since $t_i = t_0 = 0$, $s_i = s_0 = 0$, $v_i = v_0$, and $a = -g$. Because g and, of course, v_0^2 are both positive, while s_f is negative in the chosen reference frame, the quantity inside the square-root is greater than 1, so t_f has two solutions, one positive and one negative. Recall that negative time refers to a clock reading before the arbitrarily chosen t_0. The negative solution, therefore, refers to the time at which the stone would have passed s_f on its way up, so as to have an upward velocity of v_0 at $(t_0,\ s_0) = (0,\ 0)$. It is therefore irrelevant to the physical situation of the example. The positive solution

$$t_f = \frac{v_0}{g}\left(1 + \sqrt{1 - \frac{2gs_f}{v_0^2}}\right)$$

is thus the instant at which the stone arrives at the foot of the cliff, s_f, given that it was launched vertically upward at $t_0 = 0$ with initial velocity v_0 from the position $s_0 = 0$.

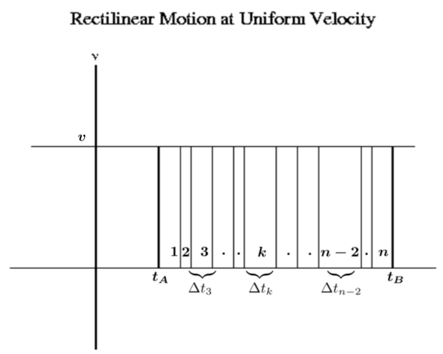
Where is the stone located at $t > t_f$, $s(t) = ??$ when $t > t_f$? Because the kinematic relations found that the stone was traveling in the negative direction before it strikes the ground at s_f at t_f, the equation

$$s(t) = -\frac{1}{2}g(\Delta t)^2 + v_i\Delta t + s_i$$

indicates that $s(t)$ at $t > t_f$ would be further in the negative direction than s_f: $s(t) < s_f$ when $t > t_f$. But the stone hit the ground at s_f and so:

The stone must be lying on the ground at $t > t_f$.

4.12 Relating $\mathbf{v_i}$, $\bar{\mathbf{v}}$, and $\mathbf{v(t)}$: "area" "under" the v-versus-t graph



Consider the v-t diagram of a motion at uniform velocity v. Any arbitrary time interval $\Delta t = t_B - t_A$, however large or small, forms the base of a rectangle with altitude v. The "area" of this rectangle is $v\Delta t = v(t_B - t_A)$. Recall the definition of uniform velocity: $v = \frac{\Delta s}{\Delta t}$. Therefore, $v\Delta t = \Delta s$ or $v(t_B - t_A) = s_B - s_A$, the displacement in the time interval $t_B - t_A$.

Please notice, and keep in mind, that this use of the term "area" is not the familiar one, surface area, of which the dimensions are $[L]^2$. Rather, this "area" is the product of velocity (altitude) and time (base), $[L]^1[T]^{-1} \times [T]^1 = [L]^1$, the dimension of displacement. The area, then, of the rectangle corresponding to any arbitrary time interval in a v-t diagram of uniform velocity is equal to the displacement which occurs in that time interval.

In addition, v could be negative, so that

1. The product of velocity and time interval is negative: this "area" is a signed quantity.

2. The "area" is not just under the curve, but, more generally, between the curve and the horizontal axis.

What is more: the large rectangle can be seen to be made up of any number of smaller rectangles. The bases of these rectangles are small time intervals which, if summed, add up to $t_B - t_A$, and the areas are small successive displacements which add up to $s_B - s_A$. That is, the interval $t_B - t_A$ is divisible into n segments (which need not be of equal size, although equal size is a perfectly acceptable special case), and each segment can be enumerated from 1 to n. Let the letter k represent any one of the numbers between 1 and n, so that each small time interval can be identified with a subscript: the first segment is thus denoted Δt_1, the second Δt_2, the kth segment Δt_k, and the last segment Δt_n. The interval $t_B - t_A$, which can be written as Δt_{AB}, is then the sum of the sub-intervals:

$$\Delta t_{AB} \equiv t_B - t_A = \Delta t_1 + \Delta t_2 + \cdots + \Delta t_k + \cdots + \Delta t_n \qquad (16)$$

The Greek capital letter sigma, Σ, is a conventional mathematical symbol for "summation," and the sum on the right-hand side of Equation 16 can be written more compactly with the aid of this symbol:

$$\Delta t_{AB} = \sum_{k=1}^{n} \Delta t_k \tag{17}$$

which is read "the summation from $k = 1$ to $k = n$ of Δt sub k."

Analogously, the kth displacement (that is, the area of the kth small rectangle) is $\Delta s_k = v\Delta t_k$, and the total displacement in Δt_{AB} is

$$\Delta s_{AB} \equiv s_B - s_A = \sum_{k=1}^{n} v\Delta t_k. \tag{18}$$

Take note that this sum is positive and corresponds to a positive displacement if v is positive (i.e., if the rectangle lies above the t-axis), while the displacement is negative if the rectangle lies below the t-axis.

18. Expand the following summation expressions:

(a) $\sum_{k=1}^{5} k$

(b) $\sum_{k=1}^{4} x_k^2$

(c) $\sum_{k=1}^{n} \frac{1}{k^2}$

(d) $\sum_{k=1}^{3} 2k\Delta t_k$

(e) $\sum_{k=1}^{n} v_k \Delta t_k$

Suppose Δt_{AB} is divided into ever more, smaller segments such that the number of them correspondingly increases. Can this process be carried on indefinitely, as n increases indefinitely and as the segments become indefinitely smallier, does the sum still have meaning? Does it still refer to a particular number $\Delta s_{AB} = s_B - s_A = v(t_B - t_A)$? The answer, at least in this case, is yes–it's the basis of integral calculus–but rigorously proving so is non-trivial, so, here, the assertion will just be accepted:

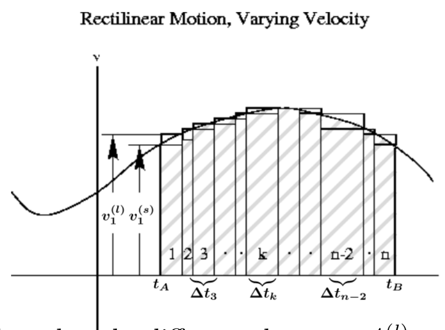
$$\Delta s_{AB} = \lim_{\Delta t_k \to 0,\ n \to \infty} \sum_{k=1}^{n} v\Delta t_k. \tag{19}$$

The expression, $\Delta t_k \to 0$ is read "the interval delta t sub k becomes indefinitely small or infinitesimal," and the expression $n \to \infty$ is read "n becomes indefinitely large or infinite."

4.13 "Area" "under" a v-versus-t curve

To see that the area under a v-t history–the shaded area in the figure, bounded by the curve, the t-axis, and t_A and t_B–in which velocity varies with time still equals displacement in the interval Δt_{AB}, as was the case for uniform velocity, the time axis is again divided into n small intervals. A number of rectangles, with velocity as altitude and time interval as base, can be erected, and the area of each represents a small displacement. For each interval, though, there are two extreme choices for the velocity (altitude). One extreme is the smallest value of v in the interval (illustrated in the figure by the "steps" which lie below the curve), while the other extreme is represented by the largest value of v in the interval (illustrated in the figure by the steps which lie above the curve). Let the altitude of the smaller rectangle for the kth interval be denoted $v_k^{(s)}$ ($v_1^{(s)}$ is marked as an example), and the altitude of the larger rectangle for the kth interval by $v_k^{(l)}$ ($v_1^{(l)}$ is marked as an example).

Adding each set of rectangles leads to two different areas, a larger area, denoted $A^{(l)}$, and a smaller area, denoted $A^{(s)}$:



$$A^{(l)} = \sum_{k=1}^{n} v_k^{(l)} \Delta t_k, \qquad (20)$$

$$A^{(s)} = \sum_{k=1}^{n} v_k^{(s)} \Delta t_k. \qquad (21)$$

Clearly, $A^{(l)} > A^{(s)}$. It should also be clear that the difference between $A^{(l)}$ and $A^{(s)}$, $A^{(l)} - A^{(s)}$, must be the sum of the areas of the small blocks seen lying partially above and partially below the curve.

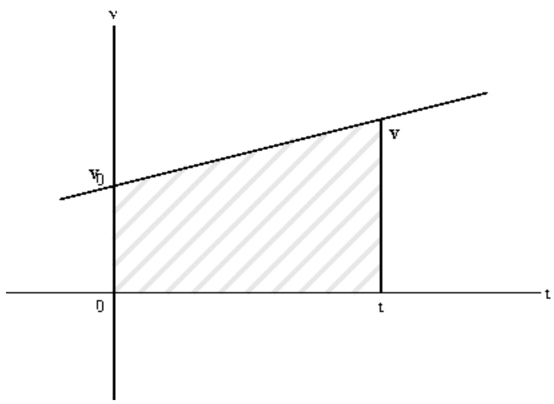
The larger area, $A^{(l)}$, can be interpreted as a displacement taking place at velocities systematically greater than ones which the particle actually experienced, and therefore $A^{(l)}$ must be greater than the actual displacement, Δs_{AB}. Similarly, $A^{(s)}$ represents a displacement smaller than Δs_{AB}:

$$A^{(l)} > \Delta s_{AB} > A^{(s)}$$

As the time intervals become indefinitely smaller (infinitesimal), that is, as $\Delta t_k \to 0$, the little blocks overlaying the curve should become ever smaller, and, therefore, $A^{(l)}$ and $A^{(s)}$ should become more nearly equal. In symbols, the numbers $\lim_{\Delta t_k \to 0} \sum_{k=1}^{n} v_k^{(l)} \Delta t_k$ and $\lim_{\Delta t_k \to 0} \sum_{k=1}^{n} v_k^{(s)} \Delta t_k$ should become equal. The resulting number, then, can be interpreted as the "area under the curve," and as the displacement, Δs_{AB}:

$$\Delta s_{AB} = \lim_{\Delta t_k \to 0} \sum_{k=1}^{n} v_k \Delta t_k. \qquad (22)$$

The superscript notation has been dropped because the limit should not depend on the value of v used in the interval, so long as it is a value for some point on the section of curve lying in that interval.



A simple test of this line of reasoning is to analyze a v-t history of uniformly accelerated motion [see Figure]. The shape of the area under the curve (which is a straight line) is a trapezoid, whose area is the average length of the parallel sides times the separation of those sides. In this case, the area is $\frac{v_0+v}{2}(t-t_0) = \frac{v_0+v}{2}t$, which should be the displacement during the interval $t - t_0 = t$:

$$\Delta s = \frac{v_0 + v}{2} t \tag{23}$$

The term $\frac{v_0+v}{2}$ should be recognized as the uniform velocity at which the displacement Δs would have occurred in the time interval $\Delta t = t - t_0 = t$. That is, $\frac{v_0+v}{2} = \frac{\Delta s}{\Delta t}$. But, recall that $\frac{\Delta s}{\Delta t} = \bar{v}$. So, both ways of looking at the average velocity–as the uniform velocity that results in the same displacement in the same interval as a varying velocity, and as the mathematical average of the initial and final velocities that bound the area under a v-t history of uniformly accelerated motion–are consistent.

19. (a) By analogy with the claim that the area under a portion of a v-t curve is to be interpreted as the displacement occurring in the given time interval, how should the area under an acceleration-versus-time curve be interpreted? Look again at the histories on page 6, but now assume that they are a-t histories (instead of s-t histories) and interpret the area between the a-t curve and the time axis for

i. graph a), between points A and B,

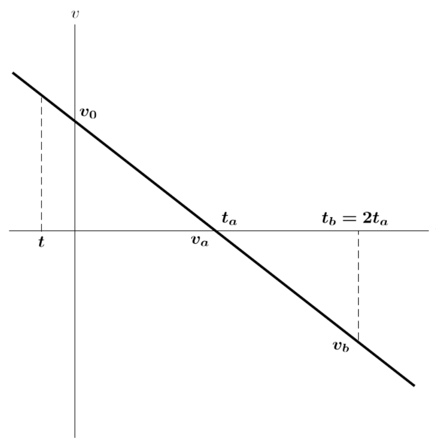
ii. graph b), for the entire history shown,

iii. graph c), between points A and B, and

iv. graph e), between points A and E.

(b) What interpretation, if any, could be given to the area under a portion of an s-t curve?

20. **Measuring areas in the figure below from $t = 0$ (the initial instant), describe and interpret the way in which the numbers behave as the value of t increases. Pay particular attention to the instants t_a and t_b. [Recall that negative values of area go with negative values of v.]**



4.14 Free fall

An object which experiences an acceleration due only to gravity, as was the case of the example of the stone thrown vertically at the edge of a cliff, is said to be in "free fall." The magnitude of that acceleration is symbolized by g, which is called, generally, the "gravitational field strength, but in free fall, the "acceleration due to gravity."[7] The direction of that acceleration defines "down," but the sign ($\pm$) of "down" depends on the (arbitrarily chosen) orientation of the reference system. That is, g has a magnitude but not a sign. Note that the object need not be moving down to be in free fall. The stone, even during its upward trajectory, was in free fall, as are the moon circling Earth and Earth circling the sun.

4.15 Superposition of motions: adapting the description of rectiliniear motion to motion in a plane or in three dimensions

Consider projectile motion, motion of an object (in this case, usually referred to as a "projectile") in a plane (two-dimensional motion) in which the projectile accelerates uniformly in only one dimension, so that its velocity is uniform in the other dimension. This motion can be analyzed as a compound of two

[7] In general, it is best to refer to g as "gravitational field strength," since g appears in equations when objects are not accelerating at magnitude g or even at all. The acceleration of an object in free fall, that is, due only to gravity, will have magnitude g, which can then be referred to as "acceleration due to gravity."

independent motions, uniform in one dimension and uniformly accelerated in the other, such that the motion in one dimension does not influence the motion in the other, each behaving as if it alone were present. The net effect is a simple combination of the two motions calculated separately. This is an hypothesis about physics, not just a matter of definition. To understand the nature of the problem, two questions must be answered.

1. Does imparting, say, a horizontal velocity to a particle in any way alter the vertical acceleration and velocities which it acquires while accelerating?
2. Conversely, does the presence of vertical acceleration and velocity alter the horizontal velocity which a particle might initially have?

These are two *separate* questions, and an answer to one does not automatically supply the answer to the other.

Please view the first and one or both of the second and first part of the third videos:

- `https://www.youtube.com/watch?v=zMF4CD7i3hg`
- `https://www.youtube.com/watch?v=KacTRPL1MtE`

  and/or the first 1 minute, 5 seconds of

  `https://www.youtube.com/watch?v=j1URC2G2qnc`

From the first: since the balls occupy identical levels at corresponding times, it may be inferred that the horizontal velocity of the projected ball has not altered the vertical motion it would have had in the absence of horizontal velocity. From the second and/or first part of the third videos: since the ball rises and falls but also remains in line with the moving cart/truck at each successive instant, it may be inferred that the initial horizontal velocity is not altered by the effects of a vertical acceleration.

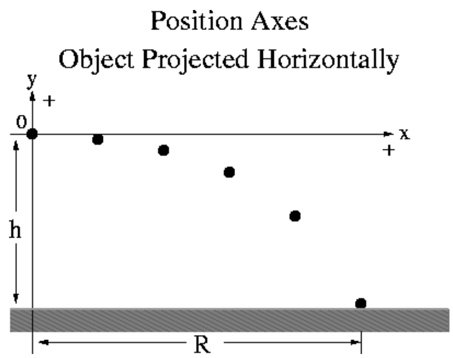
The evidence indeed suggests that motions at right angles to each other are independent, that is, one does not influence or alter the other. It should therefore be possible to "compound" or "superpose" numerical knowledge gained separately of each motion into a description of the combined two-dimensional effect.[8]

4.16 Projectile motion: initial velocity horizontal

[8]In fact, motions at right angles to each other are independent only to a certain degree of precision and only over a limited range of observation. The Special Theory of Relativity predicted–and experiment has verified–that if the horizontal velocity v_x of a falling object is very large relative to the observer, the acceleration a_y would appear smaller than it would from a frame of reference in which the horizontal velocity relative to the observer is very small or zero. The ratio of the two values of acceleration is $1 - (v_x^2/c^2)$, where c is the velocity of light (186,000 mi/sec or 3×10^8 m/sec). This fraction does not differ from unity by as much as one one-hundredth of one percent until v_x is of the order 2000 mi/sec or around 3 million m/sec. But such velocities are only accessible to subatomic particles driven by particle accelerators. Under ordinary circumstances, the dependence is unobservably small.

The videos demonstrate that, within the appropriate range of velocities,

1. horizontal and vertical motions behave as though each were alone present; and
2. in the case of the initial velocity being horizontal, the horizontal motion remains uniform and unaccelerated.



Consider the trajectory of a ball with initial horizontal velocity [see Figure]. The positions of the ball may be referred to a set of coordinate axes y and x. With the axes oriented as in the figure, let the initial position be at the origin of the coordinate system, and denote the magnitude (absolute value) of the vertical acceleration g and the uniform horizontal velocity v_{0x}. If t is measured from the instant of projection, the kinematic relations give, for horizontal motion,

$$v_x(t) = a_x\Delta t + v_{ix} = v_{0x} = \text{ constant} \tag{24}$$

$$x(t) = \frac{1}{2}a_x(\Delta t)^2 + v_{ix}\Delta t + x_i = v_{0x}t, \tag{25}$$

and for vertical motion,

$$v_y(t) = a\Delta t + v_{iy} = -gt \tag{26}$$

$$y(t) = \frac{1}{2}a_y(\Delta t)^2 + v_{iy}\Delta t + y_i = -\frac{1}{2}gt^2 \tag{27}$$

Thus, at any instant of time t, the separate values of the x and y coordinates are known, and the position of the projectile can be located in the x-y plane. This is what is meant by a simple superposition of the two motions: while the projectile undergoes a horizontal displacement from 0 to x, it simultaneously undergoes a vertical displacement from 0 to y. The total or resultant displacement is from point (0 0) in the figure to the point $(x,\ y)$. Because, for any value of x, there is a corresponding value of y, y can be regarded as a function of x, and this function can be obtained by eliminating t from the two position equations. From the x-position equation, $t = x/v_{0x}$. Substituting this expression for t into the y-position equation yields:

$$y(x) = -\frac{1}{2}g\left(\frac{x}{v_{0x}}\right)^2 = -\left(\frac{g}{2v_{0x}^2}\right)x^2 \tag{28}$$

The form of this equation should be recognized as a parabola, $y = \frac{1}{2p}x^2$,[9] with $p = -\frac{v_{0x}^2}{g}$ (which is negative in this construction), passing through the origin, opening downward, and symmetric around the y-axis. Since the physical phenomenon described by this result begins at the origin, the left branch of the parabola (in quadrant III) is not relevant. The concern of this case is with the branch in quadrant IV only. The equation represents the path or trajectory of the projectile in space, and predicts this path to be semi-parabolic.

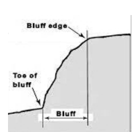
This simple physical theory produces a mathematical result which illuminates and orders a whole class of events: with larger initial velocity v_{0x}, p is larger and the parabola is broader; a larger vertical acceleration, if it were still uniform, would simply produce a narrower parabola; if the ground is a distance h below the point of projection, the horizontal range R is given by the amplitude of the parabola at a distance h from its origin:

$$R = v_{0x}\sqrt{2h/g}. \tag{29}$$

21. Derive the equation $R = v_{0x}\sqrt{2h/g}$, where h is the distance below the point of projection and R is the horizontal range of motion, by rearranging and appropriately re-identifying variables in $y = -\left(\frac{g}{2v_{0x}^2}\right)x^2$.

22. If an object is rolled or slid off the edge of a table, you should be able to determine its initial horizontal velocity from measurements of R and h. Perform such an experiment and calculate the horizontal velocity.

23. Suppose you were standing on a bluff [see Figure]. Stones and a stop watch are available, but the bluff extends out below its edge so that a vertical drop would hit the face of the bluff before landing at its toe. Is it still possible to determine the height of the bluff? How? What conditions must be satisfied?



[9] A parabola is a special conic section formed by the intersection of the surface of a right circular cone with a plane parallel to another plane tangent to the conical surface. The general equation of a conic section is the second order polynomial $ax^2 + bxy + cy^2 + dx + ey + f = 0$, where a, b, c, d, e, and f are all real numbers and at least one of a, b, and c is not zero. This equation takes the form of a parabola if $b^2 - 4ac = 0$, or, equivalently, if $ax^2 + bxy + cy^2$ is the square of the linear polynomial, $\sqrt{a}x + \sqrt{c}y = 0$. Another definition of a parabola is the collection of points equidistant from a point (called the focus) and a line (called the directrix). If the directrix is parallel to the x-axis and above or below it by a distance $p/2$, and the focus is on the y-axis on the other side of the x-axis by the same distance, $p/2$, then the collection of points equidistant from focus and directrix form a parabola with the equation $y = \frac{1}{2p}x^2$.

Whenever a particle is projected at uniform velocity in one direction and is uniformly accelerated at right angles, its trajectory must be parabolic. While gravitational effects were involved here, the same description holds when electrically charged particles move between the charged plates of a capacitor. This ordering of broad classes of events and of entirely different phenomena under one mathematical representation is characteristic of a scientific theory, especially in physics.

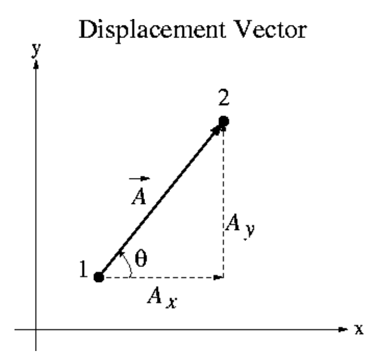
4.17 Idealizations in the description of projectile motion

This model of projectile motion includes a number of assumptions and approximations. Things like air resistance are ignored. Such models are sometimes referred to as "first-order." If higher precision is required, small corrections, sometimes referred to as second-order corrections, are applied. The terms "first-" and "second-order" refer to relative size: the correction terms are small relative to the importance of the first-order terms. Furthermore, the correction terms might not demand high accuracy: if the correction can be estimated roughly to within 20% of its value, and if the correction is about 10% of the first-order term, the final uncertainty is only 2%.

In the case of projectile motion, computational algoriths have been developed which correct the first-order model for the effects of varying air resistance, wind, and even (for very long-range projectiles) for the rotation of Earth.

The first-order model suggests that a projectile possesses two instantaneous velocities v_x and v_y. Can the projectile be thought of as having a single, resultant, instantaneous velocity, compounded of v_x and v_y, that varies in direction as well as magnitude?

4.18 Vectors and ordered pairs of numbers



The kinematics of rectilinear motion employs the entire apparatus of ordinary arithmetic: positive and negative numbers, axioms of commutation and association; and rules of addition, subtraction, and other operations. This apparatus was applied, in rectilinear motion, to points on a number line. Motion in a plane represents positions by an ordered pair of numbers, $(x,\ y)$, referred to as "coordinates" of a point in the plane.

Consider a displacement from 1 to 2 [see Figure]. This displacement is made up of horizontal and vertical components which are labeled A_x and A_y. The directed line connecting points 1 and 2 has been denoted $\vec{A}$. The arrow on the line indicates the direction of the displacement. The symbol $\vec{A}$ is called a displacement vector, which is defined by the ordered pair of numbers A_x and A_y, its horizontal and vertical "components":

$$\vec{A} \equiv (A_x,\ A_y). \tag{30}$$

Both A_x and A_y correspond to rectilinear displacements, and so can, therefore, be either positive or negative. The angle θ defines the vector's direction, determined by $\tan\theta = A_y/A_x$. The total size, or magnitude, of the displacement, regardless of direction, is the length of $\vec{A}$, which is related to A_x and A_y by the Pythagorean Theorem, so long as the space in which these ideas are applied obeys Euclidean geometry. When this is the case, the magnitude or length of $\vec{A}$ is denoted the absolute value $|\vec{A}|$, and defined by

$$|\vec{A}| \equiv \sqrt{A_x^2 + A_y^2}. \tag{31}$$

It then follows that the vector itself can be "resolved," with trigonometry, into its components, a procedure referred to as "projecting":

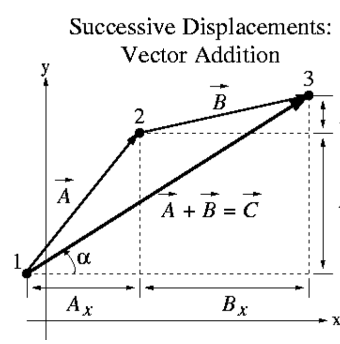
$$A_x = |\vec{A}|\cos\theta$$
$$A_y = |\vec{A}|\sin\theta.$$

These definitions have been connected, by way of illustration, with the definition of a displacement vector. The vector, however, need not be interpreted as a displacement. Regardless of the case, the equations refer to any vector $\vec{A}$, which is defined by the ordered number pair $(A_x,\ A_y)$.

4.19 Addition of vectors

Consider a succession of displacements, 1 to 2 followed by 2 to 3 [see Figure]. The overall result is equivalent to a direct displacement $\vec{C}$ from 1 to 3 with components $(A_x + B_x,\ A_y + B_y)$. The plus sign refers to ordinary arithmetic addition of the respective components' numerical values. This, then, is the definition of "vector addition," the outcome of $\vec{A} + \vec{B}$, identified by the "resultant" $\vec{C}$,

$$\vec{C} = \vec{A} + \vec{B} \equiv (A_x + B_x,\ A_y + B_y)$$
$$\equiv (C_x,\ C_y). \tag{32}$$



The notation $\vec{A} + \vec{B}$ is called a "vector sum." The plus sign here means something quite different from the plus sign of ordinary arithmetic. It refers to addition of components to produce a resultant, like that in the figure. The length of $\vec{C}$ in this example is *not* equal to the arithmetical sum of the lengths of $\vec{A}$ and $\vec{B}$, but to the Pythagorean sum of their components. The concept is related to ordinary addition, however, and rather than use a different symbol, the conventional plus sign is used, understanding that it means different things in the two separate contexts.

It follows from Equation 32 that

$$\tan\alpha = \frac{A_y + B_y}{A_x + B_x}, \tag{33}$$

$$|\vec{C}| = |\vec{A} + \vec{B}| = \sqrt{(A_x + B_x)^2 + (A_y + B_y)^2}, \tag{34}$$

$$C_x = |\vec{C}|\cos\alpha, \tag{35}$$

$$C_y = |\vec{C}|\sin\alpha. \tag{36}$$

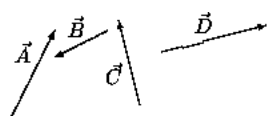
Thus, in general, the sum of two vectors is a third vector with magnitude and direction different from the first two. The definition is an arithmetical representation of what is meant by vector addition, while the figure is a geometrical representation. The geometrical equivalent of the definition is the graphical operation of placing the arrows representing the two vectors $\vec{A}$ and $\vec{B}$ head to tail and drawing the third arrow $\vec{C}$ connecting the first tail with the last head. The process of addition can, of course, be applied to 3, 4, 5, or any number of vector terms.

The sum of two or more vectors is called a "resultant vector" or simply a "resultant." The components of a "zero vector" are (0, 0). A group of vectors might very well sum to zero. This would occur geometrically when the head of the last arrow falls back on the tail of the first.

24. The definition of vector addition implies that vector addition is commutative: $\vec{A} + \vec{B} = \vec{B} + \vec{A}$. In other words, the order of addition does not alter the final result. Using ordinary arithmetic arguments, because vector components are ordinary numbers, show how this assertion follows from the definitions of vectors and of vector addition. Then illustrate the same idea geometrically.

25. The associative rule of vector addition is stated: $\vec{A} + (\vec{B} + \vec{C}) = (\vec{A} + \vec{B}) + \vec{C}$. Show that this must follow from the similar associative property of numbers in ordinary arithmetic.

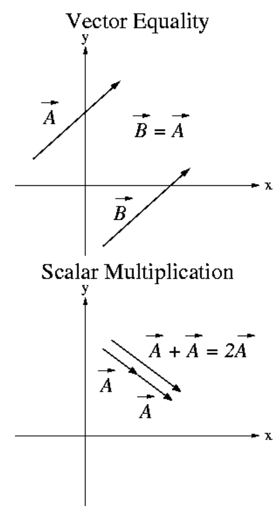
26. Generalize the definition of vector addition by writing out what would be meant by $\vec{A} + \vec{B} + \vec{C} + \vec{D} + \ldots$ in the form of the original definition. Then illustrate the same idea geometrically by adding the following vectors head to tail.



Notice that the components of a vector can themselves be treated as special vectors, parallel to the axes. Consider again $\vec{A}$ on page 33: $\vec{A}_x = (A_x, 0)$, $\vec{A}_y = (0, A_y)$. It then follows that $\vec{A} = \vec{A}_x + \vec{A}_y$, which is simply another view of a vector and its components: any vector can be thought of as the vector sum of its components.

27. Give a geometrical illustration of the identity $\vec{A} = \vec{A}_x + \vec{A}_y$ with appropriate arrows.

4.20 Multiplying a vector by a scalar



Two vectors $\vec{A}$ and $\vec{B}$ [see Figure] are said to be equal if their respective components are equal. That is, $\vec{A} = \vec{B}$ if $A_x = B_x$ and $A_y = B_y$. They are equal even though they don't lie on top of one another or have the same origin. Components are lengths–that is, differences between two positions–so origin is arbitrary. Only the difference between respective head and tail components matters.

If two equal vectors are added $\vec{A} + \vec{A}$ [see Figure], it follows from the definition of vector addition that $\vec{A} + \vec{A} = (2A_x,\ 2A_y)$. This is a vector having the same direction as $\vec{A}$ [why?] and twice the length, so $\vec{A} + \vec{A} = 2\vec{A} = (2A_x,\ 2A_y)$.

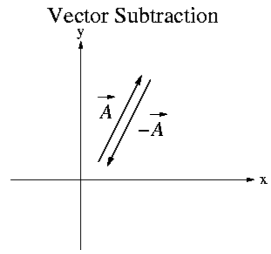
Multiplication of a vector by a scalar number k can then be defined as

$$k\vec{A} \equiv (kA_x,\ kA_y). \tag{37}$$

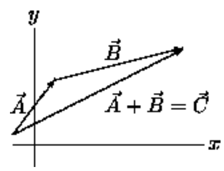
If k is a positive number, the direction of $k\vec{A}$ is identical with that of $\vec{A}$; if k is a negative number, the direction of $k\vec{A}$ is directly opposite that of $\vec{A}$. The symbol k denotes any number (not necessarily an integer) and is without directional properties. It is conventional to distinguish k from a vector by use of the contrasting term "scalar." This definition is called "the rule of vector multiplication by a scalar." Since k might be a fraction, say $1/b$, where b is some real number, this definition also includes the division of a vector by a scalar.

4.21 Subtraction of vectors

In arithmetic, the negative of a number x is that number which, when added to x, gives zero: $x + (-x) = 0$. The same pattern applies to vectors: $-\vec{A} \equiv (-A_x,\ -A_y)$. Then, from the rule for vector addition, $\vec{A} + (-\vec{A}) = \vec{0} = (0,\ 0)$ [see Figure].



28. (a) **Using the figure below, sketch the vector $-\vec{B}$. Using the graphical representation, find $\vec{A}+(-\vec{B})$, written $\vec{A}-\vec{B}$ for short, as in arithmetic.**

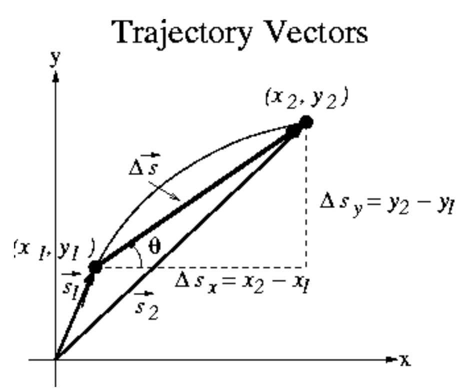


(b) **Assuming that $\vec{C}$ and $\vec{B}$ are given, as in the figure, find (sketched graphically) $\vec{C}-\vec{B}$.**

(c) **Generalize the ideas developed above to give the arithmetical rule for subtraction of vectors: $\vec{A}-\vec{B}=(A_x-B_x,\ A_y-B_y)$. Note that the meaning of the minus sign on the left of this equation differs from the meaning on the right.**

4.22 Displacement, velocity, and acceleration vectors

Consider the displacement from point $(x_1,\ y_1)$ to point $(x_2,\ y_2)$ along the trajectory shown in the figure. The vectors $\vec{s}_1$ and $\vec{s}_2$ from the origin to position points 1 and 2 are called "position vectors." They locate points 1 and 2 relative to the origin and can be thought of as equivalent displacements from the origin to the points in question . From the definition of subtraction, the displacement vector from point 1 to point 2 is



$$\Delta\vec{s}=\vec{s}_2-\vec{s}_1=(x_2-x_1,\ y_2-y_1)$$
$$=(\Delta x,\ \Delta y)=\Delta\vec{x}+\Delta\vec{y},$$

and the direction of $\Delta\vec{s}$ is determined by

$$\tan\theta = \Delta y/\Delta x.$$

Note that it makes s no difference that the object moved along the curved path rather than along the arrow $\Delta\vec{s}$. Recall that the displacement Δs in rectilinear motion is $s_2 - s_1$ which may not necessarily equal the distance traversed.

Making Δx progressively smaller brings point 2 closer to point 1 and the inclination θ keeps changing. If the appropriate limits exist, the limit of inclination of $\Delta\vec{s}$ is determined by $\lim_{\Delta x\to 0}\tan\theta = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \equiv \frac{dy}{dx}$. The right-hand side of this equation is the number associated with the slope of a tangent to a curve. Here the slope has meaning as a direction in space. This "instantaneous direction" of motion at any point along a trajectory can be thought of as being in the direction of the tangent to the trajectory at that point.

Referring again to the figure and to the definition of displacement, the average velocity can be written

$$\frac{\Delta\vec{s}}{\Delta t} = \left(\frac{\Delta x}{\Delta t},\ \frac{\Delta y}{\Delta t}\right) = \frac{\Delta\vec{x}}{\Delta t} + \frac{\Delta\vec{y}}{\Delta t}. \tag{38}$$

Since the time interval Δt is a scalar quantity, this equation follows directly from the rule for vector multiplication by a scalar. These new terms have all the vector properties previously associated with displacement. Recall that division (that is, multiplcation by the inverse of the divisor) by a positive scalar gives a new vector parallel to the first one.

Translating this result into words: the average velocity between positions 1 and 2 has the magnitude $|\Delta\vec{s}|/\Delta t$ and the direction of $\Delta\vec{s}$. The components of the velocity vector are $\Delta x/\Delta t$ and $\Delta y/\Delta t$, which can be written $\bar{v}_x$ and $\bar{v}_y$.

Providing limits exist

$$\lim_{\Delta t\to 0}\frac{\Delta\vec{s}}{\Delta t} = \left(\lim_{\Delta t\to 0}\frac{\Delta x}{\Delta t},\ \lim_{\Delta t\to 0}\frac{\Delta y}{\Delta t}\right) = \lim_{\Delta t\to 0}\frac{\Delta\vec{x}}{\Delta t} + \lim_{\Delta t\to 0}\frac{\Delta\vec{y}}{\Delta t},$$

which translates into the shorthand of velocity symbols as

$$\vec{v} = (v_x,\ v_y) = \vec{v}_x + \vec{v}_y. \tag{39}$$

Vector symbols may be used because the ordered pair of numbers in this equation must have all the vector properties associated with displacements, being derived from displacements by scalar multiplication by $1/\Delta t$.

Since $\vec{v}$ is simply shorthand for $\lim_{\Delta t\to 0}\frac{\Delta\vec{s}}{\Delta t}$, its direction should be the limiting direction of $\Delta\vec{s}$. That is, $\vec{v}$ should be tangent to the trajectory at point 1 (an assertion without rigorous proof here). Then the magnitude of the velocity vector is $|\vec{v}| = \sqrt{v_x^2 + v_y^2}$, and its angle of inclination is determined by $\tan\theta = v_y/v_x$.

An instantaneous velocity vector $\vec{v}$ can be associated with any point along a trajectory. The components of this resultant vector are the instantaneous velocities v_x and v_y. In the case of projectile motion in which the initial velocity is horizontal, a resultant vector $\vec{v} = \vec{v}_{0x} + \vec{v}_y$, and $|\vec{v}| = \sqrt{v_{0x}^2 + v_y^2}$. As motion proceeds along the trajectory, v_{0x} remains constant, but v_y becomes more negative, therefore $\vec{v}$ changes in magnitude and direction. If its velocity changes, whether in magnitude or direction, or both, the projectile undergoes an acceleration.

Since acceleration is derived from velocity changes (that is, $\Delta\vec{v}$ by another operation of multiplication by the scalar quantity $1/\Delta t$), acceleration in two dimensions must have the same vector properties as displacement and velocity:

$$\vec{a} \equiv \lim_{\Delta t \to 0} \frac{\Delta \vec{v}}{\Delta t} = (a_x,\ a_y) = \vec{a}_x + \vec{a}_y. \tag{40}$$

The direction of instantaneous acceleration, however, need not be tangent to the trajectory. In the projectile motion motion example, the acceleration vector has a constant magnitude and would be drawn directed downward at every instantaneous position. In this case, $\vec{a} = (0,\ -g)$.

29. (a) **A particle at some instant of time has a horizontal component of velocity v_x and a vertical component v_y. What are the magnitude and direction of the resultant velocity?**

(b) **What are the magnitude and direction of the resultant velocity if, at some instant of time, a particle has a horizontal component of velocity of $3v_x$ and a vertical component $-4v_y$?**

30. What are the horizontal and vertical components of a resultant velocity $\vec{v}$

(a) **directed θ above the horizontal?**

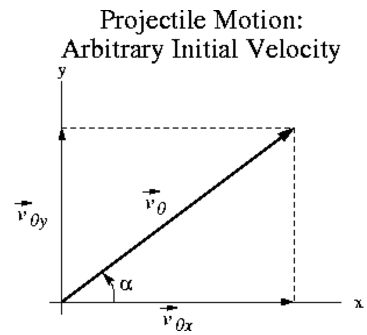
(b) **directed $\theta/2$ below the horizontal?**

Any ordered pair of numbers obeying this arithmetic is called a vector. Arbitrary vectors $\vec{A}$ and $\vec{B}$ may represent mathematical constructs defined by ordered pairs of numbers without any physical interpretation whatever, or they may be *interpreted* to represent displacements or velocities or accelerations, depending on the context in which they are applied. Any physical quantity which is so defined that it must obey vector arithmetic is called a "vector quantity."

4.23 Projectile motion: initial velocity not horizontally directed

The concept of obtaining a resultant velocity by adding components and the reverse process of resolving a velocity into its components provides a method for dealing with the general case of projectile motion, in which the particle is given an initial velocity $\vec{v}_0$ at some angle α to the horizontal [see Figure].

Denoting the magnitude of the velocity vector by v_0 rather than the more cumbersome $|\vec{v}_0|$, the motion should consist of the superposition of a horizontal part with uniform velocity component $v_{0x} = v_0 \cos\alpha$ and a vertical part with uniform acceleration and an initial vertical velocity $v_{0y} = v_0 \sin\alpha$. If the initial vertical component is directed upward, the projectile must describe a trajectory in which it rises and then falls, as in the case of the vertical stone throw example.



Here are the kinematic relations relevant to this physical situation [refer again to the figure], taking the origin of the motion at the origin (0, 0) of the $(x,\ y)$ coordinate system. The initial time is taken to be $t_i = t_0 = 0$.

<u>Kinematic Relations</u>

Horizontal Motion

Initial velocity component:	$v_{0x} = v_0 \cos\alpha$.	[1]
Velocity component in time t:	$v_x = v_{0x} = v_0 \cos\alpha$ (constant).	[2]
Position at time t:	$x = (v_0 \cos\alpha)t$.	[3]
Velocity-position relation:	—	

Vertical Motion

Initial velocity component:	$v_{0y} = v_0 \sin\alpha$.	[4]
Velocity component in time t:	$v_y = -gt + v_0 \sin\alpha$.	[5]
Position at time t:	$y = -\frac{1}{2}gt^2 + (v_0 \sin\alpha)t$.	[6]
Velocity-position relation:	$v_y^2 - v_{0y}^2 = -2gy$.	[7]

Combined Motion

Equation of trajectory: solve for t in [3] and substitute into [6)]; valid providing $\cos\alpha \neq 0$	$y = -\frac{g}{2v_0^2 \cos^2\alpha}x^2 + (\tan\alpha)x$.	[8]
Magnitude v of resultant velocity $\vec{v}$ at any instant:	$v = \sqrt{v_x^2 + v_y^2}$.	[9]
Direction θ of $\vec{v}$ at any instant:	$\tan\theta = \frac{v_y}{v_x}$.	[10]

4.24 Analysis of the projectile equations

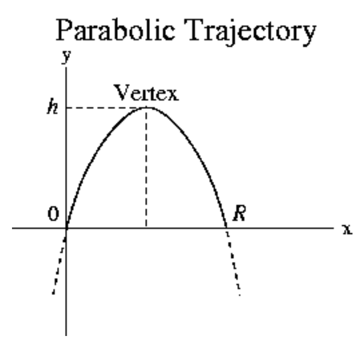
The equation of trajectory [8] is the counterpart of the equation $y = -\frac{g}{2v_{0x}^2}x^2$ derived previously for the special case of an initial horizontal motion. It should contain most of the story of general projectile motion. How might it be understood theoretically?

1. Does the equation reduce to the special cases already analyzed?

   Suppose $\alpha = 0$. This corresponds to the special case of initial horizontal motion. Since $\cos 0 = 1$ and $\tan 0 = 0$, the equation reduces to $y = -\frac{g}{2v_0^2}x^2$ which indeed corresponds to the equation for initial horizontal motion. If it hadn't, the first thing to check for is an algebra error.

   Suppose $\alpha = +\pi/2$ or $90°$. The trajectory equation no longer works, since $\cos(\pi/2) = 0$. However, $x = (v_0 \cos\alpha)t$ and $y = -\frac{1}{2}gt^2 + (v_0 \sin\alpha)t$. Because, $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$, $x = 0$ and $y = -\frac{1}{2}gt^2 + v_0 t$, the equations for motion of an object launched vertically, so the analysis is again consistent. [Suppose α is taken to be $-\pi/2$. What does this mean? Do the equations reduce to a sensible result?]

2. What is the character of the general trajectory equation? Is it still a parabola? The equation $y = ax^2 + bx$ is a parabola with its axis parallel to the y-axis, but with its vertex displaced from the origin. Since the trajectory equation is of this form, it does represent a parabola [see Figure].



3. Where is the vertex of the parabola located in the $(x,\ y)$ coordinate system, and when is this position attained? That is, what are the values of t, x, and y when $v_y = 0$? These values are to be obtained from solving simultaneously equations [5], [7], and [3]. The following conclusions should

be verified:

$$t_{vertex} = \frac{v_0 \sin\alpha}{g} \tag{41}$$

$$h = y_{vertex} = \frac{v_0^2 \sin^2\alpha}{2g} \tag{42}$$

$$x_{vertex} = \frac{v_0^2 \sin\alpha \cos\alpha}{g} = \frac{v_0^2 \sin 2\alpha}{2g} \tag{43}$$

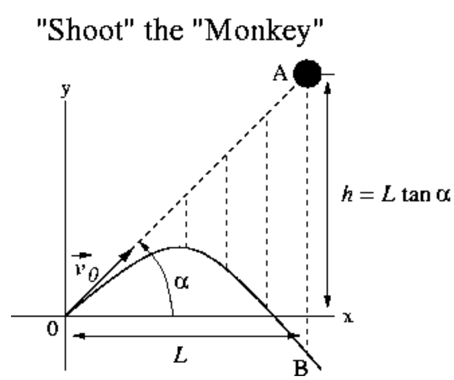
The last equation makes use of the trigonometric relation $\sin 2\alpha = 2\sin\alpha\cos\alpha$. The second equation gives the height above launch level reached by the projectile.[10]

31. (a) How does $y_{\mathbf{vertex}}$ change if the initial velocity v_0 is increased?

(b) How does $y_{\mathbf{vertex}}$ change if the angle α is decreased?

4.25 Targets and ranges

The *Shoot-n-Drop* video (simultaneous free-fall release of one ball with horizontal projection of another ball) indicates that a projectile, aimed at an object suspended at the same horizontal level, will strike the object if the latter is dropped at the instant the projectile is launched, assuming the horizontal distance to the object does not exceed the range of the projectile. What happens if instead the projectile is aimed at a target not at the same horizontal level [see Figure]? Suppose the target is released from position $(L,\ h)$ at the instant the projectile is launched. What are the vertical positions, $y(L)$ of the projectile and $y'(L)$ of the target, after the time interval



$$\Delta t = t - t_0 = t = \frac{L}{v_{0x}} = \frac{L}{v_0 \cos\alpha}$$

it takes the projectile to reach the horizontal position L of the target?

From the equation of trajectory ([8] on page 40), the projectile is at

$$y(L) = -\frac{g}{2v_0^2\cos^2\alpha}L^2 + L\tan\alpha.$$

[10]Note that if the launch level, y_0, is something other than $y_0 = 0$, this number is added to y_0.

From the kinematic equation of vertical fall, with $y_0 = h$, the target, at the same instant, will be at

$$y'(L) = -\frac{1}{2}gt^2 + h = -\frac{g}{2v_0^2\cos^2\alpha}L^2 + L\tan\alpha.$$

This results from substituting $t = \frac{L}{v_0\cos\alpha}$ (found above) and $h = L\tan\alpha$ (see figure).

Thus, $y(L) = y'(L)$–the projectile and target are at the same place at the same instant: the projectile strikes the target. Depending on the magnitude of the initial velocity v_0, the two objects may meet, if v_0 is sufficiently large, while the projectile is still rising, or, for smaller v_0, as the projectile is descending, say at point B in the figure. However, If the ground is located at level $y = 0$, v_0 would have a critical value below which the projectile and target would reach the ground before meeting:

$$v_{0,\ \mathrm{crit}} = \sqrt{\frac{gL}{\sin 2\alpha}}. \tag{44}$$

32. Interpret results of this analysis:

(a) In terms of L and α, what must v_0 equal if the objects are to meet at the projectile's maximum height? How large, then, must v_0 be for the objects to meet while the projectile is still rising?

(b) Derive Equation 44. That is, in terms of L and α, what must v_0 equal if the objects are to meet at $y = 0$? How small must v_0 be for the objects to meet below the level at which the projectile is launched?

(c) Under what conditions will the objects fail to meet?

Please view the following video:
`https://www.youtube.com/watch?v=0jGZnMf3rPo`

An alternative interpretation of these results: Motion along the parabolic trajectory may be thought of as a superposition of a uniform motion at velocity v_0 along straight line $0A$ combined with free fall through various distances from this line [see dotted lines in the previous figure].

Two particularly interesting numbers describing the flight of a projectile are its time of flight and its horizontal range. Consider the case in which a projectile is launched over horizontal terrain so that it starts at level $y = 0$ and returns to level $y = 0$. The time of flight is found by setting $y = 0$ in the equation $y = -\frac{1}{2}gt^2 + (v_0\sin\alpha)t = 0$. This equation has two solutions: $t = 0$, which corresponds to the start of the motion at $(0,\ 0)$, and a second value of t, denoted by t_R, given by

$$t_R = \frac{2v_0\sin\alpha}{g}, \tag{45}$$

which is the time of flight.

The horizontal range is obtained by setting $y = 0$ in the equation of trajectory, giving $\frac{-g}{2v_0^2 \cos^2 \alpha} x^2 + \frac{\sin \alpha}{\cos \alpha} x = 0$, which also has two solutions. The solution $x = 0$ corresponds to the starting point. The other solution, denoted by R [refer to the figure on page 41], is

$$R = \frac{2v_0^2}{g} \sin \alpha \cos \alpha = \frac{v_0^2}{g} \sin 2\alpha. \tag{46}$$

This equation demonstrates how the range of a projectile over horizontal terrain depends on the magnitude and elevation of its initial velocity. If the initial velocity v_0 (often referred to as the "muzzle velocity") of the projectile is increased while keeping α constant, the range increases. This is not surprising, but it is not obvious that the range depends on the *square* of the muzzle velocity and so increases by a factor of four when the initial velocity is doubled.

If v_0 is held constant while α is increased from low values, R increases until $\alpha = \pi/4$ or $45°$, so that 2α becomes $\pi/2$ and the sine function attains its largest value of $+1$. As α increases beyond this value, $\sin 2\alpha$ decreases. This means that the horizontal range R achieves its maximum value (for a fixed muzzle velocity) at an initial velocity elevation of $45°$.

33. Using $R = \frac{v_0^2}{g} \sin 2\alpha$, prove that launches with initial angular elevations equally bigger and smaller than 45 degrees have equal ranges, all shorter than the range at 45 degrees.. [Hint: set $\alpha = 45° + \beta$, expand $\sin 2\alpha$ with the trigonometric identity for $\sin(A + B)$, and interpret the results for cases where β has positive and negative values of equal magnitude.]

Notice, first, how mathematical this disucssion has been. Math, after all, is the language of physics. Also notice that, although the approach is based on experimental results, the discussion goes beyond the empirical. After validating the basic theory with experiment, further results and consequences are pursued with reason. But it's still worth asking: to what extent does the theory "explain" or provide understanding of, for example, the fact that maximum range is attained at a $45°$ angle of elevation? The discussion shows that this prediction is a necessary consequence of:

1. the definitions of velocity and acceleration,

2. the fact that free fall is uniformly accelerated, and

3. the fact that the two motions at right angles to each other are independent.

It is in terms of these basic ideas, which enable many other details of motion to be predicted, that the special property of a $45°$ angle is explained and the various other properties of the motion are predicted and unified by one simple theory. But why free fall is uniformly accelerated, and why the velocity components are independent, are ideas which have no explanation in the present level of inquiry. By taking these for granted (as was already pointed out, they are not absolutely correct), a partial understanding, deeper than a view of the observed phenomena as unrelated, empirical facts, has been gained. But this understanding falls short of explaining the uniformity of acceleration and the independence of velocity components.

A next level of inquiry addresses the origin of the restrictions, and, if these can thereby be explained in terms of still more fundamental ideas, understanding deepens further. This chain of inquiry has no end. Beneath each level attained is a still deeper level that may answer old–and raise new–questions.

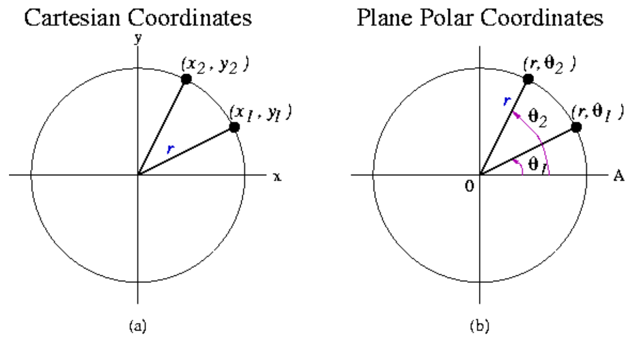
4.26 Superposition of Velocities: Circular Motion

To this point, the focus has mainly been on describing straight-line motion. Even in the curvilinear case, projectile motion, a vertically accelerated rectilinear motion was superposed on a horizontally uniform one. The intention behind limiting the scope in this way was a gradual build up of concepts and insights with the simplest physical abstractions. Not surprisingly, though, the simplest cases are neither necessarily the most prevalent nor the most interesting. In truth, rectilinear motions comprise a very small fraction of all motions. Far more frequent are repetitive motions having curved paths. Yet, even among these, there is a simplest case, that of circular motion, like, for example, a bob on a string, a rotating wheel, a hand on a(n analog) clock, Earth on its axis, electrically charged particles in a uniform magnetic field, and (to a first approximation) planets around the sun.

4.27 Coordinate Systems and the Concept of Angular Velocity

Describing the position of an object requires a system of position coordinates, For rectilinear motion, position coordinates along a number line were used. For two-dimensional projectile motion, Cartesian coordinates formed by two number lines at right angles were used.

Circular motion in which *angular* displacements are equal in successive equal time intervals is called uniform circular motion, in analogy with uniform rectilinear motion. If positions of a particle executing uniform circular motion are represented by Cartesian coordinates, $(x,\ y)$, [see Figure, left] a relatively complex equation, $x^2 + y^2 = r^2$, describes the trajectory. Note that x and y vary with time in a nonuniform way. On the other hand, if “plane polar coordinates” $(r,\ \theta)$, [see Figure, right] are employed, a very simple description emerges. More importantly, angular coordinates, θ_1 and θ_2 are precisely the numbers referred to when describing angular displacement and the uniformity of the motion.



In the plane polar reference system following the standard trigonometric convention, the positive direction for angular position is counterclockwise, and the negative direction is clockwise, from the positive half of the horizontal axis whose origin is the center of the circle. Angular position numbers, θ, then have significance and arithmetic properties completely analogous to the position numbers, s, of rectilinear motion. On the other hand, position numbers r, on any radial line at angular position θ, are marked off *positively* from the center point O; negative values of r have no geometrical significance and no physical interpretation.

With plane polar coordinates, $(r,\ \theta)$, the approach used for rectilinear motion can be followed directly to describe circular motion. Numbers such as $\frac{\Delta r}{\Delta t}$ and $\frac{\Delta\theta}{\Delta t}$ can be calculated and recognized as velocities. The condition for *circular* motion is that Δr (and therefore $\frac{\Delta r}{\Delta t}$) is zero. The quantity $\frac{\Delta\theta}{\Delta t}$ is called average angular velocity. From here, uniform, varying, and instantaneous velocities can be defined, as in the rectilinear case. The quantity $\frac{\Delta\theta}{\Delta t}$ may be positive, negative or zero, and these algebraic properties are interpreted just as they are with $\frac{\Delta s}{\Delta t}$.[11]

Angular velocity is usually symbolized with the Greek letter ω ("omega"). Thus, for average angular velocity

$$\overline{\omega} \equiv \frac{\Delta\theta}{\Delta t} \tag{47}$$

and for instantaneous angular velocity

$$\omega \equiv \lim_{\Delta t\to 0} \frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt}. \tag{48}$$

Thus, instantaneous angular velocity in circular motion is defined as the time derivative of the angular position function.[12]

[11]It may be worth reviewing that discussion.

[12]Angles, like all arguments of transcendental functionis, are dimensionless quantities. They are, however, assigned units: radians or degrees. Angular velocity is usually expressed in terms of radians (rad) per unit time. The radian is the ratio of two lengths (and therefore dimensionless), the arc length Δs subtended by an angle $\Delta\theta$ and the radius r of the circle of which the arc is a part:

$$\Delta\theta = \frac{\Delta s}{r}.$$

34. An object in uniform circular motion sweeps out 15 rad/sec (that is, $\omega = 15$ rad/sec). Sketch a polar coordinate diagram showing a few positions of the particle at successive 0.1-sec intervals. What would be the significance of $\omega = -15$ rad/sec? (Don't spend time on an accurately constructed plot; make a freehand sketch estimating sizes of relevant angles.)

4.28 Kinematics of Uniform Circular Motion (ω = Constant)

Consider a motion at uniform angular velocity: $\frac{\Delta\theta}{\Delta t} = \omega$, where ω is a constant. Rearranging:

$$\Delta\theta = \omega\Delta t \quad \text{or} \quad \theta_2 - \theta_1 = \omega(t_2 - t_1). \tag{49}$$

This equation is a kinematic equation, exactly analogous to $\Delta s = v\Delta t$. If θ is plotted against t with different values of ω, the results are straight-line diagrams exactly like those from the rectilinear equation, with similar interpretations of ω as the slope of the straight line, positive or negative depending on the direction of the circular motion.

It is often convenient to characterize or compare circular motions by means of numerical properties other than angular velocity, ω. For example, one might utilize the time interval T needed to execute a complete revolution, a displacement $\theta_2 - \theta_1 = 2\pi$ rad. Substituting these symbols in the kinematic equation 49 gives $2\pi = \omega T$, or

$$T = \frac{2\pi}{\omega}. \tag{50}$$

The interval T is called the "period" of the motion.

An alternative way of describing the motion is to cite the number of revolutions, f, executed per unit time. From the definitions of the two quantities f and T, it is evident that

$$f = \frac{1}{T} = \frac{\omega}{2\pi}. \tag{51}$$

Another view would be to argue directly that if f represents the number of revolutions each second and each revolution consists of an angular displacement of 2π rad, then $2\pi f$ rad/s are swept out in the motion. Thus,

$$\omega = 2\pi f. \tag{52}$$

The number of revolutions per unit time, f is called the "frequency."

The arc length (circumference) of a full circle is $\Delta s = 2\pi r$, and so subtends the angle $\Delta\theta = 2\pi$ rad. In terms of degrees, which is also a dimensionless unit, the angle subtended by a full circle is 360°, so

$$2\pi \text{ rad} \equiv 360^\circ$$

35. **What is the period of rotation of Earth on its axis? the period of its revolution around the sun? Calculate the frequency of each of these motions. Calculate the angular velocity of each of these motions. Do these numerical results represent instantaneous or average velocities? Explain and justify your answers carefully.**

36. **Given that a circular motion is uniform (ω constant), what is the numerical value of $d\omega/dt$? Give an interpretation of this result using the vocabulary of derivatives, acceleration, etc.**

4.29 Kinematics of Uniformly Accelerated Circular Motion

If the angular velocity of a rotating object is a function of time, the rate of change of angular velocity is described by the quantities

$$\overline{\alpha} \equiv \frac{\Delta\omega}{\Delta t}, \tag{53}$$

$$\alpha \equiv \lim_{\Delta t \to 0} \frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt}, \tag{54}$$

where α is the symbol for angular acceleration. These definitions are obviously modeled on the definitions of linear acceleration.

For the special case of motion at constant angular acceleration, $\alpha = \frac{\Delta\omega}{\Delta t}$, is a constant. Then, once again rearranging, $\Delta\omega = \alpha\Delta t \Rightarrow$

$$\omega(t) = \alpha(t - t_i) + \omega_i, \tag{55}$$

and, following the same line of reasoning used in rectilinear motion,

$$\theta(t) = \frac{1}{2}\alpha(\Delta t)^2 + \omega_i \Delta t + \theta_i. \tag{56}$$

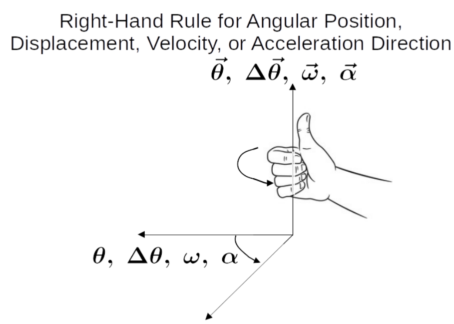
Solving $\omega(t) = \alpha\Delta t + \omega_i$ for Δt and substituting into the last result,

$$\omega(t)^2 - \omega_i^2 = 2\alpha(\theta(t) - \theta_i) = 2\alpha\Delta\theta. \tag{57}$$

These kinematic equations should be recognized as exact counterparts to the rectilinear equations derived previously. All these derivations can be arrived at more efficiently with methods of calculus. The calculus, a more powerful mathematical tool than analytic geometry, was developed to deal with continuous change and infinitesimal time, and, with it, quantitative relationships can be obtained, which, in more complex cases, are completely beyond the scope of algebra and geometry.

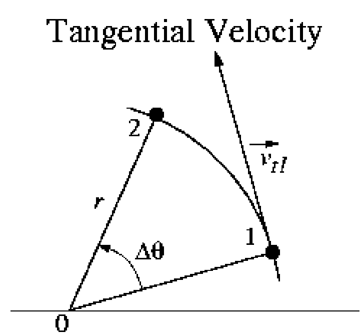
4.30 Treating Angular Position, Velocity, and Acceleration as Vectors

Angular position, angular velocity, and angular acceleration appear to have the character of vectors, requiring the specification of both magnitudes and directions. The direction, instead of being indicated directly by a straight arrow, is often shown as a curved arrow in either the clockwise or counterclockwise direction. A more formal way, used in vector algebra calculations, is to employ a so-called right-hand rule [see Figure]: while curling the fingers of the right hand in the direction of motion (for angular position and velocity) or the direction of change in rate of motion (for angular acceleration), an extended thumb gives the vector direction (which will be perpendicular to the plane of angular motion or change in angular motion). Then, angular position can be written as a vector with the symbol $\vec{\theta}$, angular velocity can be written as a vector with the symbol $\vec{\omega}$, and angular acceleration can be written as a vector with the symbol $\vec{\alpha}$.



4.31 Non-rigorous Argument Concerning the Relation between Instantaneous Angular and Linear Velocities

Consider an object in circular motion [see Figure]. The object is located at position 1 at instant t_1, and at position 2 at instant t_2, undergoing angular displacement $\Delta\theta$ in time interval Δt. What can be said about its instantaneous velocity at position 1? According to the scheme developed for trajectories, the object has an instantaneous velocity $\vec{v}_{t1}$, tangent to the circle at position 1. Let the magnitude of the vector velocity $\vec{v}_{t1}$ be denoted by the symbol v_{t1} and the magnitude of the instantaneous angular velocity by ω_1.



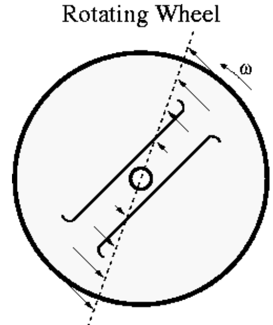
Now, ω_1 can be interpreted as the magnitude of the angle swept out in unit time, if the object continued at this angular velocity. Because arc length $\Delta s = r\Delta\theta$, and $\omega = \frac{\Delta\theta}{\Delta t}$, the angular displacement in one unit of time $\Delta t = 1$ is $\Delta\theta = \omega\Delta t = \omega$, so the length of arc swept out in one unit of time is $\Delta s = r\omega$. If, from position 1, the particle went into purely rectilinear motion at uniform velocity, it seems reasonable to expect that the magnitude of this velocity to be equal to the length of arc swept out in one second if the object instead continued in uniform circular motion. And so, it is plausable to guess that $v_{t1} = r\omega_1$. In general, for any position in the circular motion, the instantaneous "tangential" velocity would be written[13]

$$v_t = r\omega. \tag{58}$$

In light of this equation, it should be understood that in the case of a wheel, for example, although the particles of which it is comprised all have the same angular velocity, their tangential velocities vary systematically, increasing linearly as their radial distances from the axis of rotation increase.

If a wheel of outer radius R rolls on the ground without slipping at angular velocity ω about its center [see Figure], a length of arc $R\omega$ is brought into contact with the ground each succeeding unit of time. Under these circumstances, the wheel as a whole moves along the ground with a translational or linear velocity $R\omega$. Thus, when rolling takes place without slipping, the translational velocity of a wheel is equal to the magnitude of the tangential velocity of a point on its periphery.

Rotating Wheel



If angular velocity varies with time, so does the tangential velocity: $v_t(t) = r\omega(t)$, where now each velocity's dependence on time is indicated.

A concept of tangential acceleration may be defined by an exactly parallel argument, leading to the plausible relation

$$a_t = r\alpha, \tag{59}$$

where α denotes the instantaneous angular acceleration, defined by

$$\alpha \equiv \frac{d\omega}{dt}. \tag{60}$$

37. During the time interval $\boldsymbol{\Delta t}$, a wheel has an angular velocity given by the equation $\boldsymbol{\omega(t) = \omega_1 - \alpha(t - t_1)}$, where $\boldsymbol{\omega_1}$ and $\boldsymbol{\alpha}$ are both positive, and, eventually, $\boldsymbol{\omega_1 < \alpha\Delta t}$, when $\boldsymbol{\Delta t}$ becomes large enough.

[13] A more rigorous proof requires only basic command of differential calculus:

$$\Delta s = r\Delta\theta \Rightarrow \frac{\Delta s}{\Delta t} = r\frac{\Delta\theta}{\Delta t} \Rightarrow \lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t} = r\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t} \Rightarrow \frac{ds}{dt} = r\frac{d\theta}{dt}$$

Since $\frac{ds}{dt}$ is the instantaneous velocity at a point with direction tangent to the displacement function at that point, and since $\frac{d\theta}{dt}$ is the instantaneous angular velocity,

$$v_t = r\omega$$

(a) Describe what you would see the wheel doing if you watched it during the entire interval Δt for which the equation is valid.

(b) Sketch (not formal) angular acceleration versus time, angular velocity versus time, and angular position versus time plots for this motion.

(c) Obtain expressions for angular acceleration α and angular position θ as a function of t. Are the sketches and formulas consistent?

(d) Through what angle does the wheel turn from instant t_1 to the instant at which it reverses the direction of its rotation?

(e) The wheel has radius R. Find the instantaneous values of tangential velocity and tangential acceleration at radius R at an instant t_2.

(f) If the wheel rolls on the ground without slipping, how far would it roll in the interval between t_1 and t_2?

38. Watch: `https://www.youtube.com/watch?v=8H98BgRzpOM`

In a demonstration of gyroscopic behavior, a bicycle wheel is spun around a shaft. The magnitude of the wheels' angular velocity is measured to decrease from, say, ω_1 to ω_2 in an interval $\Delta t = t_2 - t_1$. The decrease is presumably caused by air resistance and friction in the bearings.

(a) To use angular kinematic relations to make calculations and predictions about the behavior of the wheel, what idealizations, assumptions, and limitations are implied?

(b) Indicating the positive direction, defining variables, explaining all steps, and pointing out where and how the assumptions and idealization just listed are invoked, predict the total angle through which the wheel will rotate in the time interval $\Delta t = t_2 - t_1$.

(c) In a second demonstration, the same wheel is given a spin such that its initial angular velocity is ω_i. What will be the angular position of a point on the wheel at clock reading t after the wheel is spun? Define all terms in you expression.

39. Earth's mean radius is 6370 km (3960 mi). Calculate the tangential velocity (in km/hr and mi/hr) associated with Earth's diurnal rotation of a point

(a) on the equator;

(b) at 30 degrees North latitude;

(c) at 45 degrees North latitude;

(d) at 60 degrees North latitude;

(e) at the North Pole.

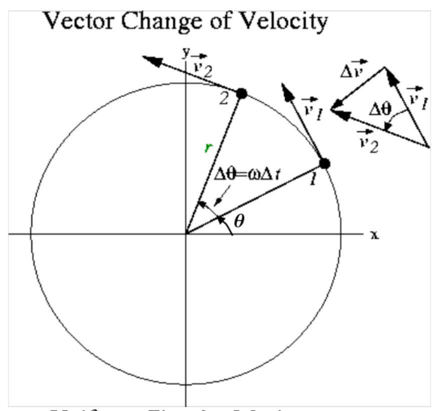
(f) Estimate this velocity at a point on the GMU campus (at roughly 40 degrees North lattitude).

40. (a) Calculate the frequency [in Hz], angular velocity [in rad/sec], and tangential velocity [in km/hr or m/s] of the moon in its orbit around Earth (mean orbital radius is about 400,000 km; orbital period is about 27 days, 8 hours).

(b) Calculate the frequency, angular velocity, and tangential velocity of an Earth-launched satellite, orbiting nearly circularly about 300 km above Earth's surface with a 90-min period.

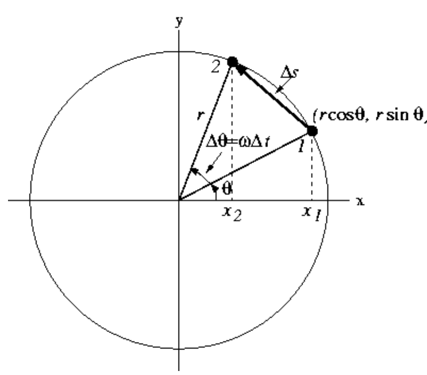
4.32 Linear Acceleration in Uniform Circular Motion

Since the instantaneous velocity in circular motion is always tangential to the circle, the direction of this velocity must be continually changing, even if its magnitude, $v_t = r\omega$, remains constant (ω = constant). When an object's velocity changes, whether in magnitude, direction, or both, the object undergoes an acceleration.



Consider the vector diagram of the figure, which illustrates the vector change of an object's tangential velocity $\Delta\vec{v}$ occurring in the time interval Δt as it undergoes circular motion.

Uniform Circular Motion



At any instant, a vector $\vec{r}$ points to the object's position on the circle. That position has plane polar coordinates, $(r,\ \theta)$, which in Cartesian coordinates are $(x,\ y)$. The relationship between plane polar and Cartesian coordinates is given by projections onto the Cartesian axes. For example at point 1 [see Figure], $(x_1,\ y_1) = (r\cos\theta,\ r\sin\theta)$. At point 2, after an angular displacement of $\Delta\theta$, the projections are $(x_2,\ y_2) = (r\cos(\theta+\Delta\theta),\ r\sin(\theta+\Delta\theta))$.

In circular motion, the magnitude of r remains constant; only θ changes with time. The average angular velocity is $\bar{\omega} = \frac{\Delta\theta}{\Delta t}$, and the average (tangential) velocity is $\bar{v} = (\bar{v}_x,\ \bar{v}_y) = (\frac{\Delta x}{\Delta t},\ \frac{\Delta y}{\Delta t})$.

Of course, the limits of these expressions as Δt becomes infinitesimal are instantaneous angular velocity and instantaneous velocity, both given by derivatives, $\omega = \frac{d\theta}{dt}$ and $v = (\frac{dx}{dt},\ \frac{dy}{dt})$.

In any case, $\Delta x = x_2 - x_1 = r\cos(\theta + \Delta\theta) - r\cos\theta$, and $\Delta y = y_2 - y_1 = r\sin(\theta + \Delta\theta) - r\sin\theta$. Expanding[14]

$$\Delta x = r(\cos\theta\cos\Delta\theta - \sin\theta\sin\Delta\theta - \cos\theta)$$
$$\Delta y = r(\sin\theta\cos\Delta\theta + \cos\theta\sin\Delta\theta - \sin\theta)$$

As Δt becomes smaller, so does $\Delta\theta$, and as $\Delta\theta$ becomes smaller $\cos\Delta\theta \rightarrow\approx 1$ and $\sin\Delta\theta \rightarrow\approx \Delta\theta$.[15] Thus,

$$\Delta x \rightarrow r(\cos\theta \times 1 - \sin\theta \times \Delta\theta - \cos\theta) = -r\sin\theta \times \Delta\theta = -y\Delta\theta$$
$$\Delta y \rightarrow r(\sin\theta \times 1 + \cos\theta \times \Delta\theta - \sin\theta) = r\cos\theta \times \Delta\theta = x\Delta\theta$$

because $x = r\cos\theta$ and $y = r\sin\theta$ (these results are written for any angle θ, not just the specific angle in the figure). Therefore,

$$\bar{v}_x = \frac{\Delta x}{\Delta t} = -y\frac{\Delta\theta}{\Delta t} \tag{61}$$
$$\bar{v}_y = \frac{\Delta y}{\Delta t} = x\frac{\Delta\theta}{\Delta t} \tag{62}$$

and, in the limit as $\Delta t \rightarrow 0$,

$$v_x = \frac{dx}{dt} = -y\omega \tag{63}$$
$$v_y = \frac{dy}{dt} = x\omega \tag{64}$$

That is,

$$\vec{v} = v_t = (v_x,\ v_y) = (-y\omega, x\omega) \tag{65}$$

which has magnitude $\sqrt{v_x^2 + v_y^2} = \sqrt{((-y)^2\omega^2 + x^2\omega^2} = \sqrt{(x^2 + y^2)\omega^2} = \sqrt{r^2\omega^2}$, or

$$v_t = r\omega \tag{66}$$

as found previously with a qualitative argument.[16] The direction of the instantaneous velocity is $\frac{v_y}{v_x} = \frac{x\omega}{-y\omega} = -\frac{x}{y}$, which shows that the direction of $\vec{v}$ is perpendicular[17] to the radius vector–that is, tangential, as was indicated.

[14] Recall, the trigonometric relations $\cos(A + B) = \cos A\cos B - \sin B\sin A$, and $\sin(A + B) = \sin A\cos B + \cos A\cos B$.

[15] Recall the series expansions, $\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots$, $\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots$. Once $\theta \ll 1$, θ^n is negligible for $n > 1$.

[16] The Calculus, using the chain rule, makes this easier and more efficient: $\frac{dx}{dt} = \frac{dx}{d\theta}\frac{d\theta}{dt} = r\frac{d\cos\theta}{d\theta}\frac{d\theta}{dt} = -r\sin\theta\frac{d\theta}{dt} = -y\omega$, and similarly for $\frac{dy}{dt}$.

[17] Two lines are perpendicular if the product of their slopes is -1.

The instantaneous components of the acceleration are

$$a_x = \frac{dv_x}{dt} = \frac{d(-y\omega)}{dt} = -\omega\frac{dy}{dt} = -v_y\omega = -x\omega^2 \tag{67}$$

$$a_y = \frac{dv_y}{dt} = \frac{d(x\omega)}{dt} = \omega\frac{dx}{dt} = v_x\omega = -y\omega^2. \tag{68}$$

or

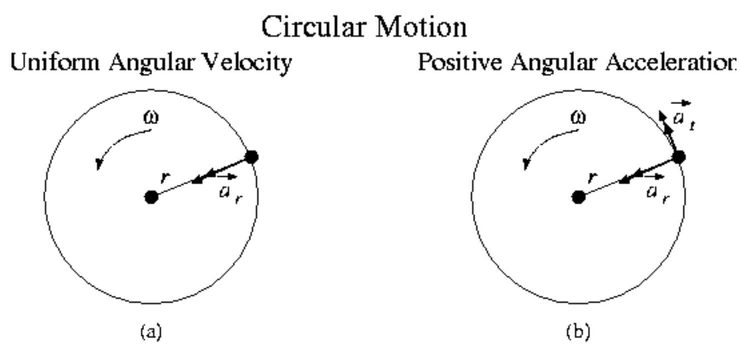
$$\vec{a} = (a_x,\ a_y) = (-x\omega^2,\ -y\omega^2) \tag{69}$$

The magnitude of the acceleration, then, is $\sqrt{a_x^2 + a_y^2} = \sqrt{(-x\omega^2)^2 + (-y\omega^2)^2} = \sqrt{(x^2 + y^2)\omega^4} = r\omega^2$, and its direction is $\frac{a_y}{a_x} = \frac{-y\omega^2}{-x\omega^2} = \frac{y}{x}$. Note, however, the canceling minus signs in the components; these indicate that the direction points opposite to that of the radius vector, that is, toward the origin. Because the acceleration is along but opposite the radial direction, the tangential component component of the acceleration is zero, while the magnitude of the radial component is

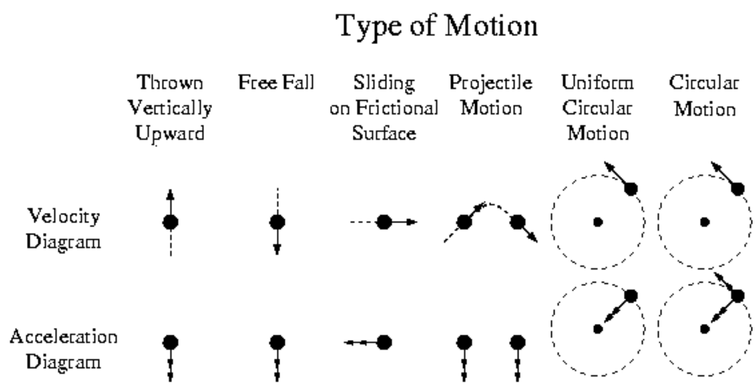
$$a_r = r\omega^2 = \frac{v_t^2}{r} = v_t\omega \tag{70}$$

since $v_t = r\omega$.

41. Show that the alternative expressions for a_r in the equation $a_r = r\omega^2 = \frac{v_t^2}{r} = v_t\omega$ follow from the relation $v_t = r\omega$.



The vector diagram of linear acceleration in uniform circular motion is drawn as in diagram (a) of the Figure. Note the use of a double arrowhead to distinguish acceleration from velocity. In nonuniform circular motion, diagram (b), with angular acceleration α, there exists a tangential component of acceleration $a_t = r\alpha$, as shown.



To recapitulate: In general, velocity and acceleration vector diagrams for a given motion differ markedly from each other [see Figure]. It is rarely the case that velocity and acceleration vectors are in the same direction.

In projectile motion, a uniform acceleration exists along one Cartesian axis and the magnitude of the linear velocity vector changes continually, In uniform circular motion, the magnitude of linear velocity does not change at all, but still the acceleration is not zero, since the direction of the linear velocity is changing continually. The magnitude of this acceleration is constant and directed toward the center of the motion *without change in the object's distance from the center of rotation.* This is the essential significance of "centripetal" acceleration. If, as part of its motion, the radial distance of an object from its center of rotation changes, the radial component v_r of its velocity would no longer be zero, and the motion would no longer be simply circular. If v_r changed with time, the object would have a radial acceleration, $\frac{dv_r}{dt}$ in addition to a centripetal acceleration, $r\omega^2$.

42. A car moving at velocity v enters a circularly curved section of road, the curve having a radius r. Calculate the centripetal acceleration.

43. Calculate the centripetal acceleration for each motion considered in questions 39 and 40. How do the centripetal accelerations at various places on Earth compare with the gravitational field strength near Earth's surface, g_E?

4.33 Relative Motion

Vector addition, employed to understand projectile and circular motions, is particularly useful when determining what are known as relative velocities. All velocities "are relative," since all vector quantities are measured relative to the coordinate system in which they are observed.

44. Reconsider the example on page 21 of the stone thrown upward at the edge of a cliff, except now as an observer in a helicopter rising at uniform velocity v_0 (the initial velocity of the

stone) relative to the ground. Indicating all observations from the helicopter with a prime symbol, $'$, assume both observers, the one who tosses the stone at the cliff edge and the one in the helicopter, begin measuring the stone's trajectory when the heliocopter reaches the cliff edge just as the stone is released–that is, at positions $s'_i = s_i = s_0 = 0$ and at clock readings $t'_i = t_i = t'_0 = t_0 = 0$. The helicopter continues to rise at the same rate, v_0, relative to the ground throughout the stone's motion. [Assume the helicopter observer also orients the position reference frame so that up is positive, but chooses the zero of that frame to be the helicopter.]

(a) What does the helicopter observer measure the stone's initial instantaneous velocity at $t'_0 = 0$, v'_0, to be?

(b) At instants $t' > t'_0$, how will the stone's motion appear to this observer? Will it appear to be always rising? to be always motionless? to rise, reverse direction, and then fall? to be always falling?

(c) To this observer, will the stone's velocity appear to be constant or varying? That is, will the stone appear to be accelerating? If so, in what direction? If measured somehow, then, what would be the value of a'?

(d) At the instant the observer on the cliff observes the stone to reverse direction, what will be the stone's motion according to the helicopter observer?

(e) What will the heliocopter observer measure the stone's velocity to be at the instant the cliff observer sees it reverse direction?

(f) What will the heliocopter observer measure the stone's velocity to be when the stone returns to s_0?

(g) Can the two observers agree on a standard time interval at which to make position measurements? That is, can they compare positions at the same instants? Will their position measurements be the same at each designated clock reading? Will their displacement determinations based on these measurements be the same?

(h) In answering these questions, what assumptions are being made regarding the relationship between the two observer's displacement and time interval measurements? Can they in principle agree on the meaning of a displacement and a time interval (on the lengths of a meter and of a second, for example)? That is, can they agree on measurement units and that these have equivalent values?

4.34 Relative Velocity

In common parlance, "relative velocity" usually refers to a velocity measured relative to a coordinate system which itself is moving relative to some "fixed" coordinate system.

Consider a state highway patrol vehicle pursuing at v_P a car traveling a straight road at $v_C < v_P$. The velocity vector for the trooper (P) *relative to the ground* (G) might look like:

$$G \xrightarrow{\qquad\qquad\qquad\qquad v_P \qquad\qquad\qquad\qquad} P$$

The speeder's vector (C) relative to the ground would then look like:

$$G \xrightarrow{\qquad\qquad\qquad\qquad v_C \qquad\qquad\qquad} C$$

And together, again, both relative to the ground:

$$G \begin{array}{l} \xrightarrow{\qquad\qquad\qquad\qquad v_P \qquad\qquad\qquad\qquad} P \\ \xrightarrow[\qquad\qquad\qquad\qquad v_C \qquad\qquad\qquad]{} C \end{array}$$

The velocity vector of *the patrol vehicle relative to the speeding car* would therefore look like:

$$C \xrightarrow{v_P - v_C} P$$

In words: the patrol vehicle approaches the car at the rate of $v_P - v_C$. Alternatively: the speeder sees in the rear-view mirror the trooper approach at $v_P - v_C$. Conversely, the trooper sees the speeder coming backward at $v_C - v_P$. The vector for the car relative to the patrol vehicle would be:

$$C \xleftarrow{v_C - v_P} P$$

This example illustrates a visual technique for dealing with relative velocity problems:

1. Treat given vector quantities literally, exactly as given.

2. Draw each vector to the same scale, representing the moving object with the arrow head and what the object is moving relative to with the tail.

3. The heads and tails of the given vectors, together with their corresponding labels, determine points in a vector space (in the example, the vector space is velocity vector space).

4. Draw vectors between any of the points in the vector space to determine any relative velocity, being sure to draw the vector toward the point representing the moving object and away from the point representing what the object moves away from.

45. A screw in an overhead grill of an elevator comes loose while the elevator is in motion. The sides of the elevator have height h, so, at the instant the screw leaves the grill, it is separated from the floor by a distance h Determine the fall time interval (that

is, the amount of time until the screw hits the elevator floor) when the motion of the elevator is

(a) moving at constant velocity, v, upward.

(b) moving at constant velocity, v, downward.

(c) accelerating upward at a rate of $\frac{g}{n}$, where g is the gravitational field strength and n is a positive integer.

(d) accelerating downward at a rate of $\frac{g}{n}$.

[Hint: As a vector, acceleration can be treated similarly to velocity.]

46. At $t = 0$, object A is at position $x_{0\mathrm{A}} = -a$, $y_{0\mathrm{A}} = 0$, and is moving with constant velocity $+v_x$. A similar object B is at position $x_{0\mathrm{B}} = 0$, $y_{0\mathrm{B}} = +b$ and moving with constant velocity $-v_y$.

(a) At $t = t$, where is A? Where is B?

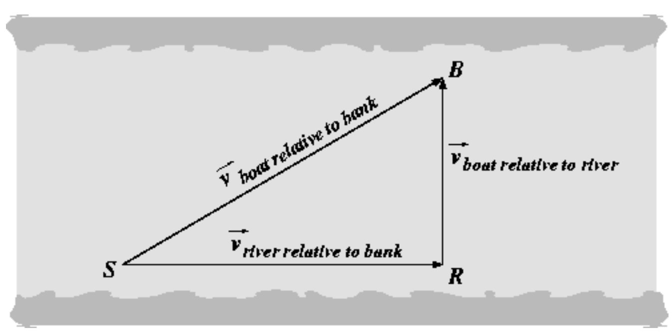
(b) What relationship must hold between a, b, v_x, and v_y if A and B are to collide regardless of their dimensions?

(c) If the relationship of (b) does not hold, how far apart are A and B at any time t?

(d) If the relationship of (b) does not hold, at what time t_d are A and B nearest each other?

(e) If the relationship of (b) does not hold, what is the minimum separation between A and B?

(f) Does the minimum separation between A and B determined in (e) reduce to zero when $av_y = bv_x$?



A row boat (B in the Figure), rowed as if intending to cross a flowing river (R) directly, moves diagonally relative to the banks of the river (S, as in "shore"), not perpendicularly as aimed, because, although the boat moves perpendicular to the river, but the river moves parallel to the banks. The result, for someone watching the boat from either bank or above, is diagonal motion, the vector sum of the two perpendicular uniform motions. In other words, the two perpendicular, uniform motions–the boat relative to the river, and the river relative to the banks–result in a uniform diagonal movement relative to the banks at the rate: $v_{\text{boat relative to bank}} = \sqrt{v^2_{\text{boat relative to river}} + v^2_{\text{river relative to bank}}}$ at an angle relative to the bank of $\theta = \tan^{-1} \frac{v_{\text{boat relative to river}}}{v_{\text{river relative to bank}}}$.

47. A river flows due south with a uniform velocity v_r relative to the bank. A rower on the west bank, who can paddle a row boat at a constant velocity v_b relative to still water desires to reach a point directly east on the opposite bank, a distance d.

(a) Explain why the rower cannot land directly opposite the launching point if $v_b \leq v_r$.

(b) Assuming $v_b > v_r$, so that landing directly across is possible, in what direction (at what angle relative to the bank) should the boat be directed? [Hint: what must the net velocity parallel to the river be in order to row straight across?]

(c) In terms of the parameters given (that is, in terms of v_r, v_b, and d), how much time will it take the rower to cross the river? [Note that $\sin(\arccos x) = \cos(\arcsin x) = \sqrt{1-x^2}$, because:

i. $-1 \leq \arccos x \leq 1$

ii. if $\phi \equiv \arccos x$ and $0 \leq \phi \leq \pi$, then $\cos\phi = x$ and $\sin(\arccos x) = \sin\phi$

iii. $\sin^2\phi + \cos^2\phi = 1$

iv. $\sin(\arccos x) = \sin\phi = \sqrt{1-\cos^2\phi} = \sqrt{1-x^2}$]

In setting up a coordinate system for such problems as these, make life easier by choosing axes parallel to at least one of the vectors–as was done in the case of projectile motion.

4.35 Reference Frames

The result of a velocity measurement depends on the relationship of the measurer to the object whose velocity is being measured. A measurer in a boat on the river in the previous example would measure the row boat's velocity just as the rower would–relative to the river only.

Consider another example which is of some historical importance. Galileo was an ardent defender of Copernicus's proposition that the preferred reference for analyzing motion in the solar system is fixed to the sun rather than to Earth. For many in Galileo's time, including scientists, the implications of this choice–to wit, that Earth both revolves around the sun and rotates on its own axis–were impossible to accept. If Earth were in motion, how could, for example, a stone dropped from a tower land at the base of the tower? The tower, attached to Earth, moving with it, would leave the stone "behind." How far behind? In question 39f, the tangential velocity of the surface of Earth near Washington, D.C. was calculated, assuming Earth completes a rotation in one 24-hour day. At this point, calculating how long it takes a stone to fall from the top of a 30-meter (100-foot) tower should be easy. The answer, of course, is that the tower would move hundreds of meters in the time it takes the stone to hit the ground, assuming a rotating Earth. And if Earth's revolution around the sun is also included? Since a dropped stone lands almost exactly below where it is released, does it make sense to say the Earth is moving?

Anyone understanding projectile motion and the superposition of motion would understand that it does make sense. The stone, in the hand of the person at the top of the tower, moves "horizontally" along with Earth, tower, and person. Its release and subsequent free fall in no way affects the horizontal motion it had at the instant of release. The falling stone behaves just like a projectile: its horizontal and vertical components of motion are independent of one another. Its horizontal velocity and that of the tower remain equal, and so the stone "keeps up."

Note that this experiment doesn't tell us whether or not Earth is moving, and certainly, therefore, not how fast, if it is. This is a very important observation: *no experiment in a uniformly moving reference frame* (i.e., one moving at constant velocity) *can measure the rate (even a rate of zero) of motion of the reference frame.* An object dropped or thrown in a jet aircraft flying at 1000 km/hr (600 mi/hr) at an altitude of 9 km (30,000 feet) appears to move just as if the plane were standing still.

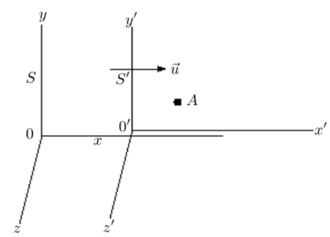
This realization serves as the basis for a fundamental rule of motion, known as the Galilean relativity principle, which states that no experiment can differentiate one uniformly moving reference frame from another uniformly moving reference frame. Another way of saying this is that the laws of physics hold identically in any laboratory moving at constant rectilinear velocity. Notice that this relativity principle, any more than the more inclusive one introduced by Einstein, does *not* say "everything is relative." Quite the contrary. It demands that certain relationships do not change–they are invariant–regardless of the uniformly moving reference frames in which they are investigated.

Uniformly moving reference frames are also referred to as "inertial reference frames," that is, an inertial reference frame is one that is not undergoing any kind of acceleration. In such a frame, an object at rest remains at rest, and an object moving in uniform rectilinear motion continues to move in a straight line at constant velocity, unless acted upon by something external to the object.

A coordinate frame fixed to Earth is not inertial, due to Earth's rotation about its axis and its revolution about the sun. This fact, however, does not alter the outcome of most experiments, since the effect is smaller than the accuracy with which many experiments are performed. The non-inertial motion of Earth, in fact, was not detectable by any measurement until 1851, when Foucault (1819-1868) suspended a heavy pendulum bob by a long flexible wire, and observed that the plane of vibration of the pendulum rotated with respect to the floor. Since there was no way the support point could twist the plane in which the pendulum oscillated once it had been started, Foucault concluded that the floor was rotating beneath the pendulum. This motion must be taken into account when trying to understand, for example, the circulation of the atmosphere, long-range flight, and the meaning of the term "centrifugal force," but, for most purposes, it may be ignored.

4.36 Transformations: Describing Physical Phenomena from Different Reference Frames

Two Inertial Reference Frames



The connection or relationship between numerical values describing motion or other physical phenomenon measured in different reference frames is called a "transformation." Consider two reference frames S and S' moving at uniform rectilinear velocity u relative to one another in the x-direction (that is, along the $x - x'$ axes) [see Figure]. Observed in the S frame, frame S' moves in the positive x-direction; observed in the S' frame, frame S moves in the negative x'-direction. The x- and x'-axes coincide, but the drawing separates them slightly. Clocks at the origins 0 and $0'$ in their respective frames are each set to zero at the instant the origins pass each other. The term "inertial," again, means that the two reference frames move only rectilinearly at constant velocity with respect to one another. An observer may choose the most convenient reference frame in which to make a measurement.

In the S frame, equations are written in terms of unprimed coordinates; in the S' frame, they are written in terms of primed coordinates. Suppose, in the S frame, a particle at position A is found to be moving with velocity v_x in the $+x$-direction, and in the S' frame its velocity is measured $v'_{x'}$. What is the connection between these two velocities? Relative to S', again, the particle moves at the rate $v'_{x'}$, and S' moves relative to S at the rate $+u$, so a resonable guess is that the velocity measured in S, v_x, will be the vector sum of these two velocities:

$$v_x = v'_{x'} + u.$$

Conversely,

$$v'_{x'} = v_x - u.$$

For motion in the y- or z-directions,

$$v_y = v'_{y'}$$
$$v_z = v'_{z'},$$

since S and S' do not move relative to one another in either of these directions.

This set of equations,

$$v_x = v'_{x'} + u \tag{71}$$
$$v_y = v'_{y'} \tag{72}$$
$$v_z = v'_{z'} \tag{73}$$

forms the transformation rules, frequently called "Galilean" or "Newtonian" transformation rules, between the two reference frames.

4.37 Length and Time in a Single Reference Frame

Implicit in the definition of reference frames has been the choice of a rigid set of coordinate axes for the measurement of position. These are scaled in some manner through the use of a rigid measuring rod. With them, positions are noted at certain instants, and then displacements, velocities, accelerations, etc., are calculated.

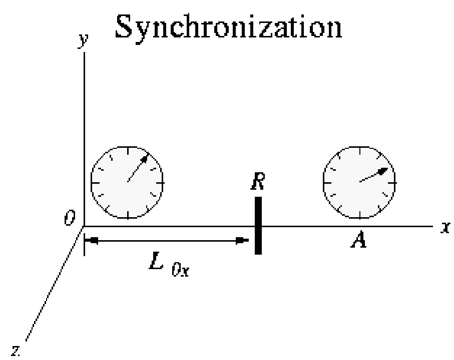
As for time, a unit based on some presumably uniform periodic process (such as the rotation of Earth, or the swinging of a pendulum, etc.) is devised. Clocks are then constructed that appear to repeat the unit interval faithfully, as indicated by their agreement with one another and with the primary reference interval. A given clock measures local time, the time at the clock's location in its reference frame. The term "event" refers to the simultaneous coincidence of a position number with a time number. For motion confined to a single dimension, events would have coordinates $(t,\ s)$: a particle is observed to be at position s on the coordinate axis and simultaneously the hand of a nearby clock points to the number t on its scale. Note this reasonable invocation of the notion of local simultaneity, a gut feeling that if something happens "here and now," then the where and when of this event are known. Perhaps so, but what can be known about "*there* and now?" What is meant when two widely separated events are said to take place simultaneously?

4.38 Synchronization

In order to accept that two events separated in space took place simultaneously, it is minimally necessary to assume that the clocks in the two locations agree with each other, that is, that they are "synchronized." How can this be assured? One apparently straight-forward means might be to build identical clocks, set them all going in the same location at exactly the same time with the same settings, and then transport them to their respective locations. This method relies on the implicit assumption that moving a clock has no influence on its time-keeping. Is this a valid assumption?

An alternative approach to synchronization is to transmit some kind of signal, say sound or light. The magnitude of the signal velocity, of course, needs to be known, but, since clocks in different locations are yet to be synchronized, this velocity can't be measured by measuring the elapsed time interval of a unidirectional displacement. Instead, a signal can be bounced off a reflector positioned at a known distance and the elapsed time of the round-trip can be measured with a single clock [see Figure]. Suppose the distance from the signal source (and the local clock) at the origin 0 to the reflector R is L_{0x}. After a pulse is sent from 0, its return after reflection from R is observed. With the clock at 0, the time interval between two events–the emission of the pulse [described by the numbers $(t_1,\ 0)$] and the return of the pulse [described by the numbers $(t_2,\ 0)$]–is recorded. The average velocity $\bar{v} = 2L_{0x}/(t_2 - t_1) \equiv c$, where the symbol c identifies the velocity of the synchronization signal.

Now, assuming that the signal velocity is the same regardless of direction, all clocks along the x-axis, none of which are running, are preset to the time it will take a signal from the origin to reach them. We arrange a mechanism to start each clock on the arrival of the signal. The clock at 0 starts on emission. The clock at, say, A in the Figure, is preset to $\overline{0A}/c$, where $\overline{0A}$ is the distance from the origin to position A, and starts with the arrival of the pulse. In this way, clocks all along the x-axis are synchronized.



Understand that this is not a matter of "seeing," or of appealing to an intuitive sense of local simultaneity. This operation for measuring time at different locations has become inextricably tied up with the measurement of space. More than just selecting a periodic device and adopting a unit, setting up a time scale involves prior definition of a position scale.

48. Image a clock on Earth and a clock on the moon. How can these clocks be synchronized? Suppose a signal from Earth is to be sent toward the moon at noon local time. To what time should the moon clock be set so that the moon clock will start synchronized with the Earth clock? [The moon is on average 384,000,000 m from Earth, and $c_{\mathbf{light}} = 3 \times 10^8$ m/s.]

4.39 Galilean (Newtonian) Transformations

Suppose an event occurs at point A of the Figure on page 61. In frame S, its position coordinates are $(x,\ y,\ z)$ at instant t; in frame S', its position coordinates are $(x',\ y',\ z')$ at instant t'. It may be assumed that clocks in each frame have been synchronized, and that a connection between clock readings in the two frames is established by calling $t = t' = 0$ the instant at which the origins 0 and $0'$ pass each other. The basis of this connection is the unexamined assumption that clock readings in the two frames will give the same number,

$$t = t'. \tag{74}$$

That is, time is assumed to flow, as Newton said, "without relation to anything external." A second is a second is a second, anywhere. Newton referred to this conception as "absolute time."

Given that the origins 0 and $0'$ coincide at the instant $t = t' = 0$ and that the relative velocity between the frames is u in the x-direction, at instant $t = t'$ the origins are displaced $ut = ut'$. Therefore, in general,

$$x = x' + ut' \quad x' = x - ut \tag{75}$$

$$y = y' \qquad y' = y \tag{76}$$

$$z = z' \qquad z' = z. \tag{77}$$

These, along with $t = t'$, constitute the Galilean or Newtonian transformation rules for position and time.

Let's say an observer in frame S observes an object displaced from x_1 to x_2 between instants t_1 and t_2, and an observer in S' observes the same object to be displaced from x'_1 to x'_2 in the interval from t'_1 to t'_2. The observer in S then determines that

$$v_x \equiv \frac{\Delta x}{\Delta t} = \frac{(x_2 - x_1)}{(t_2 - t_1)},$$

while the observer in S' determines that

$$v'_{x'} \equiv \frac{\Delta x'}{\Delta t'} = \frac{(x'_2 - x'_1)}{(t'_2 - t'_1)}.$$

But, using the position Galilean rule to transform x to x',

$$\begin{aligned} v_x &= \frac{(x_2 - x_1)}{(t_2 - t_1)} \\ &= \frac{(x'_2 + ut'_2) - (x'_1 + ut'_1)}{t'_2 - t'_1} \\ &= \frac{x'_2 - x'_1}{t'_2 - t'_1} + u\frac{t'_2 - t'_1}{t'_2 - t'_1}. \end{aligned}$$

Thus,

$$v_x = v'_{x'} + u, \tag{78}$$

which is the Galilean transformation rule for velocity in the x-direction (since there's no relative motion in the yy' or zz' directions, $v_y = v'_{y'}$ and $v_z = v'_{z'}$).

Turning to numbers that characterize acceleration in each frame,

$$a_x \equiv \frac{\Delta v_x}{\Delta t} = \frac{v_{x2} - v_{x1}}{t_2 - t_1},$$
$$a'_{x'} \equiv \frac{\Delta v'_{x'}}{\Delta t'} = \frac{v'_{x'2} - v'_{x'1}}{t'_2 - t'_1},$$

Using the velocity Galilean transformation rule,

$$\begin{aligned} a_x &= \frac{v_{x2} - v_{x1}}{t_2 - t_1} \\ &= \frac{(v'_{x'2} + u) - (v'_{x'1} + u)}{t'_2 - t'_1} \\ &= \frac{v'_{x'2} - v'_{x'1}}{t'_2 - t'_1}. \end{aligned}$$

Hence,

$$a_x = a'_{x'}. \tag{79}$$

The same number for the acceleration is found in each frame (also, obviously, $a_y = a'_{y'}$ and $a_z = a'_{z'}$.

Under the Galilean transformation rules, then, different frames give different numbers for the position and velocity components in the direction of relative motion. That is, the x-position and v_x numbers will not be the same in the two frames, but the other, orthogonal, position and velocity components, as well as all components of the acceleration will have identical values. Numbers that do not change under a transformation are referred to as being "invariant" under that transformation.

That the acceleration is invariant under the Galilean transformations is the key to understanding the impossibility of determining the absolute rate at which a uniformly moving frame moves or of distinguishing one such frame from another. The laws of mechanics are formulated in terms of accelerations. The predictions of such laws are verified in all frames moving with constant rectilinear velocity with respect to one another if they are verified in any one of them. They satisfy Galilean relativity.

49. A bar measuring L' in the S' reference frame lies (at rest) parallel to the x'-axis. If the x- and x'-axes are parallel, and the Galilean transformation rules are correct, what will be L, the length of the bar measured in the S reference frame, if

(a) the origins of the two frames are displaced but are stationary with respect to one another?

(b) the S' frame moves with velocity u in the $+x$-direction relative to the S frame?

(c) the S' frame moves with constant acceleration a in the $+x$-direction relative to the S frame?

4.40 The special, or restricted, theory of relativity

Under ordinary conditions, when physical phenomena are observed, light is utilized in a very direct way, and the inferences drawn from such observations implicitly assume that light propagates so rapidly that no account whatever need be taken of time intervals involved in its propagation. Such time intervals become significant, however, when observing the motions of objects very far apart or of objects moving at velocities comparable to that of light. This is because the speed of light is finite. Further, it is apparently constant in vacuum, regardless of the frame of reference in which it is viewed—as long as that frame moves with uniform rectilinear motion with respect to the vacuum. That is, the speed of light does not obey the Galilean transformation rule. This strange fact–that the speed of light is measured to have the same value regardless of the frame of reference–seemed reasonable enough to Albert Einstein that he made it one of the two basic assumptions underlying his special, or restricted, theory of relativity. The other assumption he made was a strict relativity: all laws of physics are of the same form in all reference frames that move relative to one another with uniform rectilinear motion. The consequences of these two postulates form the special theory of relativity.

4.41 Simultaneity

Fixing a clock at a point in a reference frame gives the time of events at that point. But how can an observer at one point obtain a clock reading at a distant point, or from a system that moves relative to the observer?

Einstein proposed an identical clock attached to every point in space. To be of any use, each must be synchronized with every other, as discussed in section 4.38. The "instant of any event" is then defined to be the reading made of the clock which is *at rest* where the event occurs. Then, if, when observing two events, two such clocks had the same reading, the two events are said to have occurred simultaneously.

50. Suppose that a light flash is emitted from the origin at the instant that points O and O' cross each other. If, as Einstein suggested, the velocity of light appears to be the same in all directions to any given observer, observers stationed at the origin in each frame will contend that, at later instants of time, the leading edge of the light pulse forms a sphere centered at the origin of the frame and that the observer in the other will have moved relative to the center of the sphere. Each observer will insist on their own contention and deny the argument of the other; there is no possibility of agreement between the two.

Sketch two separate pictures, each one illustrating the claim of one of the two observers. Note the symmetry or "reciprocity" of their points of view.

51. Suppose that an observer at position M' in the S' frame synchronizes clocks at points C' and D' equidistant to the left and right of M'. What operation would this observer perform to carry out this synchronization? How would an observer in S assess this synchronization, That is, would this observer agree that the two clock in S' are synchronized? If not, which clock would the S-based observer contend was started first?

Observers in different reference frames moving with respect to one another will not agree on the simultaneity of events.

4.42 Lorentz-Einstein Transformation Equations

Given the clock arrangements just described, consider, again, two reference frames, S and S', moving along their common x-x' axes with constant velocity of magnitude u. The clocks at the respective origins are started the instant these coincide, i.e., at $t = t' = 0$ when $x = x' = 0$, $y = y' = 0$, and $z = z' = 0$. All clocks in their respective frames are synchronized to the clocks at the respective origins. Since the motion of the frames is in the x-direction only, y and z coordinates remain coincident: $y = y'$ and $z = z'$. The problem, then, is to find the relationships among x, x', t, and t'.

Focus for the moment on the origin of the S' frame. Of course, this remains $x' = 0$ in S'. In S, it (the origin of the S' frame) will have the value $x = ut$, since it will be at coordinate x at instant t (after $t = 0$) according to the clock located at x, and its displacement after interval t will be ut because it moves at constant velocity u in the positive x-direction. A simple algebraic manipulation then gives $x - ut = 0$.

The transformation equation should be linear; otherwise, an event in one frame would correspond to two or more events in the other frame. And so, because these two equations, $x' = 0$ and $x - ut = 0$, representing the same thing, both equal zero, then the relationship between them and between (and, more generally, between any other x'-$(x - ut)$ pair must be one of proportionality (no additive term). The proportionality constant, usually signified with the Greek letter gamma, γ, must be independent of position and time, but it might depend on u, the relative velocity between the reference frames. Thus,

$$x' = \gamma(x - ut)$$

A similar argument with regard to the origin of S relative to S' gives

$$x = \gamma'(x' + ut')$$

If these two equations are going to have the same form, as required by Einstein's second postulate, $\gamma = \gamma'$. Galilean relativity gave $\gamma = 1$, but this may no longer be assumed. Rather, the first postulate, that the velocity of light is the same in both frames, must be ensured.

Solving the second equation for t' and substituting the first equation for x' yields:

$$\begin{aligned}
t' &= \frac{1}{u}\left(\frac{x}{\gamma} - x'\right) \\
&= \frac{1}{u}\left[\frac{x}{\gamma} - \gamma(x - ut)\right] \\
&= \frac{\gamma}{u}\left(ut + \frac{x}{\gamma^2} - x\right) \\
&= \gamma\left[t + \left(\frac{1-\gamma^2}{u\gamma^2}\right)x\right]
\end{aligned}$$

In order to see what these relations do with velocity, it is necessary to consider displacements and time intervals:

$$\begin{aligned}
\Delta x' &= \gamma(\Delta x - u\Delta t) \\
\Delta t' &= \gamma\left[\Delta t + \left(\frac{1-\gamma^2}{u\gamma^2}\right)\Delta x\right].
\end{aligned}$$

The average velocity $\overline{v'}$ in terms of the average velocity $\bar{v}$ is then:

$$\begin{aligned}
\overline{v'} &= \frac{\Delta x'}{\Delta t'} \\
&= \frac{\gamma(\Delta x - u\Delta t)}{\gamma\left[\Delta t + \left(\frac{1-\gamma^2}{u\gamma^2}\right)\Delta x\right]} \\
&= \frac{\frac{\Delta x}{\Delta t} - u}{1 + \left(\frac{1-\gamma^2}{u\gamma^2}\right)\frac{\Delta x}{\Delta t}} \\
&= \frac{\bar{v} - u}{1 + \left(\frac{1-\gamma^2}{u\gamma^2}\right)\bar{v}}.
\end{aligned}$$

In the coinciding limits $\Delta t' \to 0$, $\Delta t \to 0$, the relation between instantaneous velocities is found:

$$v' = \frac{v - u}{1 + \left(\frac{1-\gamma^2}{u\gamma^2}\right)v}.$$

This relation should apply in general, and so specifically when the velocity is that of the velocity of light c. According to Einstein's first postulate, this velocity should have the same value in both frames:

$$c = \frac{c - u}{1 + \left(\frac{1-\gamma^2}{u\gamma^2}\right)c},$$

from which is seen that γ can depend only on u (and c):

$$
\begin{aligned}
1+\left(\frac{1-\gamma^2}{u\gamma^2}\right)c = \frac{c-u}{c} = 1-\frac{u}{c} \\
\Rightarrow 1-\gamma^2 = -\frac{u^2}{c^2}\gamma^2 \\
\Rightarrow \gamma^2\left(1-\frac{u^2}{c^2}\right) = 1 \\
\Rightarrow \gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}} \\
= \frac{1}{\sqrt{1-\beta^2}},
\end{aligned}
\tag{80}
$$

once the assignment $\beta \equiv u/c$ is made.

With this clarification of γ, the transformation relations, known as the Lorentz-Einstein transformation equations, can be written out:

$$
t' = \gamma\left[t+\left(\frac{1-\gamma^2}{u\gamma^2}\right)x\right] = \frac{1}{\sqrt{1-\beta^2}}\left(t-\frac{\beta}{c}x\right) \tag{81}
$$

$$
x' = \gamma(x-ut) = \frac{1}{\sqrt{1-\beta^2}}(x-ut) \tag{82}
$$

$$
y' = y \tag{83}
$$

$$
z' = z. \tag{84}
$$

Recall that $\gamma = 1/\sqrt{1-\beta^2}$ and $\beta \equiv u/c$.

These may be solved for the reciprocal relations. Alternatively, the second postulate requires that the reciprical relations have the same form, and so the reciprocal relations may be arrived at simply by replacing u with $-u$:

$$
t = \frac{1}{\sqrt{1-\beta^2}}\left(t'+\frac{\beta}{c}x'\right) \tag{85}
$$

$$
x = \gamma(x'+ut') = \frac{1}{\sqrt{1-\beta^2}}(x'+ut') \tag{86}
$$

$$
y = y' \tag{87}
$$

$$
z = z'. \tag{88}
$$

These two sets of equations assign an event's position and associated clock reading values in a frame moving at constant velocity u along the x-axis with respect to the frame in which the event occurs at rest. The assignments are necessary outcomes of satisfying the postulates of special relativity. What may be deduced from them?

Measuring length in the two frames: Suppose a rigid rod AB, which is at rest in S and lies lengthwise along the x-axis, is measured. Then the length of the rod is $\ell = x_B - x_A$ in the S frame. Because the rod is at rest, it makes no difference at which instants the two ends of the rod are noted.

Using the second set of transformations to determine the length of the rod ℓ' measured by an observer in the S' frame:

$$\begin{aligned}\ell = x_B - x_A &= \frac{1}{\sqrt{1-\beta^2}}[(x'_B - x'_A) + u(t'_B - t'_A)] \\ &= \frac{1}{\sqrt{1-\beta^2}}[\ell' + u(t'_B - t'_A)].\end{aligned}$$

Therefore,

$$\ell' = \ell\sqrt{1-\beta^2} - u(t'_B - t'_A).$$

The rod's length in the frame with respect to which the rod moves with constant velocity $-u$ depends explicitly on when the positions of the ends are noted. If the position of the ends are noted simultaneously in S', $t'_B = t'_A$, and so,

$$\ell' = \ell\sqrt{1-\beta^2}, \tag{89}$$

the length measured in S' is shorter by the factor $\sqrt{1-\beta^2}$ than the length measured in S ($|\beta| \leq 1$, so $0 \leq \beta^2 \leq 1$). This effect is called "length contraction." The observer in S' sees the moving rod as being shorter.

Consider a rigid sphere of radius a at rest at the origin of S', so that it is described in that frame by the equation

$$x'^2 + y'^2 + z'^2 = a^2.$$

Viewed in S', the sphere remains spherical. But. transforming to S:

$$\frac{(x-ut)^2}{1-\beta^2} + y^2 + z^2 = a^2,$$

which represents a surface distorted from spherical in the x direction moving along the x-axis with velocity u. At $t = 0$, for example, when the origins of the two frames coincide

$$\frac{x^2}{1-\beta^2} + y^2 + z^2 = a^2,$$

so that the semi-major axes in the y and z directions still equal a, but the semi-major axis in the x direction is $a\sqrt{1-\beta^2}$.

Note that the shortening is reciprocal. If the surface were a sphere at rest in S, $x^2 + y^2 + z^2 = a^2$, then in S' it would appear an ellipsoid with the semi-major axis in the x' direction again equal to $a\sqrt{1-\beta^2}$. As u approaches c, the contraction in the x direction approaches complete flattening: that is, the semi-major axis along x approaches zero.

Measuring a time interval in the two frames: Now, consider a clock C' fixed at the origin of S' and for which $x' = 0$ always, while, from the perspective of S, the clock's position is $x = ut$. Consider, too, a set of clocks in S fixed at x_1, x_2, x_3, ..., all properly synchronized. C' reads t'_1 when it passes the clock at x_1, which itself reads t_1. Similarly, when it passes x_2, the corresponding clock readings will be t'_2 and t_2, respectively. And so forth. The transformation equations give

$$t'_1 = \frac{1}{\sqrt{1-\beta^2}}\left(t_1 - \frac{\beta}{c}x_1\right)$$
$$t'_2 = \frac{1}{\sqrt{1-\beta^2}}\left(t_2 - \frac{\beta}{c}x_2\right).$$

Thus,

$$t'_2 - t'_1 = \frac{1}{\sqrt{1-\beta^2}}\left[(t_2 - t_1) - \frac{\beta}{c}(x_2 - x_1)\right].$$

But,

$$x_2 - x_1 = u(t_2 - t_1),$$

and so

$$\begin{aligned} t'_2 - t'_1 &= \frac{1}{\sqrt{1-\beta^2}}\left[(t_2 - t_1) - \frac{\beta}{c}u(t_2 - t_1)\right] \\ &= \frac{1}{\sqrt{1-\beta^2}}(1-\beta^2)(t_2 - t_1) \\ &= \sqrt{1-\beta^2}(t_2 - t_1) \end{aligned} \tag{90}$$

Because $(1 - \beta^2) < 1$, the time interval on the clock in S' (which is the frame defined here to be moving) is shorter than those in S by the factor $\sqrt{1-\beta^2}$. The observe in S sees the clock at rest in S' moving slower.

Likewise, consider a clock C fixed in S passing in turn the clocks fixed at x'_1, x'_2, x'_3, ... in $'$, all of which are running synchronously. Denoting t_1 and t'_1 the clock readings of C and the clock at x'_1 when C passes, etc., the relation between the intervals in the two frames is, as above:

$$t_2 - t_1 = \frac{1}{\sqrt{1-\beta^2}}\left[(t'_2 - t'_1) + \frac{\beta}{c}(x'_2 - x'_1)\right].$$

But now, since the velocity of S with respect to S' is $-u$

$$x'_2 - x'_1 = -u(t'_2 - t'_1),$$

and so,

$$t_2 - t_1 = \sqrt{1-\beta^2}(t'_2 - t'_1). \tag{91}$$

Again, the time interval on the clock considered to be moving is shorter than the time interval on the clock considered stationary for the same two events. This effect, of moving clocks running slower than stationary clocks, is referred to as "time dilation."

So, is a moving rod *really* shorter? Does a moving clock actually run more slowly? Isn't it paradoxical that a clock moving along with S' appears to run slow when compared to a clock at rest with the observer in S, while this same clock moving along with S appears to run slow when compared to a clock at rest with the observer in S'? Such are the logical deductions of Einstein's special relativity theory. To be consistent with its postulates, a shorter length must be assigned to a rod moving at constant velocity with respect to an observer not moving along with it (than if it were at rest with respect to the observer). A shorter time interval between two events must be measured in a system moving uniformly relative to an observer than would be measured if the events were at rest in the observer's reference frame. Note that the practical consequences of these results–the effects on ordinary phenomenon–are negligible; they play a role only in experiments in atomic and cosmic domains. While special relativity nicely unifies apparently disparate phenomena (mechanics and electromagnetism), and remains free from internal contradictions, the outcomes of such experiments are its ultimate tests. And these have never produced a result that disagrees with the theory.

Finally, here listed are the Lorentz-Einstein velocity transformation equations:

$$v'_{x'} = \frac{dx'}{dt'} = \frac{v_x - u}{1 - \beta v_x/c} \tag{92}$$

$$v'_{y'} = \frac{dy'}{dt'} = \frac{v_y\sqrt{1-\beta^2}}{1 - \beta v_x/c} \tag{93}$$

$$v'_{z'} = \frac{dz'}{dt'} = \frac{v_z\sqrt{1-\beta^2}}{1 - \beta v_x/c} \tag{94}$$

Note the dependence of all velocity components on the relative motion of the reference frames and of the perpendicular components on the parallel component.

52. Derive the Lorentz-Einstein velocity transformation equations. [Hint: calculate $\frac{\Delta x'}{\Delta t'}$, etc.]

53. Suppose that an object is stationary in the S' frame (that is, $v'_x = v'_y = v'_z = 0$). What velocity does it have relative to the S frame according to the equations? Suppose, similarly, that an object is stationary in the S frame, what velocity do the equations give it relative to the S' frame ? Do these results make sense?

54. Verify that light traveling at $v'_x = c$ in the S' frame will be calculated as traveling at velocity $v_x = c$ relative to the S frame.

55. The Lorentz-Einstein velocity transformation rules should satisfy a "correspondence principle" (they must reduce, in the limit of low velocities, to the classical theory which is known to be successful and "correct" in that range of experience), that is, they should reduce to the classical Galilean velocity transformation rules when the velocity u is very small relative to c. Show that this is the case.

4.43 Waves

A stone dropped into a quiet pond, or the end of a rope or long helical spring (Slinky) sharply displaced, produces motions that are transmitted from one point to another in the water, or on the rope or spring, eventually affecting regions very remote from the location of the initial disturbance.

Although a disturbance is seen to move along the surface of the pond, rope, or spring, water is not displaced bodily to outer reaches of the pond from the point at which the stone fell, nor are pieces of the rope or coils of the Slinky displaced from one end to the other. Motions are communicated from one layer of the medium to the next as a "pulse"–a shape or disturbance propagating away from the original disturbance.

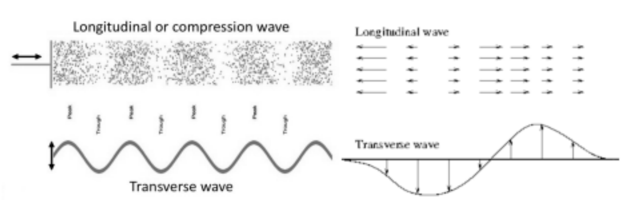
Please view the following video:

`https://www.youtube.com/watch?v=iT4KAc0Ag1E`

As the pulse moves along the medium with a finite velocity, it is apparent that two very different velocities can be distinguished. One velocity is that of the pulse as a whole, referred to as "propagation velocity." The other velocity is that of the particles of the medium, referred to as "particle velocity." The particles of water and of rope move up and down in a direction perpendicular to that in which the pulse moves, or they move back and forth in line with the pulse motion. Within the region of the moving pulse, some layers of particles are moving up while others are moving down or some are moving forward while some are moving backward.

The propagation of a pulse or disturbance in a medium is distinctly different from the bodily displacement of objects studied up to this point. This phenomenon is called "wave motion" or "wave propagation."

4.44 Longitudinal and Transverse Waves



A wave is a motion to and fro, up and down, or from side to side, usually (except for electromagnetic waves such as light) in a fluid or elastic medium, propagated continuously among the medium's constituent particles, but with no average displacement of the particles themselves in the direction of the propagation of the wave. Such repetitive motions are sometimes referred to as "vibrations" or "undulations."

For instance, water waves are (nearly) vertical undulations in the position of water particles. The oscillations in neighboring particles are phased such that a pattern moves across the water surface. Waves in a Slinky are either transverse, in that the motion of the material of the Slinky is perpendicular to the orientation of the Slinky, or they are longitudinal, with material motion back and forth in the direction of the stretched Slinky [see Figure; the arrows in the right hand diagrams indicate instantaneous displacements]. The crest and trough of a wave are the locations of maximum positive and negative displacements.

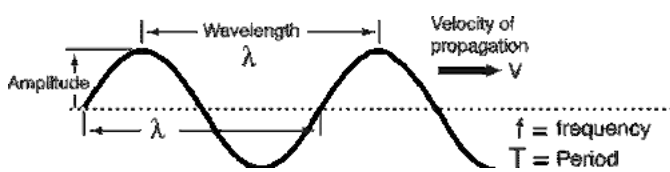
Some media support only longitudinal waves, others support only transverse waves, while yet others support both types. Light waves are purely transverse, while sound waves are purely longitudinal. Water waves are a peculiar mixture of transverse and longitudinal, with particles of water moving in elliptical trajectories as waves pass. Solids transmit both transverse and longitudinal waves.

Light is a form of electromagnetic radiation. The vibrations in an electromagnetic wave occur in the electric and magnetic fields. These oscillations are perpendicular to the direction of motion of the wave (in a vacuum), and are therefore categorized as transverse.

56. Imagine a tube closed at one end and a moveable piston at the other end. Visualize the air contained in the tube as represented by dots distributed uniformly throughout the tube. Suppose that the piston is moved abruptly toward the closed end and then back. Sketch the propagation down the tube of pulses of compression and rarefaction.

57. A long solid rod lies at rest on a table. Suppose the rod is pushed at one end hard enough to displace the rod. Does the far end move instantaneously with the push in the direction of the push? Describe what happens. In what ways is this similar to the situation of the previous problem and of the Slinky. In what ways different?

4.45 Sine wave



The sine wave sketched in the figure has the mathematical form

$$D(x) = A \sin\left(\frac{2\pi\Delta x}{\lambda}\right) \tag{95}$$

where D is the displacement at position x, either longitudinally along, or transversely from, the reference line. A is the maximum displacement or "amplitude" of the wave. $\Delta x = x - x_i$, where x_i is an inital reference position along the reference line. And λ is the wavelength, the distance along the reference line through which the sine function completes one full cycle. The oscillatory behavior of the wave is assumed to carry on to infinity in both positive and negative x directions.

The image and the equation present the sine wave as it appears at a particular instant. Most waves–and essentially all interesting waves–move with time. That is, the disturbance propagates. The movement of the disturbance to the right a distance Δs may be accounted for by replacing Δx by $\Delta x - \Delta s$ in the formula. If this movement occurs in time Δt, then the disturbance moves at velocity $v = \Delta s/\Delta t$. Solving this for Δs and substituting yields a formula for the displacement of a point as a function of both position x and time t:

$$D(x,t) = A \sin\left[\frac{2\pi(\Delta x - v\Delta t)}{\lambda}\right] \tag{96}$$

The time for a wave to move the distance of one wavelength is called the period of the wave: $T \equiv \lambda/v$. Thus, the sine wave may also be written

$$D(x,t) = A \sin\left[2\pi\left(\frac{\Delta x}{\lambda} - \frac{\Delta t}{T}\right)\right] \tag{97}$$

which itself can be rewritten in a simpler, more common form:

$$D(x,t) = A \sin\left(k\Delta x - \omega\Delta t\right) \tag{98}$$

by defining the "wavenumber" $k \equiv 2\pi/\lambda$ and (recall) angular velocity $\omega = 2\pi/T$.

The frequency of oscillatory motion is the number of cycles completed in unit time, $f = 1/T$, which is therefore related to the angular velocity by $\omega = 2\pi f$. The frequency is usually easier to measure than the angular velocity, but the angular velocity tends to be used more often in research papers and books. Both have dimensions of inverse time, but frequency is the number of cycles per unit time, while angular velocity is the angular displacement (in radians) per unit time.

The argument of the sine function is by definition an angle. The angle $\phi \equiv kx - \omega t$ is referred to as the "phase" of the wave. Because $\Delta x = x - x_i$ and $\Delta t = t - t_i$, the argument $k\Delta x - \omega\Delta t$ may be written $(kx - \omega t) - (kx_i - \omega t_i) = \phi - \phi_i$. The second term of this expression, ϕ_i, is referred to as the "initial phase" or "phase shift," which indicates that the sine wave may not be zero at $x = 0$, $t = 0$. Of course, if $x_i = 0$ and $t_i = 0$, $\phi_i = 0$, in which case there is no phase shift.

Perhaps the most important thing to realize in the discussion is that, *the difference in the phase of a wave is 2π at fixed time over a distance of one wavelength and at fixed position over a time interval of one period.*

The "intensity" of a wave, often related to the quantity of energy carried by the wave, is proportional to the square of the amplitude. The wave speed defined above, $v = \lambda/T$, often written $v = \lambda f$, is the propagation (or "phase") velocity. Since $\lambda = 2\pi/k$ and $f = \omega/2\pi$, this propagation (or phase) velocity can also be expressed in terms of the angular velocity and wavenumber:

$$v = \lambda f = \frac{\omega}{k} \tag{99}$$

58. A wave on a string can be represented by the equation $y(x,t) = a\sin(b\Delta x + c\Delta t)$. In terms of a, b, and c, what are the amplitude, wavenumber, angular velocity, frequency, wavelength, and phase (or propagation) velocity of this wave?

59. The speed of sound in air is about 350 m/s. The speed of sound in water is about 1500 m/s. The musical note middle C has a frequency of about 260 cycles/s. If a middle C is sounded in air and in water, in which medium is the wavelength shorter? By what fraction?

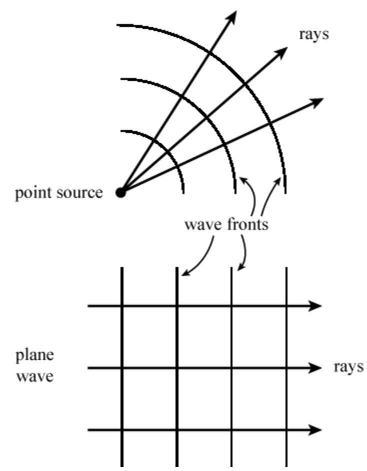
60. Measure a pulse rate (yours or somebody else's). Compute the ordinary frequency in cycles per second. Compute the angular velocity in radians per second. Compute the period.

4.46 Waves in two (or three) dimensions

Depending on the dimension of the medium, a wave disturbance spreads out in one, two, or three dimensions from its source. A quiet water surface touched with a pencil point or fingertip generates a circular wave pulse spreading out from the origin of the disturbance. If the pencil is vibrated up and down, it will generate a continuous train of waves, successive crests and troughs forming concentric circles around the source. In an exactly analogous manner, a vibrating tuning fork sends out a continuous train of sound waves. At distances large relative to the size of the fork itself, the fork appears to be a point source, with successive regions of compression and rarefaction forming concentric spherical shells around it.

A region of constant phase (a crest, a trough, a locus of zero deflection, or a locus of any arbitrarily chosen deflection from $-A$ to $+A$) is called a "wave front." Lines drawn in such a way as to be everywhere normal (perpendicular) to the wave fronts they cross are called rays [see Figure]. For ripples from a point source on a water surface, wave fronts form concentric circles, and rays are radial lines originating at the point source.

Waves may also be generated by the motion of an extended source. A long bar vibrated up and down at a water surface will generate a wave having the same phase along its entire length. At distances relatively close to the bar, the successive wave fronts have the form of parallel straight lines. Such waves are called "plane" waves; their rays are also parallel to each other and do not diverge as do the rays of circular or spherical waves due to a point source [see Figure].



At distances very large relative to the size of the bar, the wave fronts would appear to be circles of very large radius centered at the source. On the other hand, small sections of circular wave fronts at very large radii, far from the source, appear to be very nearly plane and are frequently treated to a very good approximation as plane waves with essentially parallel rays. The basic simplicity of the plane and circular waves resides, in part, in the fact that the rays are straight lines in both cases.

4.47 Refraction

In many circumstances, the velocity of wave propagation may vary from point to point in a medium. The velocity of waves decreases with decreasing depth of water; if a ripple tank has a sloping bottom, the propagation velocity becomes a function of position in the tank. Sound velocity in air or water increases with increasing temperature; if the temperature varies in the horizontal and vertical directions, as it does in both the atmosphere and the ocean, sound velocity becomes a function of position in the medium. The distribution of a quantity as a function of space and/or time is known as a "field."

When a quantity varies in space, it is said to have a "spatial gradient." The gradient in a particular direction is measured by the derivative with respect to position. This amounts to the space rate of change of the quantity in that direction. For example, a temperature gradient in the x-direction at any particular point is measured by the derivative $\frac{dT}{dx}$. Similarly, a gradient in propagation velocity is measured by $\frac{dv}{dx}$,

Any local portion of a wave front propagates forward at the local propagation velocity. Look at

`https://efsantabarbaranewsdotcom.files.wordpress.com/2018/07/rincon.jpg?w=648`

At the top right, the waves move into water of decreasing depth, with wave fronts aligned parallel to loci of constant propagation velocity. Moving into regions of lower velocity, the fronts remain parallel to each other but the wavelengths

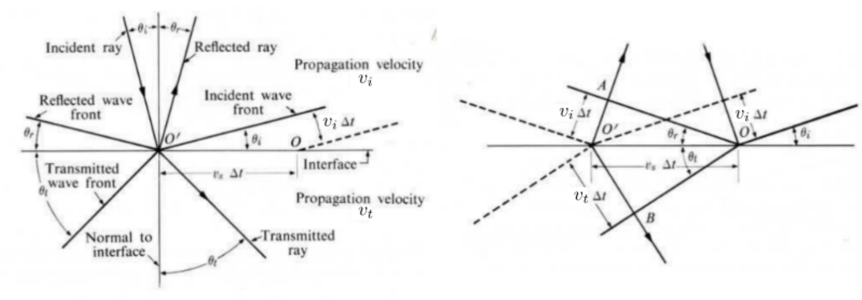
($\lambda = v/f$) decrease, since the frequency is fixed, having been established by the driving source.

At the center and front left of the photograph, the wave fronts are not parallel to regions of constant propagation velocity. As a result the left-hand portions of the waves, in deeper water, move faster than the right-hand portions. The wave fronts tend to bend around; the rays are no longer straight lines emanating from the source, but curve toward the region of lower propagation velocity.

The phenomenon of distortion of wave fronts and bending of rays under the influence of gradients in the propagation velocity is called "refraction." Refraction occurs also when a wave propagates through an interface between two uniform regions having sharply different propagation velocities: when a sound wave in air encounters a water surface or when ripples in a water layer of uniform depth encounters a "shelf" at which the depth changes to another uniform value. In these instances the propagation velocity gradient is zero in each region, but refraction occurs sharply at the boundary, where there is an abrupt change in the propagation velocity. See

`https://cdn2.webdamdb.com/md_U41wzfx9ZD94.png?1536870639`

4.48 Laws of reflection and transmission



The figures are line drawings of wave fronts and rays reflecting and refracting at a line of abrupt change in the media, such as a sudden change in depth in a ripple tank. Arrows on rays indicate direction of propagation of the respective waves. The angle θ_i = angle of incidence, θ_r = angle of reflection, and θ_t = angle of transmission or refraction. If $\theta_i = 0$, the incident ray is normal (perpendicular) to the interface and the wave is said to be at normal incidence. If $\theta_i = 90°$, the wave is said to be at "grazing" incidence. A photograph taken a short time later would show the entire pattern displaced to the right a distance $v_s \Delta t$ [see the right-hand sketch].

Experience suggests

1. θ_r and θ_i are roughly equal.

2. θ_t is greater or less than θ_i, depending on whether v_t is greater or less than v_i.

3. If θ_i is increased, the other two angles increase.

As the incident wave moves forward at velocity v_i, its point of contact $0'$ with the interface moves toward the right at a velocity denoted by v_s [see Figure]. Thus, after a time interval Δt, the incident wave front will have propagated a distance $v_i \Delta t$ and will be in the location shown by the dashed line in the left figure: point $0'$ will have moved to position 0, a distance $v_s \Delta t$. Therefore,

$$\begin{aligned} \sin\theta_i &= \frac{v_i \Delta t}{v_s \Delta t} \\ v_s &= \frac{v_i}{\sin\theta_i} \end{aligned} \tag{100}$$

What does this say?

1. When the incident wave grazes the interface, that is, when $\theta_i = 90°$, $\sin\theta_i = 1$. Therefore, $v_s = v_i$: so the point of contact moves with the propagation velocity of the wave.

2. For $\theta_i < 90°$, $\sin\theta_i < 1$, so $v_s > v_i$. Thus, in general, the point of contact moves along the interface at a velocity higher than that of the wave propagation.

61. Verify that Equation 100 implies that $v_s \to \infty$ as $\theta_i \to 0$. Describe the behavior of the point of contact as the wave (ray) approaches normal incidence. Is there anything disturbing about this result?

The sketch on the right in the previous figure shows the pattern of incident, reflected, and transmitted wave fronts at two successive instants of time. From it,

$$\sin\theta_r = \frac{v_i}{v_s} \tag{101}$$

$$\sin\theta_t = \frac{v_t}{v_s} \tag{102}$$

Substituting Equation 100 into the first of these yields,

$$\begin{aligned} \sin\theta_r &= \frac{v_i}{v_i}\sin\theta_i = \sin\theta_i \\ \theta_r &= \theta_i \end{aligned} \tag{103}$$

This result is the law of reflection. Translated into words, it says that the angle of reflection must equal the angle of incidence. Such reflection *from a smooth surface* is frequently referred to as "specular."

Similarly eliminating v_s from Equation 102,

$$\sin\theta_t = \frac{v_t}{v_i}\sin\theta_i$$
$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{v_t}{v_i} \tag{104}$$

This result is the law of refraction: The ratio of the sine of the transmitted (refracted) ray's angle with the normal to the sine of the incident ray's angle with the normal equals the ratio of the propagation velocity in the two media.

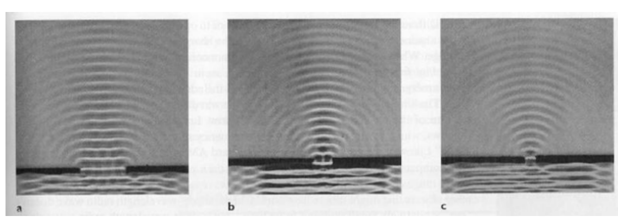
62. How is θ_t related to θ_i when $v_t > v_i$? when $v_t < v_i$ Sketch the orientation of wave fronts and rays for each case. If, in a transmitted wave, you observe the ray to have been bent away from the normal (the line perpendicular to the interface), on which side of the interface will you find the higher propagation velocity?

Frequently, propagation velocity is given as a dimensionless quantity in terms of a maximum possible velocity, usually signified c. A standard convention employs the dimensionless ratios $n_i \equiv c/v_i$ and $n_t \equiv c/v_t$ to characterize the propagation velocity in the two media forming the interface. With this definition, the law of refraction takes the form

$$\frac{\sin\theta_t}{\sin\theta_i} = \frac{v_t}{v_i} = \frac{n_i}{n_t} \tag{105}$$

The quantities n_t and n_i are called "indices of refraction" or "refractive indices" of the respective media. Waves propagate more slowly in media with larger refractive indices.

4.49 Diffraction



The figures illustrate three characteristic patterns formed when a plane wave encounters an obstacle pierced by a single opening. Waves continue to propagate through the opening, but the wave pattern on the other side is determined by the relationship between the wavelength λ of the wave and the width D of the opening. Such bending and modification of a wave front on passing through an opening is called "diffraction." Diffraction effects are small when λ is much smaller than D (Figure a) and become increasingly pronounced as λ becomes of the order of magnitude of D or larger. The pattern illustrated in Figure b is called a single-slit diffraction pattern. When λ is larger than the opening (Figure c) the slit becomes, to a good approximation, a point source of circular waves.

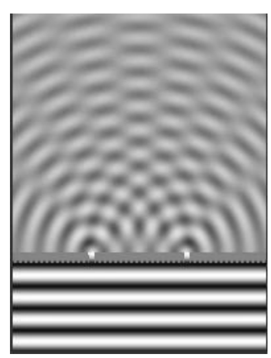
If a wide "beam" of particles, all traveling parallel to each other, were to arrive at a barrier with a single opening of width D, a narrower parallel beam, of width D, would continue on the other side, regardless of the size of D (as long as it is larger than the diameter of the particles), with perhaps a few particles deflected (or scattered) out of their original path by the edges of the slit. A wave with $\frac{\lambda}{D} \ll 1$ (Figure a) behaves in a somewhat analogous fashion, but the pronounced diffraction effects of Figures b and c, when $\frac{\lambda}{D} \geq 1$, are uniquely characteristic of wave behavior, markedly different from the behavior of a beam of particles.

4.50 Superposition Principle

It is found empirically that, in most media, as long as the amplitudes of two waves are small, the waves do not interact with on another. For example, two waves moving in the opposite direction pass through each other without their shapes or amplitudes being changed. Wherever parts of the waves are at the same position, the total wave displacement is the sum of the displacements of the individual waves. This is called the "superposition principle." At sufficiently large amplitude the superposition principle can break down: interacting large-amplitidue waves may scatter off one another, lose amplitude, or change their form.

4.51 Interference

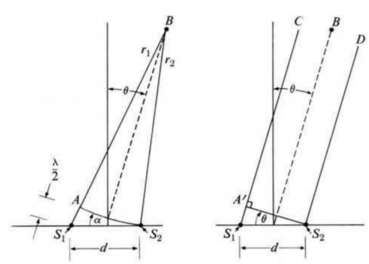
Interference is a consequence of the superposition principle. When two or more waves are superimposed, the net wave displacement is the algebraic sum of the displacements of the individual waves. Since these displacements can be positive or negative, the net displacement can either be greater or less than the individual wave displacements. When both displacements are of the same sign, the waves interfere constructively, called "constructive interference," and the resulting displacement is greater than that of either wave individuallly. If the displacements are of opposite sign, then the waves interfere destructively, called "destructive interference." and the resulting displacement is smaller than that of either wave individually. Pairs of waves are said to be "in-phase" if they interfere constructively and "out-of-phase" if they interfere destructively.



63. Suppose that two identical sinusoidal wave trains are traveling in exactly opposite directions. Sketch the waves in several suc-

cessive phases relative to each other and plot the superposition pattern. Note that the resultant pattern changes with time but does not move along the direction of propagation of the component waves. (This resultant pattern is called a standing wave. Standing wave patterns play an essential role in the description of the vibration of strings and air columns in musical instruments.)

Interference effects are barely visible in the images of single-slit diffraction [review figures on page 80], but are readily identified in double-slit diffraction [see Figure]. The two slits in the barrier struck by plane waves have openings $D \ll \lambda$, the wave length of the waves. Each slit therefore acts as a point source for circular ripples. The ripples from each slit oscillate in phase. The geometry of the pattern changes when either the wavelength of the incoming wave or the distance between the slits is changed. The lines along which the resultant wave amplitude appears to be zero are called "nodal" lines. Here the waves from the two sources overlap systematically crest-to-trough and interfere destructively. In the sectors between the nodal lines, the waves superpose crest-to-crest and trough-to-trough; they constructively interfere.



The central region of constructive interference, a line starting half way between the slits, is called the "principal axis." Along it, the waves from the two sources arrive in phase, having traveled equal distances from the sources [see Figure]. The first locus of destructive interference to the right of the axis is one along which the waves are everywhere exactly out of phase: the wave from the left slit has traveled a distance r_1, which exceeds r_2 by exactly one-half wavelength. In short, the locus is defined by the relation $r_1 - r_2 = \lambda/2$. The other loci are defined in a similar way: if the difference in distance traveled from the two sources is an odd number of half wavelengths, a nodal line forms; if the difference in distance traveled from the two sources is a multiple of the wavelength, the waves interfere constructively.

In the left figure, point B lies relatively far out along the first nodal line, the locus of which is, again $r_1 - r_2 = \lambda/2$. Imagine a circle whose center is point B and whose radius is r_2. Then consider an arc of that circle from the right slit S_2 to the point A on line r_1, that is, $\overset{\frown}{S_2A}$. Because B is on a nodal line, and the distance from A to B is the same as the distance from S_2 to B, the length $\overline{S_1A}$ equals $\lambda/2$ and, since $\overset{\frown}{S_2A}$ is very nearly a straight line, α can be thought of as an angle whose sine (side opposite over hypotenuse) is $\lambda/2d$.

In the right figure, point B still lies on first nodal line, but instead of lines r_1 and r_2 radiating from B, lines $\overline{S_1C}$ and $\overline{S_2D}$ are drawn parallel to the nodal line. Line S_2A' is then drawn perpendicular to $\overline{S_1C}$, and the angle $A'S_2S_1$ is called

θ. When point B is relatively far from the origin, the shape AS_2S_1 in the left figure is very nearly identical to triangle $A'S_2S_1$ in the right figure. $\overline{S_1A'}$ is then very nearly equal to $\lambda/2$ and θ is very nearly equal to α. Thus, $\sin\theta = \lambda/2d$, to an extremely close approximation.

This being the first nodal line from the principal axis, the angle can be labeled θ_1. Other nodal lines will occur at larger angles, for which the distance $\overline{S_1A'}$ will equal three half-wavelengths, then five half-wavelengths, and so forth. Summarizing, for the nth nodal line at angle θ_n,

$$\sin\theta_n = (2n+1)\frac{\lambda}{2d}, \tag{106}$$

with $n = 0,\ \pm1,\ \pm2,\ \pm3,\ldots$. The expression $(2n+1)$ gives successive odd numbers when the integers n are substituted.

Similarly, constructive interference occurs along lines oriented in such a way that the path difference $\overline{S_1A'}$ equals one, two, three, or any other whole number of wavelengths. Thus, the loci of constructive interference lie at angles θ_n for which,

$$\sin\theta_n = n\frac{\lambda}{d}, \tag{107}$$

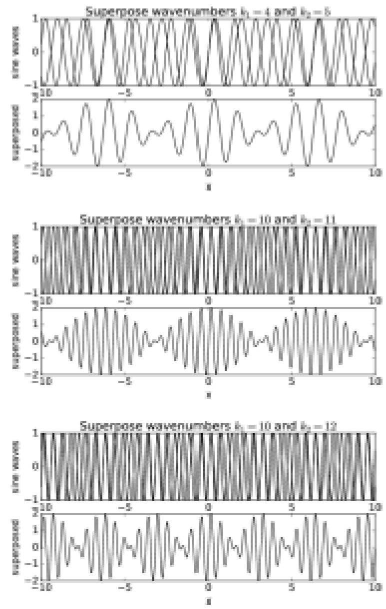
where, again, $n = 0,\ \pm1,\ \pm2,\ \pm3,\ldots$.

Thus, if λ and d are known, the angular positions of the various constructive and destructive loci can be immediately predicted. If λ is not known, but the source spacing and the angle of, say, the second nodal line, θ_2 can be measured, then the wavelength of the plane waves impinging on the barrier can be calculated.

64. For any particular source spacing d and wavelength λ, the number n ceases to have physical significance after exceeding a certain value. Why? For any given pattern, what is the highest meaningful value of n? What happens when λ becomes larger than d? Larger than $2d$? (Note that if $\lambda \gg 2d$, the two sources cannot be distinguished from a single source emitting circular waves into the medium.)

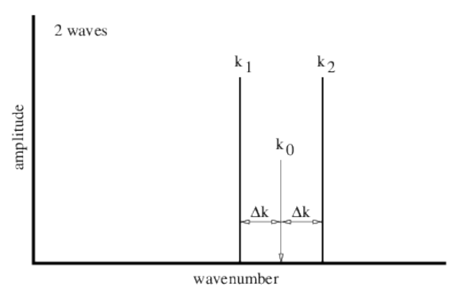
65. Suppose that the two point sources randomly shifted their phase relative to each other. Would an interference pattern emerge? Suppose that the sources were steady but the phase relation between the sources is changed, so that one slit emits a trough at precisely the moment the other emits a crest, and vice verse (i.e., the phase lag would be equivalent to $\lambda/2$). What would happen to the interference pattern?

4.52 Wave packets



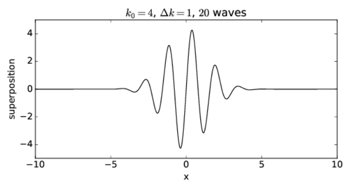
What happens when two sine waves with different wavenumbers ($k = 2\pi/\lambda$) superpose [see Figures]? The top image shows the superposition of two waves with wavenumbers $k_1 = 4$ and $k_2 = 5$, that is, with wavelengths $\lambda_1 = \pi/2$ and $\lambda_2 = 2\pi/5$. Notice that the result is a wave with about the same wavelength as the two initial waves, but varying in amplitude depending on whether the two sine waves are interfering constructively or destructively.

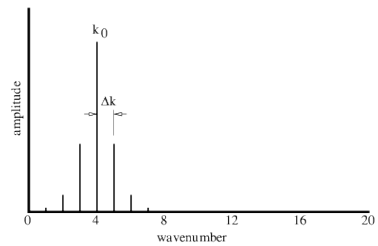
The middle image shows the result when $k_1 = 10$ and $k_2 = 11$, that is $\lambda_1 = \pi/5$ and $\lambda_2 = 2\pi/11$. Notice that the wavelength of the resultant wave is decreased, but the locations where the amplitude is maximum have the same separation in x as in the top figure.



If waves with $k_1 = 10$ and $k_2 = 12$, as shown in bottom image, are superposed, the x-spacing of the regions of maximum amplitude has decreased by a factor of two. Thus, while the wavenumber of the resultant wave seems to be related to something like the average of the wavenumbers of the component waves, the spacing between regions of maximum wave amplitude appears to go inversely with the difference of the wavenumbers. In other words, if k_1 and k_2 are close together, the amplitude maxima are far apart, while the amplitude maxima come closer together when the wavenumbers separate.

The sine waves that make up the three images can be represented symbolically by a plot such as that shown in the figure. The amplitudes and wavenumbers of each of the sine waves are indicated by vertical lines.





The regions of large wave amplitude are called "wave packets." The wave packets produced by just two sine waves follow one after the other. However, superimposing many waves can produce an isolated wave packet. For example, the left figure shows the results of superimposing 20 sine waves with wavenumbers $k = 0.4n$, $n = 1, 2 \ldots, 20$, where the amplitudes of the waves are largest for wavenumbers near $k = 4$ {because the amplitude of each sine wave is set to be proportional to $\exp[-(k - k_0)^2/(\Delta k)^2]$, where $k_0 = 4$ defines the maximum of the distribution of wavenumbers and $\Delta k = 1$ defines the "half-width" (the width at half the maximum) of this distribution}. The distribution of amplitudes of each of the sine waves making up the wave packet is shown schematically in the right figure.

The quantity Δk controls the distribution of the sine waves being superimposed: only those waves with a wavenumber k within approximately Δk of the central wavenumber k_0 of the wave packet ($3 \leq k \leq 5$ in this case) contribute significantly to the sum. If Δk is changed to 2, so that wavenumbers in the range $2 \leq k \leq 6$ contribute significantly, the wavepacket becomes narrower. Δk is called the "wavenumber spread" of the wave packet, and it evidently plays a role similar to the difference in wavenumbers in the superposition of two sine waves: the larger the wavenumber spread, the smaller the physical size of the wave packet. Furthermore, the wavenumber of the oscillations within the wave packet is given approximately by the central wavenumber, k_0.

To understand how wave packets work mathematically, consider again the case of the superposition of two sine waves. Assign $k_0 = (k_1 + k_2)/2$ and $\Delta k = (k_2 - k_1)/2$, where k_1 and k_2 are the wavenumbers of the component waves [refer, again, to the figure on page 84]. Adding and subtracting these two equations yields expressions for k_1 and k_2 in terms of k_0 and Δk, $k_1 = k_0 - \Delta k$ and $k_2 = k_0 + \Delta k$, which then may be substituted into the sum of two (time independent) sine waves:

$$\begin{aligned}\sin(k_1x) + \sin(k_2x) &= \sin[k_0x - (\Delta k)x] + \sin[k_0x + (\Delta k)x] \\ &= \sin(k_0x)\cos[(\Delta k)x] - \cos(k_0x)\sin[(\Delta k)x] \\ &+ \sin(k_0x)\cos[(\Delta k)x] + \cos(k_0x)\sin[(\Delta k)x] \\ &= 2\sin(k_0x)\cos[(\Delta k)x] \end{aligned} \tag{108}$$

where the trigonometric identity $\sin(a \pm b) = \sin a \cos b \pm \cos a \sin b$ was used.

The $\sin(k_0x)$ factor produces the oscillations within the wave packet. This oscillation has a wavenumber k_0, equal to the average of the wavenumbers of the component waves. The $\cos[(\Delta k)x]$ factor modulates the wave such that the regions of maximum amplitude are spaced at

$$\Delta x = \frac{\pi}{\Delta k} \tag{109}$$

As observed previously, the length of the wave packet, Δx is inversely related to the spread of the wavenumbers Δk (which, in this example, is the difference between the two wavenumbers) of the component waves. This relationship underlies the uncertainty principle of quantum mechanics, as will be seen.

66. **Sketch the resultant wave obtained from superimposing the waves $y_1 = \sin(2x)$ and $y_2 = \sin(3x)$.**

67. **Two sine waves with wavelengths λ_1 and λ_2 are superimposed, making wave packets of length L. To increase L, should λ_1 and λ_2 be closer together or farther apart? Explain.**

4.53 Beats

Suppose two sound waves of different frequency but equal amplitude impinge on an ear at the same time. Assume the position of the ear is fixed in space, so that the waves can be taken as independent of position. The ear perceives (that is, the eardrum is displaced by) the superposition of the two waves, with time dependence only:

$$y(t) = \sin(\omega_1 t) + \sin(\omega_2 t) = 2\sin(\omega_0 t)\cos[(\Delta\omega)t] \tag{110}$$

where the same trigonometric identity has been used, and where $\omega_0 = (\omega_1 + \omega_2)/2$ and $\Delta\omega = (\omega_2 - \omega_1)/2$. What is actually heard is a tone with angular velocity ω_0, which fades in and out with period

$$T_{\text{beat}} = \frac{\pi}{|\Delta\omega|} = \frac{2\pi}{|\omega_2 - \omega_1|} = \frac{1}{|f_2 - f_1|}. \tag{111}$$

The beat frequency is simply

$$f_{\text{beat}} = \frac{1}{T_{\text{beat}}} = |f_2 - f_1|. \tag{112}$$

Note how beats are the time analogue of wave packets: the mathematics are the same except that frequency replaces wavenumber and time replaces space.

4.54 Group velocity: how fast do wave packets move?

As will be shown, wave packets do not move at the same rate as the waves that form them.

Superpose two sine waves, one with wavenumber k_1 and angular velocity ω_1, the other with wavenumber k_2 and angular velocity ω_2 superpose:

$$\begin{aligned} y(x,\ t) &= \sin(k_1 x - \omega_1 t) + \sin(k_2 x - \omega_2 t) \\ &= 2\sin(k_0 x - \omega_0 t)\cos[(\Delta k)x - (\Delta\omega)t], \end{aligned} \tag{113}$$

where, again, $k_0 = (k_1 + k_2)/2$, $\omega_0 = (\omega_1 + \omega_2)/2$, $\Delta k = (k_2 - k_1)/2$, and $\Delta\omega = (\omega_2 - \omega_1)/2$.

A wave packet, then, is described in (mathematical) words as "a sine wave of wavenumber k_0 and angular velocity ω_0 modulated by a cosine wave of wave number Δk and angular velocity $\Delta\omega$." As the wave packet holds together, the modulation pattern moves such that the argument of the cosine function remains constant:

$$(\Delta k)x - (\Delta\omega)t = \text{constant} \Rightarrow x = \frac{\Delta\omega}{\Delta k}t + \frac{\text{constant}}{\Delta k} \tag{114}$$

Δk and $\Delta\omega$ themselves do not change with time, so, in particular, both $\frac{\Delta\omega}{\Delta k}$ and $\frac{\text{constant}}{\Delta k}$ are constant.

The velocity of the modulation ("group velocity") is, then,

$$u \equiv \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

or

$$u \equiv \frac{dx}{dt} = \frac{\Delta\omega}{\Delta k} \tag{115}$$

4.55 Electromagnetic radiation: photons

From the point of view of quantum physics, light and other forms of electromagnetic radiation consist of particle-like, discrete bundles of energy[18] called "photons" or "quanta."

Photons have energies, which depend only on the frequency of the radiation, f:

[18] It's surprisingly difficult to say what energy is. Solid, liquid, and gaseous things seem to have it, as do fields (electrical, magnetic, gravitational), but it has no single, indentifiable property. To be sure, it is a scalar quantity and a substance-like quantity that can neither be created nor destroyed. Although an object has much energy when it is hot, moves or rotates fast, or is under high pressure, the amount of energy is not simply proportional to temperature, velocity, or pressure. Energy seems to assume many different forms, all of which are ultimately different manifestations of either kinetic or potential energy. Energy associated with motion is called "kinetic energy." Energy associated with the relative position of two objects is called "potential energy."

$$E = hf = \frac{hc}{\lambda} \tag{116}$$

since the velocity of a wave, $v = \lambda f$, and $v = c$ for light. Further, h is "Planck's constant," and has dimensions Mass · Length^2 · Time^{-1} [$\mathrm{ML^2T^{-1}}$]. This is the first appearance of the dimension Mass.[19]

Special relativity stipulates that photons, because they travel at light speed, have zero mass. A photon's energy, therefore, is purely kinetic. If a photon exists, it moves at light speed; if it ceases to move a light speed, it ceases to exist.

A beam of electromagnetic radiation is, in this view, a collection of photons traveling at light speed. The intensity (or power per unit area) of such a beam is proportional to the number of photons crossing a unit area per unit time. Power is the amount of energy transferred or converted per unit time. Putting this together, then, if the beam has only one frequency (is monochromatic), then

$$I = (\text{energy of one photon}) \times \frac{\text{number of photons}}{\text{area} \times \text{time interval}} \tag{117}$$

A moving object has, besides kinetic energy, momentum, $\vec{p}$, which is a vector quantity having (Euclidean) components, $\vec{p} = [p_x,\ p_y,\ p_z]$, such that the magnitude of the momentum is $|\vec{p}| = \sqrt{p_x^2 + p_y^2 + p_z^2}$, exactly similar to, for example, velocity, to which momentum is related.

Special relativity combines the scalar energy, E, and the vector momentum, $\vec{p}$, into a "four-vector," called the "four-momentum," $\boldsymbol{P} = [\frac{E}{c},\ p_x,\ p_y,\ p_z]$. Special relativity also shows that the magnitude squared of a four-momentum is a constant:

$$|\boldsymbol{P}|^2 \equiv M^2c^2 = \left(\frac{E}{c}\right)^2 - |\vec{p}|^2. \tag{118}$$

Notice that the squared magnitude of a four-vector is the *difference* between the square of the scalar and the squared magnitude of the (three-)vector. The quantity M is known as the invariant mass, which happens to be the mass of an object which has energy E and momentum $\vec{p}$.

Since photons are massless, $M = 0$, so

$$|\vec{p}| = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda}, \tag{119}$$

since $E = hf = hc/\lambda$, and $c = \lambda f$.

68. What are the wavelength, frequency, and magnitude of momentum of a photon whose energy is E, in terms of E?

[19] Perhaps surprising, since photons have no mass.

69. If the maximum wavelength of a photon with enough energy to disassociate a diatomic molecule is λ_{max}, what is the binding energy of the molecule? ["Binding energy" is the minimum energy required to remove a particle from a system of particles or to disassemble a system of particles into individual parts.]

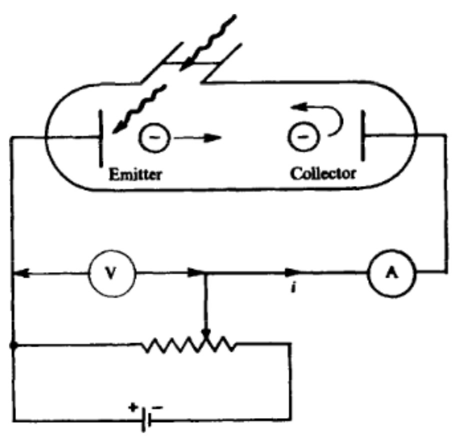
70. An electron has mass m_e and energy E_e.

(a) What is the magnitude of the electron's momentum, $\vec{p}_e$?

(b) What energy does a photon have if the magnitude of its momentum equals that of the electron?

71. Monochromatic light of wavelength λ is incident normally on (perpendicular to) a surface of area A. If the intensity of the light is I, what is the rate (photons/unit time) at which photons strike the surface?

4.56 Photoelectric effect



In what is known as a photoelectric experiment, light is shined on a metal surface housed inside an evacuated tube [see Figure]. The light can cause electrons to be emitted from the surface of the metal. In the experiment, the frequency f and intensity I of the light, the retarding voltage V, and the kind of metal can be varied. Sufficiently energetic electrons can overcome the retarding voltage V and reach the collector to be recorded as current $i = I_Q$ in the ammeter A. That is, if the kinetic energy K is greater than or equal to the electron charge e times the retarding voltage V ($K \geq eV$), then a current i will be indicated in the ammeter A. If their kinetic energy K is less than the product eV, ($K < eV$), then the electrons will be turned back before reaching the collector and will register no current in the ammeter.

The results of the experiment are:

1. The current begins almost instantaneously, even for light of very low intensity (the delay between the light striking the surface and current being registered is of order 10^{-9} s, independent of the intensity).

2. When the frequency and retarding potential are fixed, the current is directly proportional to the intensity of the incident light.

3. When the frequency and intensity are fixed, the current decreases as the retarding voltage is increased, reaching zero at a certain "stopping voltage," V_s, independent of the intensity.

4. For a given metal and surface condition, the stopping voltage varies linearly with the frequency of the light:

$$eV_s = hf - \Phi \tag{120}$$

   where e is the charge of the electron, V_s is the stopping voltage, h is Planck's constant, f is the frequency of the light, and Φ is the "work function," characteristic of the metal and the condition of the surface. Note that this is a linear relationship, with the independent variable f, the dependent variable eV_s, slope h, and intercept $-\Phi$.

5. For a given metal and surface condition, there exists a "threshold frequency," f_{th}, below which no electrons will be emitted, no matter the intensity.

Of these results, only (2) is explicable in terms of a wave model of light: increased intensity means increased energy transmission, so more electrons should be emitted. On the other hand, interpreting light as photons or quanta explains all the results. In this model, a single electron absorbs the energy of a single photon. If this energy is sufficient to overcome the energy with which the electron is bound to the surface, then the electron will be emitted with kinetic energy equaling the difference between the absorbed energy and the binding energy. Electrons are bound to the surface with a variety of energies, but the binding energy of the least tightly bound electrons depends only on the metal and the condition of the surface. The energy required to remove these least-tightly-bound electrons is the work function, Φ, of the surface of the metal. Electrons will therefore be emitted with varied kinetic energies ranging from zero to a maximum value given by:

> Maximum kinetic energy of emitted electron = (energy carried by photon) − (binding energy of the least tightly bound electron)

which explains result (3). Since $K_{\max} = eV_s$, the maximum-energy relation becomes

$$eV_s = hf - \Phi \tag{121}$$

explaining results (4) and (5), where the threshold frequency is that which gives the photon an energy equal to the work function:

$$hf_{\text{th}} = \Phi \tag{122}$$

Below this threshold energy (frequency), the incident photon energy is insufficient to release even the least tightly bound electrons, regardless of the light intensity.

Photon absorption occurs almost instantaneously, explaining the short delay time [experimental result (1)]. Finally, result (2) is explained by the relationship between light intensity and photon density: the greater the intensity the larger the number of photons per unit area, and the more electrons are emitted.

72. The kinetic energies of emitted electrons range from 0 to K_{max} when light of wavelength λ falls on a surface. Answer the following questions using only the information given by making appropriate substitutions into and manipulating Equations 121 and 122.

   (a) What is the stopping potential V_s for this light?

   (b) What are the threshold frequency f_{th} and threshold wavelength λ_{th} for the surface?

73. The threshold wavelength of the emitting surface in a photoelectric tube is λ_{th}. What is the wavelength of the incident light if its stopping potential is V_s? Again, use only the information given.

4.57 Matter waves

It has just been demonstrated that, in certain circumstances (for example, the photoelectric effect), electromagnetic radiation exhibits the characteristics of matter, while in other circumstances (for example, two-slit interference and diffraction) it exhibits wave characteristics. It seems that electromagnetic radiation exhibits what has been referred to as "wave-particle duality."

Because they both have energy and momentum, it's worth emphasizing the distinction between waves and objects (particles). A classical object has position, momentum, kinetic energy, mass, and perhaps electric charge. A classical wave has a wavelength, a frequency, phase and group velocities, an amplitude, an intensity, energy, and momentum. Perhaps the primary distinction between these is that objects can be localized whereas waves spread out and can occupy large portions of space.

In 1924, Louis de Broglie proposed that an object could, under certain circumstances, act like a wave. In particular, de Broglie proposed that an object passing through a slit of width comparable to the wavelength associated with the object would undergo diffraction.

What wavelength is associated with an object? For a photon, $f = E/h$ and $\lambda = h/|\vec{p}|$, where f is the frequency, E is the energy, λ is the wavelength, p is the momentum, and h is Planck's constant. Notice that the left sides of these relationships involve wave characteristics, while the right sides involve characteristics that objects, too, possess. Planck's constant connects the sides. Arguing from the symmetry of nature, de Broglie conjectured that frequencies and wavelengths associated with objects satisfy the same relationships as photons. In particular, the frequency f and wavelength λ of a wave associated with an object would be

$$f = \frac{E}{h} \tag{123}$$

$$\lambda = \frac{h}{|\vec{p}|} \tag{124}$$

where, again, E is the object's total energy, $\vec{p} = m\vec{v}$ is its momentum (a vector in classical physics), m is the object's mass and $\vec{v}$ its velocity. The *group* velocity of the associated wave equals the magnitude of the object's velocity. Note, this implies that the wave is not a simple sine wave, but a wave packet, localized to the position of the object. In the case of photons and other massless objects, which move at light speed, these relationships are related by $c = \lambda f$ and $|\vec{p}| = E/c$, but the motion of massive objects and their associated group velocities must be less than c, and the relationships for frequency and wavelength are independent.

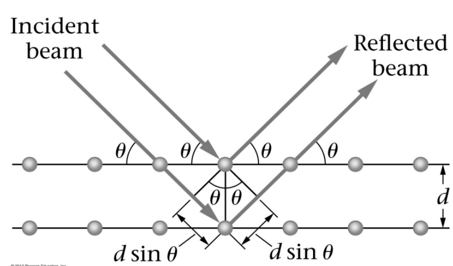
Take a look at these short videos:

`https://www.youtube.com/watch?v=MbLzh1Y9POQ&list=PLwpAKFo3GUVud-sc-NnqldKkhlbYoE35v&index=7`

`https://www.youtube.com/watch?v=ZqS8Jjkk1HI`

4.58 Bragg's law

When X-rays–a form of electromagnetic radiation with wavelengths comparable to the distance between atoms in a solid–interact with a crystal, they diffract, a process known as Bragg diffraction. The scattering they undergo from atoms in a regular crystalline (lattice) structure results in an interference pattern.



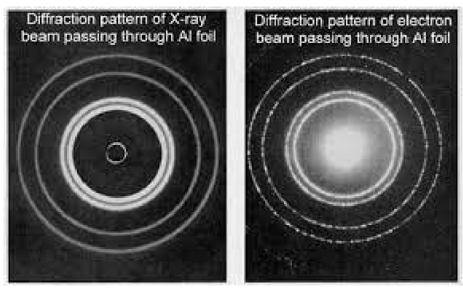
Electrons also have wave-like properties and can undergo Bragg diffraction by crystals. Bragg diffraction thus connects wave and particle phenomena.

The figure is a sketch of interference between waves scattering from two adjacent rows of atoms in a crystal. The net effect of scattering from a single row is equivalent to partial reflection from a mirror: the angle of "reflection" equals the angle of incidence for each row. Interference then occurs between the beams reflecting off different rows of atoms in the crystal.

For the two adjacent rows shown in the figure, the path difference between beams is twice the length of the side opposite to the angle θ. Remember that the sine of an angle is the ratio of the side opposite the angle to the hypotenuse. Here, the hypotenuse is d, so, because the path difference results from reaching and then reflecting from the next layer, the total path difference is $2d \sin\theta$. For constructive interference, this must be an integer number of wavelengths, $m\lambda$, where the integer m is called the "order of interference." The result is Bragg's law of diffraction:

$$m\lambda = 2d \sin\theta_m \tag{125}$$

where $m = 1,\ 2,\ 3, \ldots.$ If only two rows are involved, the transition from constructive to destructive interference as θ changes is gradual. However, if interference from many rows occurs, then the constructive interference peaks become very sharp with mostly destructive interference in between. The image on the left shows Bragg diffraction in a crystal with X-rays, and the image on the right shows Bragg diffraction in the same crystal with electrons.



74. (a) **In order for a wave to scatter, its wavelength must be roughly the same size or smaller than the size of the object being observed. If the linear dimension of an object (say, its diameter) is d, what is the minimum energy of a photon that would make the object "visible?"**

(b) **An object's kinetic energy and momentum are related non-relativistically by $p = \sqrt{2mK}$, where p is the object's momentum, m its mass, and K its kinetic energy. All electrons have the same mass, m_e. What is the minimum kinetic energy of an electron whose associated wave will scatter from an object of linear dimension d?**

4.59 The Meaning of Quantum Wave Functions

Bragg diffraction illustrates perhaps the most basic peculiarity of the peculiar subject of quantum mechanics: objects can have wave-like properties and waves can have object-like properties. The variation of X-ray intensity with angle seen in a Bragg diffraction apparatus is very difficult to explain in any terms other than wave interference. Yet, X-rays are typically detected by a device such as a Geiger counter, which produces a pulse of electricity for each photon, which hits it. Thus, X-rays sometimes act like matter and sometimes like waves. Light isn't alone in having both matter and wave properties. As demonstrated by their undergoing Bragg diffraction in crystals, electrons also can act like waves. much the way X-rays do.

But what exactly is "waving" when an electron (or other massive object) exhibits wave properties? The answer to this question is given by quantum mechanics. Now, many people, include many physicists (even Albert Einstein), have found quantum mechanics to be bizarre. Though it may be strange, it is useful for solving problems.

Consider again two-slit interference (and recall the two videos showing the resulting pattern for light and electrons). The emergent pattern depends on the superposition of waves emanating from both slits. What happens if the slit each object passes through is identified, which should be possible if light (and electrons) have both matter and wave properties? Such identification is indeed possible. Experiment demonstrates, though, that the very act of making this measurement–identifying the slit each object passes through–alters the form of the associated wave.

According to the wave view of electromagnetic radiation, the intensity I (energy per unit area per unit time) at a point is proportional to the electric field squared, $\mathcal{E}^2$, where $\mathcal{E}$ is the value of the electric field at the point. According to the photon view of electromagnetic radiation, the intensity at the point is proportional to the number of photons per unit area per unit time, N. Because N and $\mathcal{E}^2$ are both proportional to the intensity, they are proportional to one another, $N \propto \mathcal{E}^2$, where the symbol $\propto$ means "proportional to."

It is not possible to predict where an individual photon will strike the recording screen, but, because the final pattern consists of alternating bright and dark bands, individual photons have a high probability of arriving at a bright band and zero probability of arriving at a dark band. N, then, at a point on the screen is a measure of finding a photon in the vicinity of that point. But, since $N \propto \mathcal{E}^2$, the square of the electric field gives the probability of finding a photon near the point, and, in the wave interpretation of electromagnetic radiation, the electric field is what oscillates. The square of the function describing these oscillations at a point gives the probability of finding a photon in the vicinity of this point.

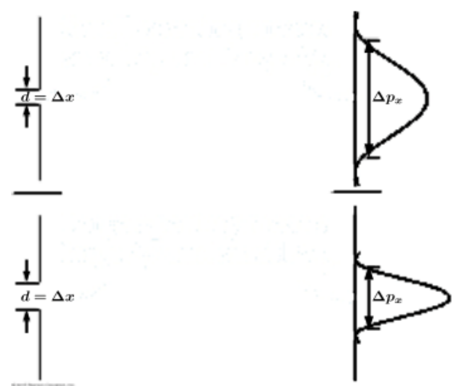
Matter waves produce similar interference patterns. The quantity oscillating with a de Broglie wavelength $\lambda = \frac{h}{|\vec{p}|}$ is the "wave function," commonly denoted $\psi(x)$, whose absolute square, $|\psi(x)|^2$ gives the probability of finding the massive object in the vicinity of the point in question. According to quantum mechanics, the exact position of an object is impossible to specify. Rather, only the *probability* of finding the object at some position at a particular time can be given.

Now it is possible to understand what happens to the interference pattern when the slit each object passes through is identified. Since the absolute square of the wave function $|\psi|^2$ represents the probability of locating the object, once a object has been found–in this case, to have passed through one slit or the other–the wave function "collapses" into a very small wave packet located at the observed position of the object. Thus, the wave function becomes zero at the slit not passed through. Because an interference pattern results from the superposition of waves emanating from both slits, if no wave comes from one of the slits (because the wave displacement is zero there), then there is no interference pattern.

To turn the argument around: If there is an interference pattern, then the wave function is non-zero at both slits. Because the absolute square of the wave function is interpreted as a probability, the existence of an interference pattern means it is not possible, even in principle, to know through which slit the object passed if it can pass through either. No experiment can identify the slit without destroying the interference pattern.

4.60 The Heisenberg Uncertainty Principle

How might the position of an electron be determined? One way might be to place a gap in a barrier along the suspected path of an electron of known energy. If the electron is recorded on a screen beyond the gap, then it will be known that the electron passed through the gap. The position, at least in one direction (call it x–in any case, perpendicular to the original direction), is known to within the size of the gap, $\Delta x = d$ [see Figure]. The smaller the gap, the smaller Δx, and hence the more precisely the position is known.



Again, if the electron is recorded on the screen beyond the gap, then the electron will be known to have passed through the gap. But the wave associated with the object will be diffracted on passing through the gap, the more so the smaller the gap [refer again to the figures on page 80]. It is therefore not possible to predict where on the screen an electron will strike. Diffraction implies that the electron's velocity, and therefore its momentum, is changed (recall that velocity and momentum are vectors, so "change" could refer to magnitude or direction or both).

Prior to passing through the gap, the electron's position was unknown, but its momentum was known, both the magnitude (because the energy was known) and the direction (perpendicular to the barrier). On passing through the gap, the electron's position is localized, but its momentum may have obtained an x-component (which had been zero), as it moves to some arbitrary point within the diffraction pattern. Because the point at which the object strikes the screen can not be known, the x-component of the electron's momentum must become uncertain, Δp_x. Increasing the size of the gap will minimize Δp_x, but wll increase Δx. Uncertainties in a object's position component and corresponding momentum component cannot both be made arbitrarily small: precision in one of these quantities can be increased only at the expense of precision in the other.

All such experiments lead to the same conclusion, codified in 1927 by W. Heisenberg:

$$\Delta p_x \Delta x \geq \frac{h}{4\pi} \tag{126}$$

The "Heisenberg uncertainty principle" can also be formulated in terms of other conjugate or complementary variables[20]. Another example is energy and time: In order to measure the energy E of an object, the measurement must be made over a certain time interval Δt. The uncertainty in the energy, ΔE, is related to the time interval Δt during which the energy is measured by

$$\Delta E \Delta t \geq \frac{h}{4\pi} \tag{127}$$

That is, the energy can be known with perfect precision ($\Delta E = 0$) only if measured over an infinite time interval ($\Delta t = \infty$).

The principle has physical consequences for systems,[21] such as excited atoms and resonance states, that transform in a finite amount of time, that is, with a mean lifetime τ. These systems have a natural minimum energy uncertainty given by $\Delta E = h/(4\pi\tau)$.

The uncertainty principles show that a single experiment cannot simultaneously measure both variables of conjugate or complementary pairs (such as p_x and x, and E and t) to arbitrary precision. A fundamental consequence of such principles is that matter and wave aspects cannot be measured in the same experiment. An experiment intending to measure matter properties must minimize position and time uncertainties, since an object can, by definition, be localized in position and time. But then the wave properties, wavelength and frequency (related to momentum and energy, $\lambda = h/p$ and $f = E/h$) will be essentially unknown. Displaying matter properties suppresses wave properties, and vice versa. This is codified as the "principle of complementarity" (N. Bohr, 1928): while both wave and matter properties must be measured to completely understand matter, they cannot be observed simultaneously.

75. Suppose the uncertainty in the momentum of an object equals the particle's momentum, $\Delta p_x = p_x$. How is the minimum uncertainty in the object's position, Δx, related to the de Broglie wavelength, $\lambda_{deB} = h/p$?

76. If the uncertainty in the energy of a nuclear state (and therefore its mass) is ΔE , what is the state's average lifetime, τ?

[20] Conjugate or complementary variables are pairs of variables about which precise knowledge of one implies little knowledge of the other.

[21] Recall: A system in physics is a specified collection of objects. The rest of the universe is considered the environment.

4.61 The point of quantum mechanics

Quantum mechanics was invented to understand the very, very small, where, it seems, ordinary–and even classical physics–insights fail. Although metaphorical images are often used to try to describe quantum behaviors, these, too, fail. These physical phenomena can only be modeled as mathematical objects. Fortunately, recipes, to be presented and illustrated here, are available to work through the equations relating these mathematical objects.

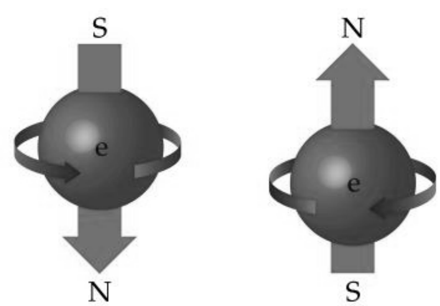
A generic name for these objects is "quanta," or "quantum" in the singular. Photons and electrons are examples of quanta. Each species of quanta has certain intrinsic characteristics, such as mass, charge, and something called "spin" which distinguish one species from another. Electrons, for example, all have the same mass, charge, and spin, while no photon has mass or charge, but all have the same spin that differs from that of the electron. Individual quanta can be in different "states," in which the values of "observables," such as position, energy, or (spin) angular momentum components, can be measured in experiments.

Two basic types of observables characterize a quantum moving in one dimension: spin observables (related to spin angular momentum) and spatial observables (related to position, momentum and energy).

Quantum mechanics has been developed to predict values of experimental observables when an ensemble of quanta are prepared in a certain way.

4.62 Spin

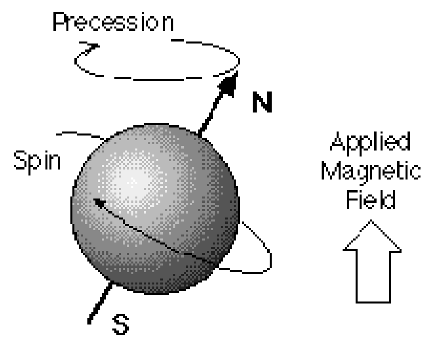
An electron behaves as if it were a spinning charged ball,[22] and thus acts like an electromagnet due to its circulating current. Because the electron's charge is negative, the poles are aligned opposite to the direction of the spin (using a right-hand rule) [see Figure].



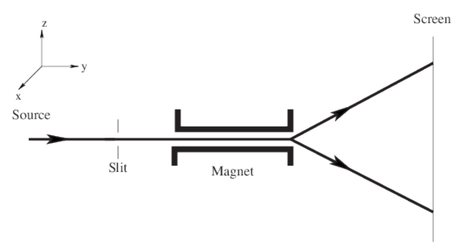
Magnets have two poles, designated north and south. The north pole of a magnet is the one attracted by the North Pole of the Earth (which, because opposite poles attract and like poles repel, is therefore a magnetic south pole). The strength of this attraction falls off rapidly with distance. In general, the pole of a large magnet will attract the opposite pole of a smaller magnet passing by and repel the like pole, but the effect will be larger on the nearer pole than on the farther pole. The strength of the effect (either attraction or repulsion) is proportional to the cosine of the angle between the

[22]It's not a ball. As far as anyone knows, it is a mathematical point, smaller than 10^{-18} m.

north-to-south orientation of the small magnet and the direction of the magnetic field of the large magnet (the external field). If the effect is attraction, the path of the small magnet will be diverted in the direction of the pole of the large magnet; if the effect is repulsion, the small magnet will be diverted away from the large magnet's pole.



Because it is spinning, an electron (a small magnet) will not simply line up with the large magnet's field, but, like a top spinning on the ground, will instead precess (that is, the orientation of its rotational axis will orbit in a circle around the direction of the external field [see Figure]) in such a way that the projection of the spin angular velocity (recall the right-hand rule for the direction of angular velocity) onto the direction of the external field (that is, the magnitude of the spin vector times the cosine of the angle between the north-to-south orientation of the electron and the direction of the external field) remains constant. Thus, the small magnet will not align with the field, but will precess around the direction of the external field, preserving the projection of its spin vector onto the direction of the external field, as its trajectory along the magnet is diverted either toward or away from the pole.

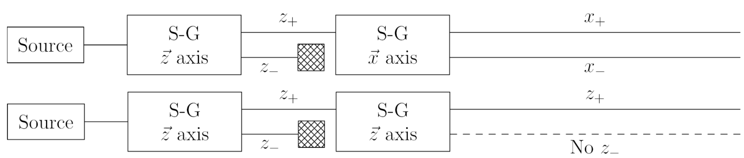


This sort of experiment is called a "Stern-Gerlach" experiment, after the scientists who discovered the intrinsic spin of quanta. The interesting thing is that the diverted trajectories were expected to create a broad smear of interactions on the screen, but what happened was that the deflections split into two, implying that the projection of the electron spin vector takes on only two values, assigned $+\frac{1}{2}$ (known as "spin-up," or aligned with the external field) and $-\frac{1}{2}$ (known as "spin-down," or anti-aligned with the external field). Further, reorienting the large magnet, so that the external field lies along another axis, changes nothing: the projection of the electron's spin vector on any axis yields just two values, $+\frac{1}{2}$ (spin-up) and $-\frac{1}{2}$ (spin-down). From a random (unpolarized) beam of electrons, about half will have the positive value and half will have the negative value, but it's impossible to predict which any given electron will have before making the measurement.

It is possible, then, to create a model of a Stern-Gerlach experiment as a box (a SG box) that takes in electrons and sends them out positive or negative outputs. If an electron's spin orientation is not known at the input, then there will be a 50-50 chance of it being either positive or negative. If the orientation of an electron relative to one axis is known, then its probabilities of being spin-up and spin-down relative to another axis rotated by angle θ relative to the first axis are:

$$P(+) = \cos^2(\frac{\theta}{2}) \tag{128}$$

$$P(-) = \sin^2(\frac{\theta}{2}) \tag{129}$$

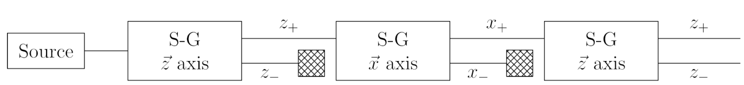


Note that if the second axis is perpendicular to the first, $\theta = 90°$, so $\frac{\theta}{2} = 45°$, and $\cos 45° = \sin 45° = \frac{1}{\sqrt{(2}}$, so $P(+) = P(-) = \frac{1}{2}$ [see top Figure]. Note further that if an output of a SG box runs through an identical SG box, the probability of getting the same result is 100% (this is called self-consistency) [see bottom Figure].

77. Use Equations 128 and 129 to show that the model is self-consistent, for both spin-up and spin-down electrons, realizing that the angle between the external field and the spin-up orientation is $\theta = 0$ and that between the external field and the spin-down orientation is $\theta = 180°$.

It is possible, then, to model any series of these sorts of experiments with a small number of boxes:

1. Aligned with the external field, which is conventionally designated the z-axis, SG_z.
2. At an angle θ with respect to the external field, SG_θ.
3. (Not strictly necessary, but convenient) Perpendicular to the external field, say, SG_x (the x- and y- axes are both perpendicular to the z-axis, so they are interchangeable, and x substitutes as well for y).



This being quantum mechanics, though, the output of a series of these boxes depends on whether or not the output of an intermediate box is known. For example, if a SG_z box selects only those electrons whose spin vector projection along the external field is aligned with the field (that is, that are spin-up with respect to the z-axis) and sends them through a SG_x box which selects only those electrons whose spin vector projection is aligned with the x-axis, then a subsequent SG_z will find that half of the electrons are spin-up and half spin-down, as if the initial selection of spin-up electrons never happened (although the number of electrons of each orientation will have been reduced by a factor of 4 rather than a factor of 2 relative to the initial beam) [see Figure]. However, if spin-up and spin-down electrons from the SG_x box are superposed (recombined, so that the intermediate state remains undetermined) and then passed through the last SG_z box, only spin-up electrons would emerge, just as they were selected by the first SG_z box, as if the SG_x box weren't there. This is exactly the case in two-slit interference: determining–or not–an intermediate state affects the final result.

78. **If a SG_x box selects spin-up electrons and sends them through a SG_z box, which selects spin-up electrons and sends them through a SG_θ box, what are the SG_θ output probabilities for spin-up and spin down electrons?**

Mathematical notes: Although only real wave functions are used in this discussion, quantum mechanics more generally requires wave functions to be complex, that is, to have real and imaginary parts, similar to a complex number. The complex number z is the sum of a real number a and an imaginary number ib. An imaginary number is a real number multiplied by $i \equiv \sqrt{-1}$, so both a and b are real numbers. Thus, $z = a + ib$ for any complex z, where a is the real part of z and b is the imaginary part, sometimes written $a = \Re(z)$ and $b = \Im(z)$. The "complex conjugate of z is $z^* = a - ib$. That is, the complex conjugate of a complex number transforms all is into $-i$s. The product of a complex number and its conjugate gives the magnitude squared of the complex number: $zz^* = z^*z = a^2 + b^2$.

Quantum mechanical wave functions are represented by vectors whose length depends on the number of possible values of the observables, and whose components are in general complex (although, again, in this discussion, they will be real). A quantum mechanical vector can be written as a column vector or a row vector. The row vector is the complex conjugate of a column vector. Short-hand notation, known as bra-ket notation, represents the vectors, with the ket being the column and the bra being the row:

$$|\psi\rangle \equiv \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \end{bmatrix} \tag{130}$$

$$\langle\psi| \equiv \begin{bmatrix} \psi_1^*, & \psi_2^*, & \ldots \end{bmatrix} \tag{131}$$

The product of a row and column vector is known as the "inner product":

$$\langle\varphi|\psi\rangle \equiv \begin{bmatrix}\varphi_1^*, & \varphi_2^*, & \ldots\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\\ \vdots\end{bmatrix} = \varphi_1^*\psi_1 + \varphi_2^*\psi_2 + \cdots \tag{132}$$

(analogous to what, in linear algebra, is called a "dot" or "scalar" product.). The inner product is not generally commutative, $\langle\varphi|\psi\rangle \neq \langle\psi|\varphi\rangle$, except when all components of both vectors are real (as they will be here).

Notice that the inner product of a vector with itself gives the magnitude squared of the vector, as with complex numbers. A quantum mechanical vector is "normalized" if $\langle\psi|\psi\rangle = 1$ (its magnitude squared is 1). Two quantum mechanical vectors are orthogonal (perpendicular in n-space) if $\langle\varphi|\psi\rangle = 0$.

4.63 Quantum mechanics: a basic how-to

Recall that quantum mechanics has been developed to predict values of experimental observables when an ensemble of quanta are prepared in a certain way. Here is a sort of recipe, using spin variables, for how this works:

1. Preparing a quantum puts it into a well-defined quantum state represented by the normalized vector $|\psi\rangle$ ($\langle\psi|\psi\rangle = 1$).

   As discussed, an electron has only two possible spin values, $+\frac{1}{2}$ and $-\frac{1}{2}$, so the (spin) state vector has two components:

   $$|\psi\rangle = \begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix} \tag{133}$$

   Normalization means $\langle\psi|\psi\rangle = |\psi|^2 = \begin{bmatrix}\psi_1*, & \psi_2*\end{bmatrix}\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix} = |\psi_1|^2 + |\psi_2|^2 = 1$.

   A quantum state vector doesn't specify the values of, for example, a quantum's spin projections, as would be the case in classical physics, but rather provides probabilities for possible outcomes of a measurement of an observable.

2. The possible numerical values of a given observable are known as eigenvalues. Associated with each eigenvalue is an eigenvector, each of which is normalized and orthogonal to every other eigenvector. That is, each measurement yields a unique value for each observable.

   A measurement of the projection of an electron's spin onto the z-axis, S_z, (or onto any other axis) yields only two possible results (eigenvalues), $+\frac{1}{2}$ and $-\frac{1}{2}$. The associated eigenvectors are usually labeled $|+z\rangle$ and $|-z\rangle$ (or $|+x\rangle$ and $|-x\rangle$ for a measurement of S_x, etc.).

   The following table shows the conventional eigenvectors for an electron's spin projection observables:

Eigenvalue \ Observable	S_z	S_x	S_θ
$+\frac{1}{2}$	$\lvert +z\rangle = \begin{bmatrix}1\\0\end{bmatrix}$	$\lvert +x\rangle = \begin{bmatrix}\sqrt{\frac{1}{2}}\\\sqrt{\frac{1}{2}}\end{bmatrix}$	$\lvert +\theta\rangle = \begin{bmatrix}\cos\frac{\theta}{2}\\\sin\frac{\theta}{2}\end{bmatrix}$
$-\frac{1}{2}$	$\lvert -z\rangle = \begin{bmatrix}0\\1\end{bmatrix}$	$\lvert -x\rangle = \begin{bmatrix}\sqrt{\frac{1}{2}}\\-\sqrt{\frac{1}{2}}\end{bmatrix}$	$\lvert -\theta\rangle = \begin{bmatrix}-\sin\frac{\theta}{2}\\\cos\frac{\theta}{2}\end{bmatrix}$

79. Show that the conventional eigenvectors for an electron's spin projection observables are normalized and orthogonal. That is, show that $\langle +z|+z\rangle = 1$, $\langle -z|-z\rangle = 1$, $\langle +x|+x\rangle = 1$, etc., and that $\langle +z|-z\rangle = 0$, etc.

3. Measuring the value of an observable sets (prepares) the quantum's state function to be the eigenvector associated with whatever value is measured. For example, an electron leaving a SG_z box with a measured spin projection of $+\frac{1}{2}$ is in a $|+z\rangle$ (eigen)state, while one leaving a SG_z box with a measured spin projection of $-\frac{1}{2}$ is in a $|-z\rangle$ (eigen)state. This effect of measurement is frequently referred to as "collapsing the wavefunction."

4. Consider some observable, call it A, of a quantum prepared in state $|\psi\rangle$. Say there are N different measurable values of observable A, so there will be N different eigenvectors associated with N different eigenvalues, $|a_1\rangle$, $|a_2\rangle$, $\ldots$, $|a_k\rangle$, $\ldots|a_N\rangle$. The "amplitude" for measuring, say, the eigenvalue associated with $|a_k\rangle$ is given by the inner product $\langle a_k|\psi\rangle$ [note that the outcome of the measurement is the bra (on the left), while the initial state is the ket (on the right)]. The probability of measuring this value is the absolute square of the amplitude $|\langle a_k|\psi\rangle|^2$.

   For example, say an electron is prepared to be in a spin-up state $|+z\rangle$ by selecting $+\frac{1}{2}$ spin projections from a SG_z box. It is then sent through a SG_θ box. The amplitude for the electron to emerge from the SG_θ box in a $|+\theta\rangle$ eigenstate is

$$\langle +\theta|+z\rangle = \begin{bmatrix}\cos\frac{\theta}{2}, & \sin\frac{\theta}{2}\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = (\cos\frac{\theta}{2})(1) + (\sin\frac{\theta}{2})(0) = \cos\frac{\theta}{2} \quad (134)$$

   so the probability of the electron emerging from the SG_θ box in a $|+\theta\rangle$ eigenstate is

$$|\langle +\theta|+z\rangle|^2 = \cos^2\frac{\theta}{2} \quad (135)$$

80. Show that the probability of an electron prepared in a spin-up $|+z\rangle$ state emerging from the SG_θ box in a $|-\theta\rangle$ eigenstate is $\sin^2\frac{\theta}{2}$.

81. An electron emerging spin-down from a SG_x box, $|-x\rangle$, passes through a SG_θ box. In terms of θ, what is the probability that it emerges spin-down from the SG_θ box, $|-\theta\rangle$?

5. Probability theory holds that the probability of independent events occurring in sequence is the product of each of the event probabilities[23]. In quantum mechanics, the amplitude for the outcome of successive independent measurements is the product of the individual amplitudes, and the probability is the magnitude squared of this product.

   Consider a quantum prepared in state $|\psi\rangle$. The amplitude for measuring the value a_k of observable A followed by a measurement of b_k of the observable B is $\langle b_k|a_k\rangle\,\langle a_k|\psi\rangle$, where $|a_k\rangle$ and $|b_k\rangle$ are the eigenvectors associated with the eigenvalues a_k and b_k. Note that the amplitude is written so that the evolution of events is read right to left: the state vector $|\psi\rangle$ is prepared, then the observable A is measured, and then the observable B is measured.

   Suppose a beam of electrons is prepared by sending it through a SG_z box and sorting spin-up and spin-down, and the resulting wavefunction is:

$$|\psi\rangle = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$$

82. Show that $|\psi\rangle$ is normalized.

If the spin-up electrons selected by the SG_z box are subsequently sent through a SG_θ box, the amplitude for electrons to emerge spin-up from this new orientation is

$$\begin{aligned}\langle +\theta| + z\rangle\,\langle +z|\psi\rangle &= \left(\begin{bmatrix}\cos\frac{\theta}{2}, & \sin\frac{\theta}{2}\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right)\left(\begin{bmatrix}1, & 0\end{bmatrix}\begin{bmatrix}\frac{3}{5}\\ \frac{4}{5}\end{bmatrix}\right)\\ &= \left(\cos\frac{\theta}{2}+0\right)\left(\frac{3}{5}+0\right) = \frac{3}{5}\cos\frac{\theta}{2}\end{aligned}$$

so the probability that an electron from the prepared beam emerges from a SG_z box spin-up and then from a SG_θ box also spin-up is $\frac{9}{25}\cos^2\frac{\theta}{2}$.

83. (a) What is the probability that an electron from the prepared beam emerges from a SG_z box spin-up and then from a SG_θ box spin down?

(b) What is the probability that an electron from the prepared beam emerges from a SG_z box spin-down and then from a SG_θ box spin-up?

[23] The probability of two consecutive heads in successive flips of a fair coin is $(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$.

(c) **What is the probability that an electron from the prepared beam emerges from a SG_z box spin-down and then from a SG_θ box also spin down?**

(d) **Since this covers all possible outcomes, what should the sum of the probabilities be? Is that the case here?**

84. Let

$$|\psi\rangle = \begin{bmatrix} -3a \\ 4a \end{bmatrix}$$

(a) **What must be the value of a for $|\psi\rangle$ to be normalized?**

(b) **If $|\psi\rangle$ represents a spin state, what is the probability that a measurement of S_z yields $+\frac{1}{2}$ (spin-up)?**

85. Suppose a SG box prepares electrons in a definite but unknown spin state $|\psi\rangle = \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}$. Sent through a SG_z box, the probability that an electron emerges spin-up $|+z\rangle$ is 1/10 while the probability that it emerges spin-down $|-z\rangle$ is 9/10. Sent through a SG_x box, the probability that an electron emerges spin-up $|+x\rangle$ is 1/5 while the probability that it emerges spin-down $|-x\rangle$ is 4/5. Assuming they are real, what are the values of the components of $|\psi\rangle$? [The overall sign of a state vector cannot be determined (the magnitude squared is positive definite regardless), so there are two, equivalent answers.]

6. If a quantum can take a number of different paths from preparation to final measurement and it is not possible, even in principle, to know which path has been taken, then the total amplitude is the *sum* of the amplitudes for each path.

Recall a previous discussion on page 99: the outcome of an experiment depends on whether or not an intermediate state is known. In the example given there, spin-up electrons from a SG_z box are sent through a SG_x box. Either spin-up electrons from this second box are sent into another SG_z box, or both spin-up and spin-down electrons from this second box are superposed (remixed) before entering another SG_z box. The output of the second SG_z box differs in the two cases. If the spin-up output of the SG_x are selected, and therefore intermediate information is gleaned, then the output of the second SG_z box would be 50-50 spin-up and spin-down, as if the initial selection never happened, except that only half the original beam would come out.

$$
\begin{aligned}
\langle +z | +x \rangle \langle +x | +z \rangle &= \left(\begin{bmatrix} 1, & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} \end{bmatrix} \right) \left(\begin{bmatrix} \sqrt{\frac{1}{2}}, & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\
&= \left(\sqrt{\frac{1}{2}} + 0 \right) \left(\sqrt{\frac{1}{2}} + 0 \right) = \frac{1}{2} \\
\langle -z | +x \rangle \langle +x | +z \rangle &= \left(\begin{bmatrix} 0, & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} \end{bmatrix} \right) \left(\begin{bmatrix} \sqrt{\frac{1}{2}}, & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\
&= \left(0 + \sqrt{\frac{1}{2}} \right) \left(\sqrt{\frac{1}{2}} + 0 \right) = \frac{1}{2}
\end{aligned}
$$

But if the outputs of SG_x are superposed, so that no S_x information is known, then the sequence through $|-x\rangle$ has to be added to these results.

$$
\begin{aligned}
\langle +z | -x \rangle \langle -x | +z \rangle &= \left(\begin{bmatrix} 1, & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{2}} \end{bmatrix} \right) \left(\begin{bmatrix} \sqrt{\frac{1}{2}}, & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\
&= \left(\sqrt{\frac{1}{2}} + 0 \right) \left(\sqrt{\frac{1}{2}} + 0 \right) = \frac{1}{2} \\
\langle -z | -x \rangle \langle -x | +z \rangle &= \left(\begin{bmatrix} 0, & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \\ -\sqrt{\frac{1}{2}} \end{bmatrix} \right) \left(\begin{bmatrix} \sqrt{\frac{1}{2}}, & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\
&= \left(0 - \sqrt{\frac{1}{2}} \right) \left(\sqrt{\frac{1}{2}} + 0 \right) = -\frac{1}{2}
\end{aligned}
$$

leading to

$$
\begin{aligned}
\langle +z | +x \rangle \langle +x | +z \rangle + \langle +z | -x \rangle \langle -x | +z \rangle &= \frac{1}{2} + \frac{1}{2} = 1 \\
\langle -z | +x \rangle \langle +x | +z \rangle + \langle -z | -x \rangle \langle -x | +z \rangle &= \frac{1}{2} - \frac{1}{2} = 0
\end{aligned}
$$

demonstrating that, if spin-up and spin-down electrons from the SG_x box are superposed (recombined, so that the intermediate state remains undetermined) and then passed through the last SG_z box, only spin-up electrons would emerge, just as they were selected by the first SG_z box, as if the SG_x box weren't there.

86. A beam of electrons is prepared in the $|+z\rangle$ state by a SG_z box and sent through a SG_θ box. The outputs of the SG_θ box are superposed (remixed) and sent through a second SG_z box. What is the amplitude for electrons to emerge spin-down $|-z\rangle$ from the second SG_z box?

87. A beam of electrons is prepared in the $|+z\rangle$ state by a SG_z box and sent through a SG_θ box. The outputs of the SG_θ box are superposed (remixed) and sent through a SG_x box. What is the probability that an electron emerges spin-up $|+x\rangle$ from the SG_x box? Spin-down $|-x\rangle$?

7. The final step in the recipe is the time evolution of the state if the measurement of an observable is delayed for some interval after preparation. Time dependence of the state is indicated by $|\psi(t)\rangle$, where it's assumed that the clock starts, $t = 0$, at preparation.

The laws of physics are time translation invariant, which means that they involve Δt, not any specific t; the designation of $t = 0$ is an arbitrary choice. This invariance of time translation in physics laws is related to energy conservation by a theorem proved by Emily Nörther. This close association between time and energy manifests in this final step of the recipe.

Three ingredients, then, are necessary:

(1) Write the just prepared (initial) state as a sum of energy eigenstates[24]

$$|\psi(0)\rangle = b_1\,|E_1\rangle + b_2\,|E_2\rangle + \cdots$$

where the coefficients b_n assume values that properly represent the preparation.[25]

(2) Calculate the time evolved state as two additive series, one of cosines, the other of sines:

$$|\psi(t)\rangle_C = b_1 \cos\left(\frac{2\pi E_1 t}{h}\right)|E_1\rangle + b_2 \cos\left(\frac{2\pi E_2 t}{h}\right)|E_2\rangle + \cdots \quad (136)$$

$$|\psi(t)\rangle_S = b_1 \sin\left(\frac{2\pi E_1 t}{h}\right)|E_1\rangle + b_2 \sin\left(\frac{2\pi E_2 t}{h}\right)|E_2\rangle + \cdots \quad (137)$$

where the coefficients b_1, b_2, …, are the same in all cases.[26]

[24] An energy eigenstate (or energy eigenvector or stationary state) is one in which all observables are independent of time. A quantum in an energy eigentstate remains unchanged with time. A state which does vary with time can be written as a linear combination of all possible energy eigenstates available to that system.

[25] If $|\psi(0)\rangle$ is in an energy eigenstate $|E_k\rangle$, then $|\psi(0)\rangle = |E_k\rangle$, and the only coefficient is $b_k = 1$.

[26] For $|\psi(0)\rangle = |E_k\rangle$, $|\psi(t)\rangle_C = \cos\left(\frac{2\pi E_k t}{h}\right)|E_k\rangle$ and $|\psi(t)\rangle_S = \sin\left(\frac{2\pi E_k t}{h}\right)|E_k\rangle$.

(3) Calculate the probability for a certain outcome:[27]

$$|\langle a_k|\psi(t)\rangle|^2 = |\langle a_k|\psi(t)\rangle_C|^2 + |\langle a_k|\psi(t)\rangle_S|^2 \tag{138}$$

that is, the sum of the magnitudes squared of the inner product of the measured eigenvector with the cosine and sine series.[28]

Say an electron has been prepared at $t = 0$ in a $|+x\rangle$ state. Furthermore, assume that, if the electron has energy E, it would emerge spin-up, $|+z\rangle$, from a SG_z box. Thus, assume $+z$ is an energy state, $|E\rangle = |+z\rangle$. Again, if the electron has energy 0, it would emerge from a SG_z box spin-down, $|-z\rangle$, and so $-z$ would also be an energy state $|0\rangle = |-z\rangle$. With these assumptions,

$$|\psi(0)\rangle = |+x\rangle = \begin{bmatrix}\sqrt{\frac{1}{2}}\\ \sqrt{\frac{1}{2}}\end{bmatrix} = \sqrt{\frac{1}{2}}\begin{bmatrix}1\\0\end{bmatrix} + \sqrt{\frac{1}{2}}\begin{bmatrix}0\\1\end{bmatrix} = \sqrt{\frac{1}{2}}\,|+z\rangle + \sqrt{\frac{1}{2}}\,|-z\rangle$$

Then

$$\begin{aligned}
|\psi(t)\rangle_C &= \sqrt{\frac{1}{2}}\cos\left(\frac{2\pi Et}{h}\right)|+z\rangle + \sqrt{\frac{1}{2}}\cos(0)\,|-z\rangle \\
&= \sqrt{\frac{1}{2}}\cos\left(\frac{2\pi Et}{h}\right)\begin{bmatrix}1\\0\end{bmatrix} + \sqrt{\frac{1}{2}}(1)\begin{bmatrix}0\\1\end{bmatrix} \\
&= \sqrt{\frac{1}{2}}\begin{bmatrix}\cos\left(\frac{2\pi Et}{h}\right)\\ 1\end{bmatrix} \\
|\psi(t)\rangle_S &= \sqrt{\frac{1}{2}}\sin\left(\frac{2\pi Et}{h}\right)|+z\rangle + \sqrt{\frac{1}{2}}\sin(0)\,|-z\rangle \\
&= \sqrt{\frac{1}{2}}\sin\left(\frac{2\pi Et}{h}\right)\begin{bmatrix}1\\0\end{bmatrix} + \sqrt{\frac{1}{2}}(0)\begin{bmatrix}0\\1\end{bmatrix} \\
&= \sqrt{\frac{1}{2}}\begin{bmatrix}\sin\left(\frac{2\pi Et}{h}\right)\\ 0\end{bmatrix}
\end{aligned}$$

Therefore, the probability of finding the electron in a $+x$ state at some time t after its preparation in a x state is

[27] This piece is far from obvious, but would be clearer if complex functions were used: $e^{i\theta} = \cos\theta + i\sin\theta$, so $|e^{i\theta}|^2 = e^{i\theta}e^{-}i\theta = e^0 = 1 = (\cos\theta + i\sin\theta)(\cos\theta - i\sin\theta) = \cos^2\theta + \sin^2\theta$, since $(i)(-i) = (\sqrt{-1})(-\sqrt{-1}) = -(-1) = 1$.

[28] For $|\psi(0)\rangle = |E_k\rangle$, $|\langle a_j|\psi(t)\rangle|^2 = \left[\langle a_j|E_k\rangle\cos\left(\frac{2\pi E_k t}{h}\right)\right]^2 + \left[\langle a_j|E_k\rangle\sin\left(\frac{2\pi E_k t}{h}\right)\right] = [\langle a_j|E_k\rangle]^2$, which is constant for any energy E_k and for any observable eigenvalue a_j. When a quantum is in an energy eigenstate, the probability of a certain measurement is constant in time.

$$\langle +x|\psi(t)\rangle_C = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \cos\left(\frac{2\pi E t}{h}\right) \\ \sqrt{\frac{1}{2}} \end{bmatrix} = \frac{1}{2}\cos\left(\frac{2\pi E t}{h}\right) + \frac{1}{2}$$

$$\langle +x|\psi(t)\rangle_S = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} \sin\left(\frac{2\pi E t}{h}\right) \\ 0 \end{bmatrix} = \frac{1}{2}\sin\left(\frac{2\pi E t}{h}\right)$$

$$\begin{aligned} \Rightarrow |\langle +x|\psi(t)\rangle|^2 &= \frac{1}{4}\left[\cos\left(\frac{2\pi E t}{h}\right) + 1\right]^2 + \frac{1}{4}\sin^2\left(\frac{2\pi E t}{h}\right) \\ &= \frac{1}{4}\left[\cos^2\left(\frac{2\pi E t}{h}\right) + 2\cos\left(\frac{2\pi E t}{h}\right) + 1\right] + \frac{1}{4}\sin^2\left(\frac{2\pi E t}{h}\right) \\ &= \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi E t}{h}\right) \end{aligned}$$

The probability oscillates sinusoidally between 1 and 0.

88. Suppose the spin states $|+x\rangle$ and $|-x\rangle$ are energy eigenstates with energies E and 0. A beam of electrons is prepared in a spin-up state by a SG_z box:

$$|\psi(0)\rangle = |+z\rangle = b_1\,|+x\rangle + b_2\,|-x\rangle$$

(a) What are the values of the coefficients b_1 and b_2?

(b) What, then, are $|\psi(t)\rangle_C$ and $|\psi(t)\rangle_S$?