CHAPTER 1

The Problem of Calculus

INTRODUCTION

As a mathematical theory of wide scope and great power, calculus is remarkable for the unity, simplicity, and intelligibility of its elements. The heart of it can be regarded as a direct answer to a natural question, yet from that answer springs the modern world of exact science. Such ramification from a single stem does not occur without roots that extend throughout our experience of nature, geometry, and number. To expose some of them is the purpose of this first chapter, in which a variety of physical and mathematical perplexities will be shown to issue in a common problem. The solution to this problem will be developed in the next three chapters, and the last four will be devoted to extensions and applications of the fundamental theory.

1.1 MASS AND LINE DENSITY

1.1.1 Mass of a segment and line density at a point. Imagine a straight wire of unvarying cross-sectional area. I wish to consider two quantities associated with places on the wire, one of them pertaining to segments of the wire and the other to points along it. Let me say at the outset that in this example and the following ones my intention is to propose lines of thought rather than to give adequate accounts of phenomena. Some of the objections that can be raised to the treatment here will be dealt with in subsequent chapters.

CALCULUS - The Elements

The first quantity is **mass**, which belongs to each segment of the wire. It will be measured in kilograms (kg).[1] Mass is that property of a body by virtue of which it has weight and inertia, and in familiar speech we often treat the kilogram as a measure of weight. Since weight, unlike mass, can vary with location—we weigh less on the moon—this use of the kilogram is valid only locally, where weight keeps a fixed proportion to mass.

If the wire is **uniform**, in the sense of being everywhere the same in composition, the mass of a segment is proportional to its length. In general this will not be the case; for example, one segment of a wire of copper and silver may contain a higher proportion of silver than another segment of the same length and consequently outweigh it.

The second quantity is **line density**, which is associated with each point of the wire. (By "point" I mean a location along the wire, not a geometrical point on or within it.) Line density at a point is measured in kilograms per meter (kg/m), the meter being the measure of length, but does not in general refer to whole meters of the wire; rather, it describes the state of affairs at the single point in question.[2] We might say, of a copper and silver wire: at point A, where the wire is mostly copper, the line density is 0.08 kg/m; from that value it rises continuously to 0.09 kg/m at point B, half a meter away, where the wire is mostly silver. This requires an explanation.

To begin with, in the case of a uniform wire every meter of it is just like every other, so the number of kilograms per meter of wire is unambiguously defined. The line density at any point is said to be that number of kilograms per meter. Thus we know the meaning of line density for the uniform wire, and understand that it is the same at all points.

Now let P be a point of an arbitrary wire; what meaning can be attached to "the line density at P"? A cross-section taken at P reveals the composition of the wire at that point—the ratio of copper to silver, say. Imagine another wire, a uniform one, which has *everywhere* the same composition as the given wire has *at* P. As a uniform wire it has a definite line density (kg/m); this line density we take to be the line density at P. You see that the idea is to describe the state of affairs at P by imagining it extended along a whole wire. To summarize,

---

[1] See the note at the end of §1.1.1.

[2] Line density is to be distinguished from "density" without qualification, which ordinarily signifies volume density (kilograms per cubic meter). Silver is denser than copper: it has more mass, volume for volume.

> the line density of a wire at a point is the line density (kg/m) of the *uniform* wire which *matches* the given wire at the given point.

Or, in other words: the line density of a wire at a point is the line density (kg/m) the wire would have if it were everywhere the same in composition as at the point.

The notion of line density is of doubtful meaning where composition changes abruptly, as at the boundary between a copper and a silver segment. In order to avoid this difficulty, and analogous difficulties in the succeeding examples, we can assume that all quantities vary smoothly.

Here and elsewhere I have freely assumed that *exact numerical values can be assigned to properties of a physical object*, even though in some cases, at least, this is clearly an idealization (as it is for line density at a geometrical cross-section through atomic matter). By this assumption the “physical object” is rendered mathematical to suit our convenience in the present study.

1.1.2 Relation between mass and line density. Total and specific quantity.

Having defined the two quantities, let us consider the relation between them. For a uniform wire it is easily stated. If M is the mass of a segment of length L, and λ is the line density, we have

$$M = \lambda L$$

—the number of kilograms in a segment is the number of kilograms per meter times the number of meters in the segment. Equivalently,

$$\lambda = \frac{M}{L}$$

—the number of kilograms per meter is the number of kilograms in a segment divided by the number of meters in the segment. Mass is the *product* of line density and length, line density the *quotient* of mass by length, or *ratio* of mass to length. The elementary mathematical concepts of product and quotient suffice to express either of the quantities in terms of the other, when the wire is uniform.

What if it is not uniform? Plainly, mass and line density cannot be as simply related as before. Yet there can be little doubt that they still mutually determine one another. Line density indicates the *specific concentration* of matter at each *point*, the quantitative aspect of the wire by which mass is accumulated over distance. Mass represents the *totality* of matter in each *interval*, however small, and so establishes the distribution of matter along the wire of which line density is the ultimate measure. If either is known everywhere, the other must be determined; but by what mathematical relations?

We have arrived at a first formulation of the central problem of calculus: to express the relation between what may be called a **total** quantity (mass) and a corresponding **specific** quantity (line density), by means of suitable generalizations of the notions of product and quotient. The importance of this task will become clearer as we proceed with the examination of a series of diverse situations in which the same problem presents itself again and again to the eye that knows how to look. The ability to recognize its essential features in a given set of circumstances is the greater part of understanding calculus, and effort directed to grasping what is common to the different examples will be amply repaid.

1.1.3 QUESTIONS

Q1. (a) A uniform wire 3 m long has a mass of 9 kg. What is its line density at any point?[3] (b) The line density of a uniform wire 8 m long is 4 kg/m. What is its mass? What is the mass of a 2 m segment?

Q2. A length of round copper and silver wire, uniform in diameter, is pure copper at its midpoint, but the proportion of copper falls off steadily in both directions until pure silver is reached at the ends. How is the variation in composition reflected by the line density at different points, and by the mass of different segments?

Q3. Consider the collection of all the possible segments that could be marked off on a given wire. Does the line density at a particular point P depend on the mass of every segment? Explain.

1.2 VOLUME AND GROWTH RATE

1.2.1 Volume and growth rate. Suppose that a quantity is growing by continuous addition to its substance—a line being drawn, a shadow spreading, water rising in a tank fed by a pipe. To be definite I will consider the last of these instances, although the discussion will apply to growth generally. The water will be measured by **volume** in cubic meters (m^3), and time in seconds (sec). By a **uniform** flow of water into the tank is meant a steady flow, an unchanging motion of liquid.

---

[3] To keep the numbers simple, here and subsequently we must admit wires which if made of real metal would be very thick. Even platinum (round) wire with this line density would be half an inch in diameter.

In each interval of time a certain volume of water is added to the tank. If the flow into the tank is uniform, the volume added in a time interval is proportional to the length of the interval. In general this is not the case, since a stronger flow will add more water in a given time than a weaker one.

At each moment of time the water in the tank has a certain **rate of growth**, or **growth rate**, which we measure in cubic meters per second ($\mathrm{m}^3/\mathrm{sec}$). If the flow is uniform, the rate of growth is the same at all times, and equals the number of cubic meters that enters the tank each second. For an arbitrary flow, the growth rate at a moment t is understood to be the growth rate ($\mathrm{m}^3/\mathrm{sec}$) produced by a uniform flow which matches the given flow at the moment t—that is, the growth rate the volume of water would have if the flow were always the same as at t. (Matching flows can be defined directly in terms of matching velocities of their particles, for which an image is suggested in §1.3.1.)

Evidently volume and growth rate form a pair of quantities analogous to mass and line density. Volume is the total quantity, accumulated over a time interval according to the rate of growth, growth rate the specific quantity, indicating the tempo of change in volume at a particular instant. Time plays here the role that was taken in the previous situation by space—the limited one-dimensional "space" along a line, that is.

As before, we can express each quantity in terms of the other when the flow is uniform. Let U be the volume[4] that enters the tank in time interval T, and let γ be the rate of growth; then

$$U = \gamma T\,,$$

or equivalently

$$\gamma = \frac{U}{T}\,.$$

In the uniform case volume is the product of growth rate and time, growth rate the quotient of volume by time. How the two quantities of the pair determine one another when growth is not uniform is the problem of calculus.

1.2.2 QUESTIONS

Q1. Reformulate Q1 of §1.1.3 as an analogous problem about water flowing into a tank, using the same numbers. Solve in the new terms.

---

[4] The letter V is reserved for another use.

Q2. The same problem with respect to Q3 of §1.1.3.

Q3. (a) A water tank has two inlets. When only inlet 1 is open, the volume of water in the tank grows uniformly at γ_1 m^3/sec; in a time interval of length T, U_1 m^3 of water is added. When only inlet 2 is open, the corresponding numbers are γ_2 and U_2. Find the growth rate and the volume added in time T when both inlets are open. (b) What can be said if the flows are not assumed uniform?

1.3 MEASURES OF RECTILINEAR MOTION

1.3.1 Distance and velocity. Let a particle move—what is a "**particle**"? A movable point, an imaginary physical object which has location without dimension, and may be assigned mass, electric charge, or other qualities as convenient. Small things, such as gas molecules, are often regarded as particles in order to facilitate the study of their behavior. Since a particle has an exact location and no parts, its motion admits of simpler mathematical description than that required for the motion of an extended body.

Let a particle move along a straight line, in a definite direction—towards the right, say. In any interval of time it traverses a certain **distance**, or length of line; at any moment it is moving at a certain **speed**, or **velocity**, which we measure in meters per second (m/sec).[5] If the motion is **uniform**, or steady, distance covered is proportional to length of time interval and speed is constant, the number of meters covered each second. In variable motion, the velocity at an instant t means the velocity of a uniformly moving particle whose motion matches that of the given one at t.

To assist the imagination with the idea of matching motions intended here it is useful to think not of two particles but of two long parallel trains, of which one advances at a uniform pace while the other accelerates steadily from an initial state of rest so as to pass it. At a certain moment the trains are moving exactly together (to a passenger in one, the other momentarily appears motionless); at that moment their motions match, in the sense we need. The velocity of the accelerating train at that moment is the velocity of the uniformly moving train.

If the particle is regarded as tracing out or drawing the line of its motion, it becomes clear that this situation differs only in terminology from the one last considered. Distance covered is length of a segment, and speed is rate of

---

[5] "Speed" and "velocity" are synonymous here, although a distinction between them, irrelevant at present, will be made later (§2.10.5).

growth of length. Thus we already know that distance and velocity are a total-and-specific-quantity pair. Nevertheless, in view of the importance of motion it is advisable to think the argument through again with the moving particle in mind (Q1).

The relations that hold when motion is uniform between distance L, velocity v, and time T are well known:

$$L = vT\,,$$

$$v = \frac{L}{T}\,.$$

That distance and speed mutually determine one another in non-uniform motion as well is shown by the cooperative behavior of the odometer and speedometer of an automobile. The relation between such instruments is mechanical or electrical; what is the mathematics of it?

1.3.2 QUESTIONS

Q1. Express in careful language the character of distance as a total quantity, and of velocity as a specific one, without referring to growth.

Q2. (a) A particle travels at the speed of sound, 330 m/sec. How long does it take it to go 1 km (= 1000 m, about 0.6 mile)? (b) If a particle has traveled 1 km in that time, was it moving at the speed of sound?

1.3.3 Velocity and acceleration. Suppose now that our particle is not only moving towards the right, but moving more rapidly as it goes. In any interval of time it receives a certain **increase in velocity**, and at any moment its velocity is increasing at a certain rate, which we call the **acceleration** of the particle. Although this new pair of quantities is, formally speaking, just another instance of growth, as a rather difficult one it deserves careful examination.

In a given time interval a certain quantity of velocity, if I may use the phrase, is acquired by the particle, just as the water tank acquired a certain quantity of volume—quantity of water, that is, which was regarded then as a thing with volume, much as the particle's motion is being regarded now as a thing with velocity. Velocity added, or **increment** of velocity, is measured in meters per second: in any time interval a certain number of meters per second is added to the particle's motion. If this augmentation takes place uniformly over time, we have so-called **uniformly accelerated** motion, in which velocity added during a time interval is proportional to the length of the interval. In particular, the same amount of velocity is added during each second. The acceleration, or rate of increase of velocity, is then a certain number of *meters per second*, *per second*, or m/sec^2 as the unit is written.

Let the velocity acquired by a uniformly accelerating particle in time T be V; if a is the acceleration, we have

$$V = aT\,,$$

$$a = \frac{V}{T}\,.$$

The meaning of acceleration at an instant for non-uniformly accelerated motion can be explained in the way by now familiar, through an instantaneous match between the given motion and a uniformly accelerated one (Q2).

1.3.4 Velocity as specific and total quantity.

What may be surprising here is that velocity, which in the first discussion of the particle's motion (§1.3.1) was a "specific" quantity, now turns up as a "total" one. Instead of velocity v at an instant we have velocity V added over a time interval—which is why I changed the notation, as I have been using capital letters for quantities belonging to intervals of time or space, small letters for quantities belonging to points and moments. Quantities are not, in fact, intrinsically total or specific, but are one or the other in relation to other quantities. Our objects of study at present are *pairs* of quantities, in which a single quantity can in principle play either of the two roles.

> At the same time there may be mathematical limitations on this possibility (cf. the note at the end of §1.1.1), and in physics it often happens that only one pair involving a given quantity arises naturally. Mass, for example, normally occurs as a total quantity only, as in our first example (Art. 1.1).

In speaking of the two roles of velocity in this way I am being a little careless. Instantaneous velocity is not really the same quantity as added velocity, for the very reason that the one is associated with a moment, and the other with an interval, of time. But they appear to be two aspects of the same thing, and there is a simple expression for their exact relation: the increment of velocity over a time interval is the *difference* of the velocities at its final and initial moments,

$$V = v_2 - v_1\,.$$

For example, a particle that accelerates from 3 m/sec to 5 m/sec has gained 2 m/sec of velocity.

1.3.5 Application to a falling body. When the problem of calculus is solved we will know how to determine velocity from acceleration, and distance from velocity; and likewise in the other direction, velocity from distance and acceleration from velocity. The use of this is illustrated by its most famous application, to free fall under the influence of gravity. A brief description is as follows.

Bodies are observed to fall together, independently of their masses. Upon measurement it is found that distances and times of fall indicate a certain uniform acceleration—the connection between distance and acceleration being made by calculus. From this **acceleration of gravity** g, whose value is about 9.8 m/sec^2, the distance of fall in any given time can then be predicted. Although accurate for terrestrial purposes, this description of gravitation proves to hold good only near the earth's surface; over greater distances astronomy shows a more complex law of acceleration to be at work. Nevertheless, calculus is again able to determine it and account for the motion it prescribes.

In Chapter 5 the arguments from the acceleration laws to the motions will be worked out in detail. We next take up the general subject of motion under the influence of force, of which gravitational motion is only a noteworthy species.

1.3.6 QUESTIONS

Q1. Let a heavy particle fall from rest with the acceleration of gravity, 9.8 m/sec^2. (a) What is its velocity after 1 sec? After 2 sec, 3 sec, n sec? After $n + \frac{1}{2}$ sec? (b) Attempt to describe how the distance of fall increases with time.

Q2. Use the two trains (§1.3.1) to illustrate the idea of uniform and non-uniform accelerations matching at an instant.

1.4 FORCE AND ITS TOTALS[6]

1.4.1 Velocity produced by force over time. Let a particle move in a straight line as before, but now suppose it has mass m and is under the influence of some force which urges it on in the direction of motion. The particle may be drawn downward by its weight, or pulled along by a thread, or pushed by an expanding spring; whatever the nature of the force, I mean it

---

[6] Article 1.4 is more difficult than the rest of the chapter, and can be omitted at a first reading. It will be needed for Chapter 5. If it is skipped now, one should also ignore a few passages of Chapter 2: §§2.3.4 and 2.8.4, and allusions to force and work in §§2.10.2 and 2.10.7. There is a summary of the article at the end (§1.4.9).

to operate continuously. I further assume that this force is the only one acting on the particle so as to affect its motion. As an ideal example we might think of a small heavy block pulled along a frictionless horizontal table, in vacuum, by a cord in which tension is maintained by a winch. The downward force of gravity on the block, and the upward force exerted on it by the table, are in balance and have no effect on its motion.

As the particle moves along, it picks up speed. How much velocity does it acquire in a given time interval, say from a time t_1 to a later time t_2? That depends not only on the strength of the force, but also on m, for everyone knows by experience that a given force will impart a greater velocity in a given time to a smaller mass than to a larger one.

> This does not contradict what is also known from experience, that bodies of different masses subject to the same force of gravity fall together (§1.3.5), because the expression "same force of gravity" is misleading. Bodies of different masses have different weights, that is, feel different downward pulls; in fact gravitation nicely proportions weight to mass (§1.1.1), and that is why the bodies fall together.
>
> I am taking for granted the fundamental physical principle that force *accelerates* mass. If this seems to be denied by the experience of pushing a block along a horizontal surface at *constant* speed by application of a steady force, reflect that two forces, not one, are really at work: the surface acts on the block by friction so as to resist the applied force. Our ideal example excludes friction.

During the interval from t_1 to t_2 the force on the particle may vary with time, or be **constant**, or **uniform**, as when exerted by an unchanging cause. However it varies, the force will impart some definite velocity V. If we let it act in exactly the same way, over an equal time interval, upon a particle of any other mass m', we will find experimentally that the velocity V' acquired by that particle stands to V in the inverse ratio of the masses:

$$\frac{V'}{V} = \frac{m}{m'}\,,$$

or

$$mV = m'V'\,.$$

For example, suppose that a winch pulls unevenly on a block of mass m, so that the tension in the cord varies; if it pulls on another block of mass m' with the same irregularity over the same length of time, so that the histories of the actions on the blocks are identical, the added velocities V and V' will bear the stated relation to the masses.

1.4.2 Momentum and impulse. It is clear from the preceding result that the quantity mV may be thought of as generated, or amassed, by the action of force over time. This product, rather than the added velocity V alone, depends only on the force and the time interval over which it operates. If the velocity of the particle is v_1 at time t_1 and v_2 at time t_2, then (§1.3.4) $V = v_2 - v_1$ and consequently

$$mV = mv_2 - mv_1 \,,$$

the increase in the product mv of mass and velocity. We call this latter quantity the **momentum** of the particle, and denote it by p:

$$p = mv \,.$$

Like velocity v, momentum is associated with an instant of time. The capital letter P can be used to represent its increase over time,

$$P = mV = p_2 - p_1 \,,$$

where $p_1 = mv_1$, $p_2 = mv_2$.

Directing our attention to the force rather than the particle we can also speak of the **impulse** I of the force on the particle over the time interval. Then impulse, which refers to force and time, equals increment of momentum, defined by mass and velocity of a particle: $I = P$.

Momentum is measured in kilogram meters per second: one kg m/sec is the momentum possessed by a mass of 1 kg moving with a velocity of 1 m/sec.

1.4.3 Increment of momentum, and force. The identification of momentum as a thing increased by force over time raises the question whether **increment of momentum** and **force** form another total-and-specific-quantity pair. Increment of momentum is associated with a time interval; let us review how force is associated with an instant. One kind of *constant* force, namely weight, can be measured with a balance, and other constant forces by equating them to weights—e.g., the force of a compressed spring equals the weight that compresses it. The magnitude of a variable force at an instant can then be understood in our usual way in terms of a matching constant force. Thus the irregular operation of the winch pulling on the block (§1.4.1) produces a definite force at each moment.

If now we restrict ourselves to constant forces, we find by experiment that:

(1) for any given force, momentum P added during a time interval is proportional to the length T of the interval,[7]

$$P \propto T\,;$$

(2) in any given length of time, momentum P added is proportional to force f,

$$P \propto f\,.$$

From (1) and (2) it follows that in general the increment of momentum is proportional to the product of (constant) force and time,

$$P \propto fT$$

(Q5). Equivalently, $f \propto P/T$, a relation which makes it possible to measure force as *momentum added per time*. Accordingly we define the **newton**, our unit of force, to be one kilogram-meter-per-second (of momentum) per second (of time):

$$1 \text{ newton} = 1 \text{ kg m/sec}^2\,.$$

Otherwise said,

> a force of one newton is one which, acting steadily for one second, imparts one kg m/sec of momentum.

The impulse of such an action can then be called one newton second—one newton acting for one second. Thus 1 newton sec of impulse produces 1 kg m/sec of momentum.

With the newton as unit our proportionalities become equalities, so that when force is constant we have

$$P = fT\,,$$

---

[7] That is, P/T is constant (for $T \neq 0$). The symbol $\propto$ means "is proportional to."

or

$$f = \frac{P}{T}\,.$$

The first equation shows that the impulse of a constant force on a particle, by which momentum P is gained, is the *product* of force and time, $I = fT$; the second equation exhibits force as momentum gained per time. These relations strongly suggest, even if they do not demonstrate, that we do have a new pair of quantities analogous to the pairs already considered. Although variable momentum and force are both too difficult of intuitive access for us to be quite confident that this is so, the fact is asserted as a fundamental postulate of physics, part of Newton's second law of motion. As momentum is accumulated by the instantaneous action of force over time, so force specifies the rate of momentum-acquisition at an instant; and the mathematics, as yet unexplained here, that connects the quantities is the same as that needed to relate the mass of a wire to its line density.

1.4.4 Force the product of mass and acceleration. On the one hand, we have now identified force as the time-specific quantity corresponding to the increment of momentum,

$$mV \leftrightarrow f\,.$$

On the other, in §1.3.3 we recognized acceleration as similarly corresponding to the increment of velocity,

$$V \leftrightarrow a\,.$$

From the latter relation it follows that mass times acceleration corresponds to the increment of momentum,

$$mV \leftrightarrow ma\,;$$

for multiplication by the constant m cannot affect the relation of total to specific quantity (Q6). Hence

$$f = ma\,,$$

the better-known form of Newton's law.

1.4.5 QUESTIONS

Q1. Express in proper units the momentum of a 5 kg particle traveling at 10 m/sec.

Q2. Identical time-varying forces act during the same time interval on particles of mass 2 kg and 3 kg. The first is accelerated from rest to 6 m/sec. If the second was initially moving at 4 m/sec, to what speed is it accelerated?

Q3. In the same ten-second time interval force f_1 imparts momentum of 20 kg m/sec to one particle, while force f_2 imparts momentum of 30 kg m/sec to another. What can be said about f_1 and f_2 (a) if they are constant forces? (b) if they are not constant?

Q4. Since weight is a force, it is correctly measured in newtons, not kilograms. How many newtons does a kilogram weigh at the surface of the earth?

Q5 (§1.4.3). In order to establish the proportionality of P and fT, prove the following lemma.

> Lemma. *If Q depends on x and y, which vary independently of one another, in such a way that*
>
> (i) *Q is proportional to x, when y is fixed, and*
> (ii) *Q is proportional to y, when x is fixed,*
>
> *then Q is proportional to xy.*

Q6 (§1.4.4). Explain.

Q7. Prove $f = ma$ when force and acceleration are uniform by considering the velocity and momentum acquired in a time T.

1.4.6 Squared velocity produced by force over distance.

In the preceding discussion we took the force on a particle to be associated with moments of time, and identified momentum as the corresponding quantity accumulated over a time interval. Let us now change our point of view, and regard the force as associated with points of space instead—that is, points of the line on which the particle travels. This may at first seem to be no change at all, since as the particle moves it occupies moments of time and points of space simultaneously; can it matter which the force is associated with in our thought?

The answer is yes, because intervals of time and intervals of space do not maintain a constant relation to one another. When velocity is uniform, time elapsed and space traversed are proportional; but owing to the presence of the force, the velocity of our particle is continuously increasing. If the particle moves a distance L in a time interval of length T, in the next time interval of the same length T it will move not L, but a greater distance L'. This means that momentum cannot be the total quantity accumulated over *distance*. For consider the uniform case, in which the force f is constant. Under its impulsion the particle acquires in each of the two equal time intervals the same momentum, equal to fT. But whatever the total quantity accumulated over distance may be (assuming that there is one), its value for the first interval is fL, and for the second fL', since in the uniform case a total quantity is always a product. These values being different from one another, the total quantity cannot be the same as momentum, or even determined by it.

We must therefore begin afresh in order to determine the total quantity that corresponds to force regarded as specific to points of the line.

Following §1.4.1, we allow a given force, possibly varying with location, to operate over an interval of the line from x_1 to x_2 on particles of mass m and m', and look for an inverse relation between the masses and the changes in velocity. The result is similar to the earlier one, but not the same. Instead of the added velocities it is now the added *squares* of velocity that are inversely as the masses. For the particle of mass m let q denote the square of instantaneous velocity, $q = v^2$, and let Q be the increase in the square of velocity as the particle travels from x_1 to x_2: $Q = q_2 - q_1$, where q_1 is measured at x_1 and q_2 at x_2. Similarly, for the other particle let $Q' = q'_2 - q'_1$. Then experiment shows that

$$\frac{Q'}{Q} = \frac{m}{m'}\,,$$

or

$$mQ = m'Q'\,.$$

One should not confuse the increase in the square of velocity with the square of the increase in velocity: in general, $Q \neq V^2$. For example, suppose that a force causes a particle of mass m to accelerate from $v_1 = 3$ m/sec to $v_2 = 5$ m/sec. The increase in velocity is $V = 2$, whose square is 4. On the other hand $q_1 = 3^2 = 9$ and $q_2 = 5^2 = 25$, so the increase in the square of velocity is $Q = q_2 - q_1 = 16$.

If another particle of mass m' is accelerated by the same force, over the same interval of the line, from 4 m/sec to 6 m/sec, we will have

$$Q' = q'_2 - q'_1 = 6^2 - 4^2 = 20\,,$$

and our assertion is that $m/m' = 20/16$.

1.4.7 Kinetic energy and work. Continuing as in §1.4.2, we think of mQ as a quantity accumulated by the particle by the action of force over distance. It depends only on the force and the interval of the line traversed by the particle. The same is true, of course, of any constant multiple of mQ; and there happens to be advantage in working with $\frac{1}{2}mQ$ rather than mQ itself. We have

$$\frac{1}{2}\,mQ = \frac{1}{2}\,mq_2 - \frac{1}{2}\,mq_1 = \frac{1}{2}\,mv_2^2 - \frac{1}{2}\,mv_1^2\,;$$

this is the increase in the quantity

$$k = \frac{1}{2}\,mv^2\,,$$

which is called the **kinetic energy**, or energy of motion, of the particle. Its increment can be denoted by K:

$$K = \frac{1}{2}\,mQ = k_2 - k_1\,.$$

As we did when defining impulse, we can direct our attention to the force rather than the particle. We then speak of the **work** W done by the force on the particle over the interval of the line. Work, which refers to force and distance, equals increment of kinetic energy, defined by mass and velocity: $W = K$. (This is a restricted meaning of the word "work," by the standard of ordinary English usage (Q4).)

The unit of kinetic energy is the kg m^2/sec^2, twice the kinetic energy possessed by a mass of 1 kg moving with a velocity of 1 m/sec. It is also called the **joule** (rhymes with *cool*).

1.4.8 Increment of kinetic energy, and force. When force is constant we can show that the increment of kinetic energy is proportional to force times distance,

$$K \propto fL\,,$$

in the same way as we earlier showed the increment of momentum proportional to force times time (§1.4.3). Then also $f \propto K/L$, so force can be measured by *kinetic energy added per distance.* If the newton is retained as the unit of force, the constant of proportionality turns out to be unity:

$$K = fL$$

and

$$f = \frac{K}{L} \,,$$

where K is measured in joules.

The same result in terms of force alone is that the work done by a constant force on a particle (by which kinetic energy K is gained) is the product of force and distance, $W = fL$. A force of 1 newton, working for 1 m, is said to do 1 newton m of work; the effect on the particle is to increase its kinetic energy by 1 joule.

We now have reason to believe that **increment of kinetic energy** and **force** form yet another of our pairs. In Chapter 4 (§4.2.8) it will be shown that this statement is in fact equivalent to the corresponding one about momentum and force. Force at a point specifies the intensity with which the particle is gathering kinetic energy; kinetic energy is accumulated by the action of force on the particle at each point.

1.4.9 Summary.

The conclusions of this article can be summarized as follows.

Let a continuously acting force cause a particle of mass m to accelerate along a straight line. The force can be regarded as a specific quantity in two ways: with respect to time, and with respect to place.

1. During an interval of time the force gives a certain **impulse** I to the particle equal to its **increment of momentum**

$$P = mV = mv_2 - mv_1 \,,$$

momentum being the instantaneous quantity $p = mv$. When the force f is constant, its impulse over time T is fT; hence $P = fT$, or $f = P/T$. In general, P is the total quantity corresponding to the time-specific quantity f—this is Newton's second law (as applied to the particle), also expressible as $f = ma$.

The unit of momentum is the kg m/sec, the momentum of 1 kg moving at 1 m/sec. The unit of force is the newton, which produces 1 kg m/sec of momentum in 1 sec. One newton acting for 1 sec has 1 newton sec of impulse, which equals 1 kg m/sec of momentum.

2. Along an interval of the line the force does a certain amount of **work** W on the particle equal to its **increment of kinetic energy**

$$K = \frac{1}{2}\,mQ = \frac{1}{2}\,mv_2^2 - \frac{1}{2}\,mv_1^2\,,$$

kinetic energy being the instantaneous quantity $k = \frac{1}{2}mv^2$. When the force f is constant, the work it does over distance L is fL; hence $K = fL$, or $f = K/L$. In general, K is the total quantity corresponding to the location-specific quantity f—a fact equivalent to $f = ma$.

The unit of kinetic energy is the joule, or kg $\mathrm{m}^2/\mathrm{sec}^2$, twice the kinetic energy of 1 kg moving at 1 m/sec. One newton acting over 1 m does 1 newton m of work, which equals 1 joule of kinetic energy.

1.4.10 QUESTIONS

Q1. A constant force of 1 newton accelerates a particle of mass 1 kg from rest. When the particle has traveled 1 m, how fast is it moving?

Q2. (a) Near the earth's surface, how far must a body of mass m fall (from rest) to reach a speed of 5 m/sec? (b) Explain why the answer is independent of m.

Q3. Calculate the increase in kinetic energy of a 1400 kg car when it accelerates from 0 to 10 mph; from 55 to 65 mph (1 mph = 0.447 m/sec).

Q4. The "work" done by a force, as we have employed the word, involves motion and does not depend on time. Give examples of ordinary English usage in which "work" seems not to involve motion and does depend on time.

1.5 GEOMETRICAL QUANTITIES ASSOCIATED WITH CURVES

1.5.1 Curves and quantities. From physics we now turn to geometry, where the problem of calculus arises in connection with features of curves in the plane. Because of the capacity of plane curves to represent physical quantities in graphs, this study has more than a purely geometrical interest: whatever can be discovered about curves has application to the quantities whose graphs they are. We will be able to supplement the conceptual structure that has guided us so far, the scheme of total and specific quantities relative to a third quantity (space or time), by a geometrical scheme that represents it—or rather, two schemes, in one of which it is the specific quantity, in the other the total quantity, that appears as a graph.

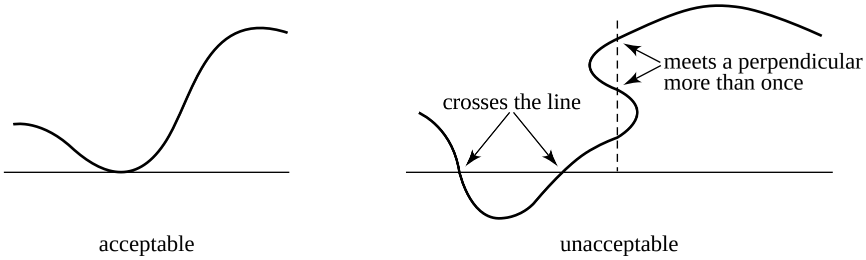

Fig. 1-1

1.5.2 Area and height. A. The uniform case. Let there be a straight line in the plane, and a continuous curve lying on one side of it; the curve may touch the line but not cross or lie along it, and must not have more than one point in common with any perpendicular to the line (Fig. 1-1). It is convenient to draw the line horizontally and call it the x-axis, even though we will not be using Cartesian coordinates as yet. The word "curve" is to be understood as comprehending straight lines and lines with corners, but (in view of the assumed continuity) not broken lines.

With each point of the line let us associate the **height** of the curve above the line at that point; that is, the length of the perpendicular from the point cut off by the curve (Fig. 1-2). With each segment of the line let us associate the **area** under the curve on that segment; that is, the area of the figure bounded by the line, the curve, and the perpendiculars from the endpoints of the segment to the curve (Fig. 1-3). I assert that area and height form a total-specific pair: area is the total quantity relative to height, height the specific quantity relative to area.

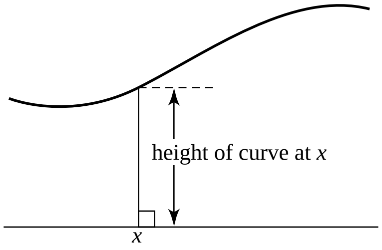

Fig. 1-2

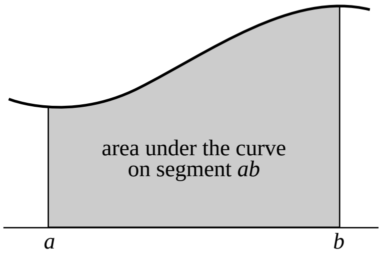

Fig. 1-3

The uniform case is that in which the curve is a straight line parallel to the x-axis (Fig. 1-4). Obviously the height is the same everywhere and the area on a segment is proportional to its length. If the height is h and the area on a segment of length L is A, we have

$$A = hL\,,$$

or

$$h = \frac{A}{L}\,,$$

the usual relations between total and specific quantity.

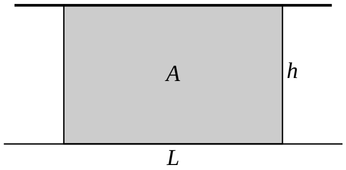

Fig. 1-4

1.5.3 B. The general case. In the general case, it is evident that height determines area, since the heights at all points of a segment together fix the upper boundary of the figure whose area is to be found, and thereby determine that figure and its area completely. Area is accumulated along a segment according to the heights of the curve above it, much as mass is accumulated along a segment of wire according to the line densities at its points.

The converse, that area determines height, is a little more difficult to see. Why is the height at a point x determined by areas under the curve? The answer is perhaps best given indirectly: if the height were other than it is, certain of the areas would be other than they are. For let h be the height at x. The continuity of the curve implies that its height at points *near* x is *near* h—it cannot *suddenly* rise or fall. If the height at x were $k \neq h$, the height near x would similarly be near k. This would produce a different area over a little segment containing x, as Fig. 1-5 illustrates. Much as the distribution of mass in the vicinity of a point of a wire determines the line density at the point, the presence of area above the vicinity of x determines the height of the curve at x.

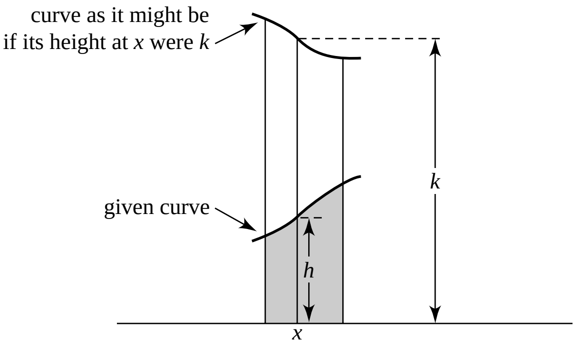

Fig. 1-5

In Fig. 1-5 the segment has been taken so short that above it the two curves (the true one at height h and the imagined one at height k) are wholly separated from one another. This must be possible, although it may require a *very* short segment.

1.5.4 Applications to figures and graphs. We can say, then, that solving the problem of calculus will enable us to express area and height in terms of one another, and determine one from the other. Of great interest to the geometer is the prospect of a general method of finding areas of curved figures. For example, the curved portion of the boundary of a parabolic figure as in Fig. 1-6 can be defined by specifying its height above all points of the straight portion (that is, giving a rule whereby h is determined by x); therefore we expect calculus to supply the means of calculating the area of the figure, a problem which, in the absence of a general method, once demanded the ingenuity of Archimedes. Areas that are not "under curves" can be calculated as sums and differences of areas that are, so we will not be confined to a special class of problems (Q2).

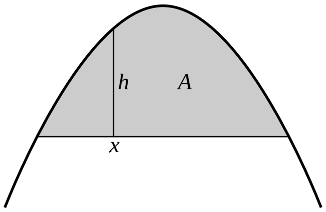

Fig. 1-6

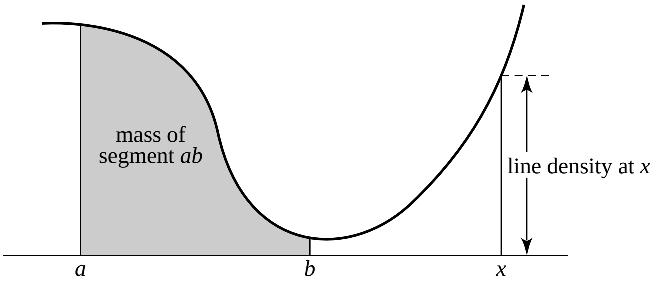

Fig. 1-7

A different application is to graphs of physical quantities. Recognition of area and height as a total-specific pair invites the simultaneous representation of two quantities in a single picture. In Fig. 1-7, the x-axis is identified with a wire of varying line density. The height of the curve at any x represents the line density there, and the area under it on any segment ab represents the mass of the segment. It is as though the mass were piled up on the line like sand. You will agree that this is a great aid to the imagination—and will hasten to produce like representations of the other pairs of quantities we have considered (Q3, Q4). When the quantities are associated with time t rather than place x, the axis may be called the t-axis, and is to be thought of as a geometrical representation of the continuum of time.

1.5.5 QUESTIONS

Q1. Find the area under the straight line $y = \mu x$, $\mu > 0$, between $x = a > 0$ and $x = b > a$ (Fig. 1-8).

Q2. (a) If areas under curves can be found, how can the area of a circle be determined? (b) A closed curve is **convex** if it bulges outward everywhere; that is, if every chord, apart from its endpoints, lies entirely inside the curve (Fig. 1-9). Show that the area enclosed by a convex curve is a difference of two areas under curves. (c) What if the closed curve is not convex?

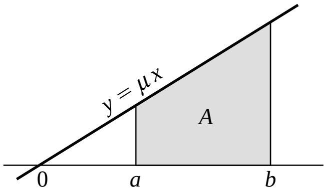

Fig. 1-8

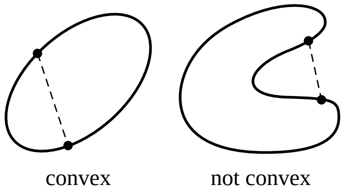

Fig. 1-9

Q3. Draw figures like Fig. 1-7 and label them so that they apply to the other pairs of physical quantities we have identified. Include the appropriate physical units.

Q4. The velocity of a certain particle is always proportional to the time it has been traveling since $t = 0$. How far does it go between $t = a$ and $t = b$?

1.5.6 Rise and slope. A. The uniform case.

The second geometrical scheme also begins with the x-axis and a curve which has at most one point in common with any perpendicular to the axis. However, instead of requiring the curve to lie above the axis, we will now assume that it always rises as we move along the axis towards the right (Fig. 1-10).

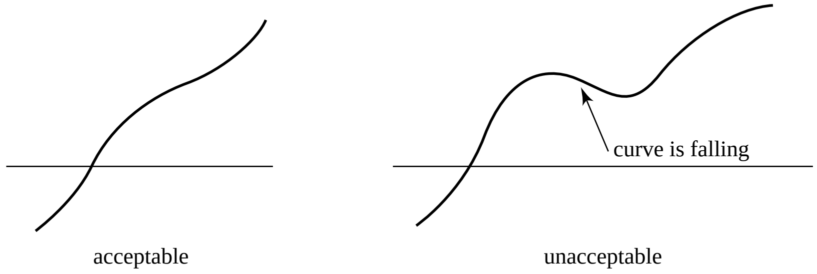

Fig. 1-10

With each segment of the x-axis we associate the **rise** of the curve that takes place over that segment: the distance by which the curve rises between the perpendiculars to the axis at the left and right endpoints of the segment (Fig. 1-11). With each point x of the axis we associate the **slope** of the curve at the point P on it determined by the perpendicular at x; that is, the slope of the tangent line at P (Fig. 1-12). I assert that rise and slope form another

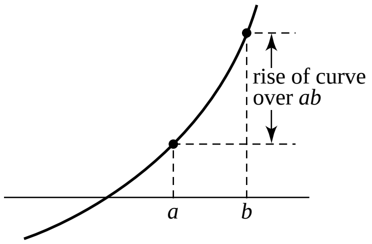

Fig. 1-11

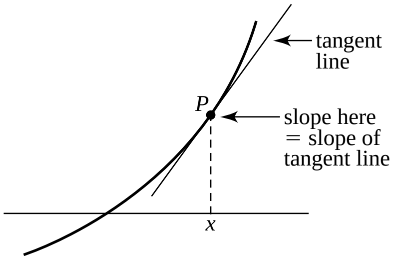

Fig. 1-12

pair: that rise is the total quantity relative to slope, slope the specific quantity relative to rise.

> The slope is represented in Fig. 1-12 only in so far as the visible inclination of the tangent line is able to indicate it. We are assuming that the curve has a definite tangent line and slope at each point P; hence we exclude corners, where there is no well-defined tangent, as well as vertical tangents, which have no slope.

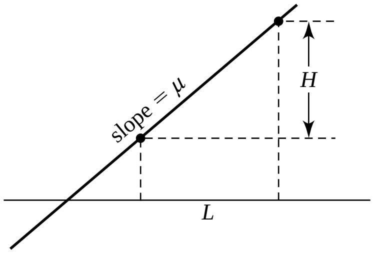

Fig. 1-13

The uniform case is again that in which the curve is a straight line, but this time the line rises (Fig. 1-13). The slope of the line is everywhere the same, and the rise over a segment is proportional to its length. If the slope is[8] μ and the rise over a segment of length L is H, we have

$$H = \mu L\,,$$

or

$$\mu = \frac{H}{L}\,,$$

as expected.

1.5.7 B. The general case. The definition of slope in the general case, by means of a tangent line, is in keeping with our usual method of describing a non-uniform thing at a point by matching it to a uniform one: the tangent line at P fits the curve better than any other straight line through P. For

---

[8] Avoiding the letter m, already used for mass.

the tangent line, as for any straight line, slope signifies rise per **run**, or distance measured along the x-axis; hence we interpret the slope of a curve at P in the same way, as rise per run "at P"—the slope of the uniform curve, or straight line, that matches the given curve at P. It should be noted that we can speak of the slope "at x" rather than "at P," regarding the point of the curve as specified by the corresponding point on the axis.

This definition of slope shows quite clearly that its relation to rise is the same as that of line density (mass per length) to mass, or velocity (distance per time) to distance. In order to exhibit rise and slope as determining one another, we can proceed as follows (for simplicity confining our description to points where the curve lies above the axis). Suppose first that the rise is known for every segment. The slope at an arbitrary point x will surely be determined if the course of the curve is known in the vicinity of x. Let a be any point to the left of x. Since the rise is known over every interval whose left endpoint is a, the height of the curve is known at every point to the right of a (Fig. 1-14). This means that the course of the curve is determined, and with it the slope at x.

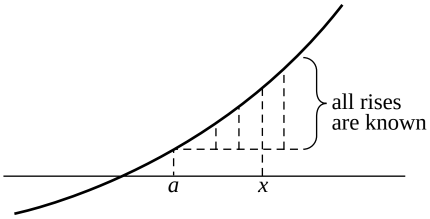

Fig. 1-14

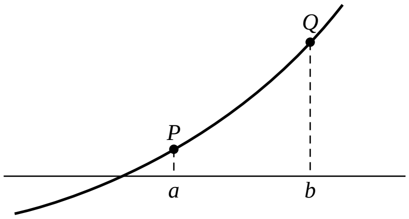

Fig. 1-15

Conversely, suppose that the slope is known at every point. Imagine flying in an airplane above the x-axis along the curve from point P, above a, up to point Q, above b (Fig. 1-15). At each point of the flight the plane is aiming upward at a certain angle to the horizontal. This angle of climb, known at all points, determines the course of the flight from P, since to deviate upward or downward from arc PQ the plane would have to turn so as to ascend more or less steeply. Consequently it determines the total increase in altitude from P to Q, which is the rise over ab. Now slope is not the same as angle, but it determines angle, and hence rise; moreover, as rise *per horizontal distance* it determines rise exactly as a quantity gained relative to distance along the x-axis: at each point the slope directs the airplane as to the rate at which it is to gain altitude relative to its horizontal progress. Thus slope not only determines rise, but determines rise as its corresponding total quantity.

1.5.8 Rise and height. You will have noticed that the total quantity "rise," denoted H, is closely related to the specific quantity "height," denoted h, which was introduced in §1.5.2. Their relation is analogous to that between velocity V added over a time interval and velocity v at an instant (§1.3.4): for a rising curve above the axis, the rise over a line segment is the difference between the heights at its endpoints, $H = h_2 - h_1$ (Fig. 1-16). If we allow ourselves to regard H and h as aspects of the same quantity, then we find this quantity appearing in one geometrical pair (together with area) as a specific quantity, and in another (together with slope) as a total quantity, in the same way as velocity occurred in two physical pairs.

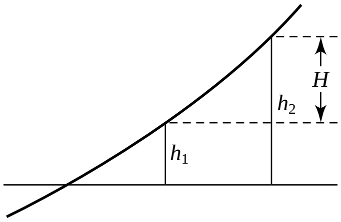

Fig. 1-16

1.5.9 Applications to curves and graphs. With the solution of the problem of calculus will be solved the geometrical problem of finding the tangent line to a curve at a given point, the curve being defined in relation to some straight line. The parabola shown in Fig. 1-17, for example, can be described by specifying its height h above an axis at each point; from this its

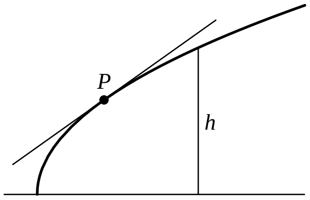

Fig. 1-17

rise over every interval is known, and thereby the slope of the tangent line through P, and the line itself. Like the problem of finding areas of figures, the tangent problem is difficult or impossible when the general method of calculus is unknown.

Rise and slope serve also to represent pairs of physical quantities in a new way. Like Fig. 1-7, Fig. 1-18 has the x-axis identified with a wire. This time, however, the mass of segment ab is represented by the rise between a and b, and the line density at x by the slope at that point. The physical situation portrayed does not differ from the one pictured in Fig. 1-7; what is different is the mode of portrayal.[9] Which mode to prefer in a given case will depend on which curve can be drawn from available data, among other considerations.

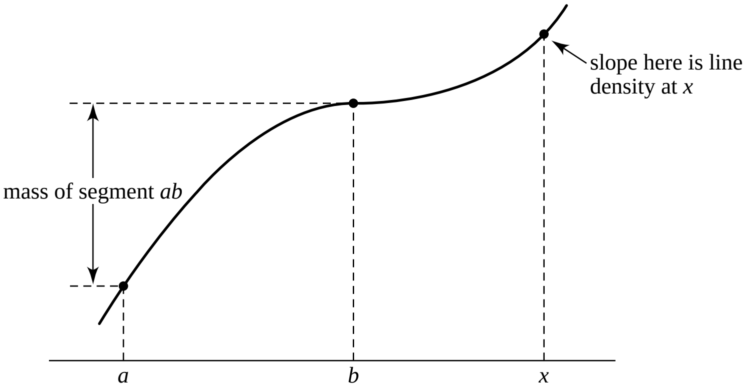

Fig. 1-18

1.5.10 QUESTIONS

Q1. What correlation is there between the shapes of the curves in Figs. 1-7 and 1-18?

Q2. Draw figures like Fig. 1-18 for each of the other pairs of physical quantities. Include the appropriate physical units.

Q3. Suppose that the curve in Fig. 1-18 is shifted vertically upward or downward—that is, its height is everywhere increased or decreased by the same quantity. What effect does this have on its representation of mass and line density by rise and slope?

---

[9] The curve in Fig. 1-18 is so drawn that its rise and slope correspond to the area and height in Fig. 1-7, except that the vertical scale is reduced.

Q4. In Fig. 1-19, PQ is a quarter-circle, PR a horizontal tangent. As x moves from P to R, how do the following quantities change? (i) α, the angle of inclination to the horizontal (the sloping line is a tangent); (ii) the slope at x.

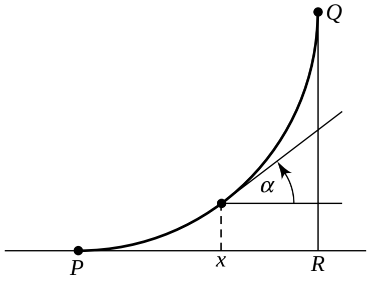

Fig. 1-19

1.6 NUMBERS, VARIABLES, AND FUNCTIONS

1.6.1 Variables. Independent variables. Each of the pairs of quantities we have looked at has been founded upon a linear continuum—a line in physical space, the continuum of time, or a geometrical line—with whose intervals and points the total and specific quantities have respectively been associated. As you know, it is customary to represent such a continuum by a coordinate axis, usually an x-axis for a line in physical or geometrical space and a t-axis for time. The **variable** x or t assumes as its values the numbers that are assigned to the points of the axis. Coordinate representation requires certain choices to be made; thus in an example involving time, where the values of t are numbers that name moments or points of time, we have to choose a zero-point, a unit, and a positive direction. The positive direction is always taken to be towards the future, and we will always employ the second as unit; commonly $t = 0$ is chosen as the moment when motion begins.

Particular values of a variable that are not specified as numbers can be denoted by other letters or with the aid of subscripts: "the value a of x," or "$x = a$"; "the value x_1 of x"; "time t_2." The plain letters x and t ought in strictness to be reserved for denoting the variables, but I will constantly violate this rule and say, e.g., "the value x," in order to avoid multiplication of symbols. Such deliberate carelessness is often expedient in calculus, provided one remembers that there is no imprecision in the mathematics, only in writing and speech.

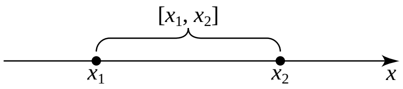

Fig. 1-20

Variables such as x and t take their numerical values at points. As for intervals, these are designated by reference to their endpoints: given $x_1 < x_2$, the interval (or segment) from x_1 to x_2, which includes all x-values such that $x_1 \le x \le x_2$, is denoted $[x_1, x_2]$ (Fig. 1-20). (Read "x_1, x_2"; the simpler notation x_1x_2 would cause confusion with a product.) So, for example, we might say, referring to some given wire:[10] the line density at x_1 is 3 kg/m; the length of $[x_1, x_2]$ is 5 m, and its mass is 6 kg. Points of an interval that are not endpoints are called **interior** points. The word "interval" can also apply to a region such as the non-negative x-axis (all x such that $x \ge 0$), for which we need not introduce a notation.

If the values assumed by x or t can be arbitrarily specified, as when we choose a point x at which to evaluate line density, the variables are called **independent**, a designation which serves to distinguish them from the variables to be considered next.

1.6.2 Dependent variables. Functions.

Quantities that depend on x or t are, or can be made into, numerical variables as well, each denoted by its own letter. They are called **dependent** variables, because we think of their values as determined by values of x or t. Dependent variables which are *total* quantities have values determined by *intervals* of x or t, ones which are *specific* quantities have values determined by *single values* of x or t. Table 1-1 lists the variables we have encountered so far.

Each of the specific quantities is a **function** of its independent variable, which means that to each value of the independent variable is assigned a value of the dependent one. If $x = 3$, λ has a certain value, say 2; if $x = 4$, it has again a certain value, possibly the same one; in general, each value of x determines a value of λ. We write $\lambda(3)$ (read "λ of 3") for the value of λ associated with 3; then $\lambda(3) = 2$. Similarly $\lambda(4)$ is the line density at $x = 4$, and in general $\lambda(x)$ ("λ of x") denotes the value of λ at any value x. We may also write $\lambda(x)$ instead of λ just to remind ourselves that λ is a function

---

[10] Recall the footnote to §1.1.3.Q1(a).

Table 1-1

Where symbols for dependent variables were introduced	Independent variable	Dependent variables Total quantity	 Specific quantity
§1.1.2	x (location on wire)	M (mass; kg)	λ (line density; kg/m)
§1.2.1	t (time)	U (volume; m^3)	γ (growth rate; m^3/sec)
§1.3.1	t (time)	L (distance; m)	v (velocity; m/sec)
§1.3.3	t (time)	V (velocity added, or increment of v; m/sec)	a (acceleration; m/sec^2)
§§1.4.2–3	t (time)	P (momentum added, or increment of momentum p; kg m/sec) I (impulse; newton sec)	f (force; newton)
§§1.4.7–8	x (location on line of motion)	K (kinetic energy added, or increment of kinetic energy k; joule) W (work; newton m)	f (force; newton)
§1.5.2	x (abscissa)	A (area)	h (height)
§1.5.6	x (abscissa)	H (rise, or increment of h)	μ (slope)

of the variable x; then we refer to "the function $\lambda(x)$," and similarly "the function $v(t)$."

The relation between a total quantity and its independent variable is a little more complicated. A value of the total quantity is associated not with each value, but with each interval of values of the independent variable. The quantity M, for instance, is a function not of x, but of the intervals of values of x.

To the interval $[3, 5]$, M assigns a certain value, say 4—the mass, in kilograms, of the 2 m segment of wire from $x = 3$ to $x = 5$; to the interval $[5.7, 6.1]$ it likewise assigns a certain value, possibly the same one; so in general, to each interval of x-values corresponds a value of M. We write $M([3, 5])$ ("M of 3, 5") for the value of M associated with $[3, 5]$, so that $M([3, 5]) = 4$. In general, $M([x_1, x_2])$ is the value of M on any interval $[x_1, x_2]$.[11]

1.6.3 Cumulative quantities and functions. The notation last introduced is clumsy, and the truth is that functions of intervals are themselves not as handy as functions of points (that is, of single numbers). Fortunately it is possible to express functions of intervals in terms of functions of points; I will now show how this is done.

Consider first the example of the wire. Let its left end be at $x = 0$, and for any x let $m(x)$ be the mass of the segment of wire from 0 to x; that is,

$$m(x) = M([0, x])$$

(Fig. 1-21). So defined, m is a function of x: to each x it assigns a number $m(x)$. We can call m a **cumulative** quantity or function, since $m(x)$ is the mass accumulated in passing along the wire from 0 up to x. The mass of any segment $[x_1, x_2]$ is evidently given by

$$M([x_1, x_2]) = m(x_2) - m(x_1)$$

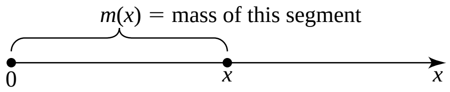

Fig. 1-21

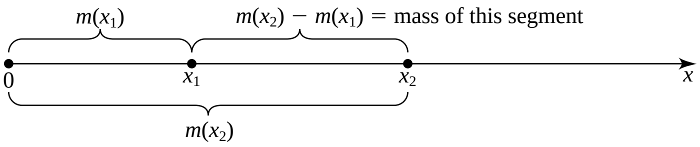

Fig. 1-22

---

[11] Here letters with subscripts are being used to denote arbitrary values of the variable.

(Fig. 1-22). Thus M is fully expressed by means of m: the value of M on any interval is the difference of the values of m on the endpoints of the interval.

In the uniform case, since the length of $[0, x]$ is x, its mass is λx, where λ is the constant line density; that is, the cumulative function has the simple form

$$m(x) = \lambda x\,,$$

a so-called **linear** function of x.

> We are at liberty to choose a different point from 0 as the starting-point for accumulation; the resulting cumulative function will differ by a constant from m as defined above (Q3). If we do this, or more generally if accumulation does not begin at the left end of a wire, we have to deal with points to the left, as well as to the right, of the starting-point. This complication will be addressed in Chapter 2 (§2.10.4).

Other total quantities can be handled similarly. In the case of growth, we obtain a cumulative function by defining $u(t)$ to be the volume added to the water tank from the beginning of filling at $t = 0$ until time t, $u(t) = U([0, t])$; then $u(t)$ is the volume at time t, and the volume added in any interval $[t_1, t_2]$ is $u(t_2) - u(t_1)$. In the case of added velocity V, the cumulative function $v(t)$ is the velocity acquired since the beginning of motion at $t = 0$, or simply the velocity at time t, and V added between t_1 and t_2 is $v(t_2) - v(t_1)$, as we noticed before (§1.3.4). Similarly, for a curve which rises from the axis at $x = 0$ its height $h(x)$ serves as cumulative function with respect to rise H (§1.5.8 and Fig. 1-23).

In view of the close connection between total and cumulative functions, it is apparent that our examination of pairs of corresponding total and specific quantities might have been carried out with regard to pairs of cumulative and specific quantities instead. The latter type of pair will play the chief part in the work to come, as most of it will have to do with functions of numbers.

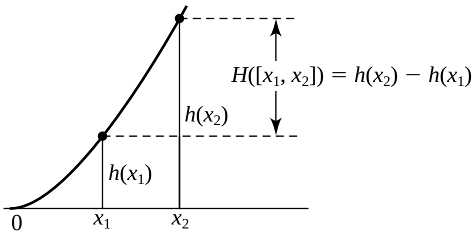

Fig. 1-23

1.6.4 QUESTIONS

Q1. Referring to Table 1-1 and our examples, express in words the meaning of the following symbols: $\gamma(3.5), L([5,6]), a(1), f(t_2), K([a,b]), W([a,b]), A([0,1]), \mu(\pi)$.

Q2. (a) Find the cumulative function $m(x)$ for a uniform wire of line density 3 kg/m with left end at $x=0$. (b) For the same wire, find $m_1(x)$, the cumulative mass measured towards the right from $x=1$.

Q3. Let x_0 be a point on an arbitrary wire with left end at $x=0$, and let $m_0(x)$ be the cumulative mass measured towards the right from x_0. Express $m_0(x)$ in terms of the function m.

Q4. In the pair area and height, what is the geometrical interpretation of the cumulative function?

Q5. Is momentum $p = mv$ a cumulative function?

1.6.5 The meaning of "function." When we say of two variables x and y that y is a function of x, we are making mention not of two entities, but three. There are

(i) the independent variable x,
(ii) the dependent variable y, and
(iii) the relation between them, or law of dependence of y on x.

The third of these is what the word "function" primarily refers to.

In the most general meaning of the term, a function is any law by which, to each thing in one class of things, a thing in another class of things (possibly the same class) is assigned. The function is said to be **defined** for the members of the first class and to take **values** in the second class. Each body has a definite mass, in the sense of a number of kilograms; hence the assignment of mass to body specifies a function defined for all bodies and taking values in the class of numbers (of kilograms). That two bodies may have the same mass is immaterial; what matters is that the law specifies one definite mass for each given body. Of interest to us in the realm of purely numerical variables are functions that assign numbers to numbers, or numbers to intervals of numbers. The former of these types is the most common kind of function, and is always intended when the term is used without qualification.

One can think of a function as a black box with a hopper for input and a chute for output (Fig. 1-24). If the function is the one that assigns mass to a body, we imagine putting in any body—a piece of wire, say—and getting out its mass in kilograms, or more precisely a statement of the mass—"0.5 kg," perhaps printed on a slip of paper. An ordinary metric weighing scale is an (imperfect) embodiment of this function.

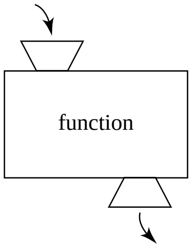

Fig. 1-24

An example of a numerical function is "square," which can be written $(\ )^2$, the blank indicating where the number to be squared, or input, will go (Fig. 1-25). If "y is a function of x" by the law of squaring, the variable x denotes an input *in general*, the variable y denotes an output *in general*, and we write

$$y = x^2$$

to indicate the same thing as is meant by Fig. 1-26. Here the box is labeled with the symbol for the function, and the hopper and chute with the general symbols for the input and the output. Put in any particular value x_0 of x and you get out a certain value y_0 of y, namely the square of x_0. A simpler but less exact picture (recall the remark on symbols, §1.6.1) is Fig. 1-27.

Since I do not mean to suggest by the image of the box that an input vanishes into it and is used up, or that an output once produced is gone from it and no longer available, perhaps it would be better to replace the hopper by a "reader" or "scanner" which examines inputs, and the chute by a "writer" or "producer" which supplies outputs. One might then compare a function to an ordinary bar-code device, which reads numbers written in one form, as bar codes, and

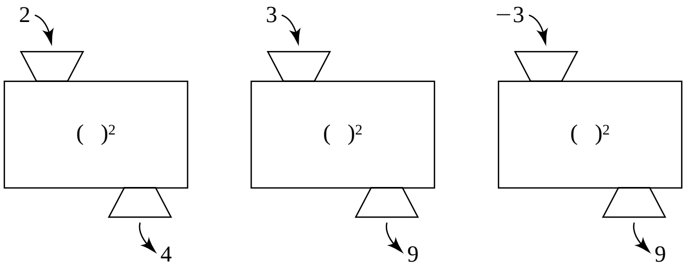

Fig. 1-25

CALCULUS - The Elements
© World Scientific Publishing Co. Pte. Ltd.
http://www.worldscibooks.com/mathematics/4920.html

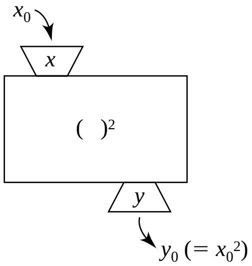

Fig. 1-26

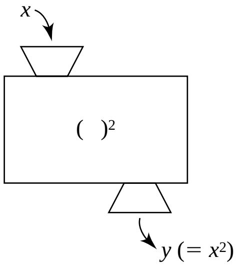

Fig. 1-27

produces corresponding numbers written in another form, as printed prices on a sales slip. Whichever way it is conceived, a function is nothing but a certain correlation of inputs and outputs. The purpose of the images is only to assist the mind in recognizing that such a correlation is an entity in its own right.

1.6.6 Domain and range. The set of all the allowable inputs to a function, the class of things for which it is defined, is known as its **domain**. The set of all its possible outputs, the collection of values it takes, is called its **range**.

Specifying the domain is part of defining a function. Sometimes the domain is limited by the functional law of dependence, as in the case of $y = 1/x$, for which it cannot include 0. (We may then choose to augment the law in order to extend its domain, as in Example 2, §4.2.9.) A restriction can also be imposed. If we were interested in whole numbers only, say, we might view "square" as applying to them alone. We would then have a different function from "square" as defined for all numbers: its domain would be smaller, and consequently its range too.

The range is determined by the domain and the functional law together. In a given case it may be the law, rather than the domain, that is so chosen as to yield a desired range. An example of this is the definition of "square root" (Q3; see also §4.3.2).

The sign of independent and dependent variables is treated in Art. 2.10.

1.6.7 Notation for functions and representation by graphs.

A function is often assigned its own symbol, usually a letter, and frequently the letter f. If we have assigned f to the squaring function, then

$$y = f(x)$$

means the same thing as

$$y = x^2\,;$$

thus $4 = f(2)$, etc. In other words, f, or $f(\ )$, is the same as $(\ )^2$. If, in a given context, the function is understood, we can simply write

$$y = y(x)\,,$$

which expresses the fact that y is a function of x—dependent on x by some law—without naming the function. This notation has already been used in writing $m(x)$, $v(t)$, etc.

It is important to realize that a function as such can be thought of apart from the names of any particular variables it may relate. The function $(\ )^2$ is the same whether we write $y = x^2$ or $z = w^2$ (where both x and w can be any number)—the same squaring machine is at work in both cases.

Functions which relate numbers to numbers admit the familiar geometrical representation provided by a Cartesian graph. A graph is a kind of picture of a function, in which the whole law of dependence of one variable on another is exhibited by a pattern in the plane, ordinarily a curve (Fig. 1-28). From the input x the output y is derived by following the indicated path. Since to each x must correspond one determinate y, no vertical line can intersect the curve more than once.

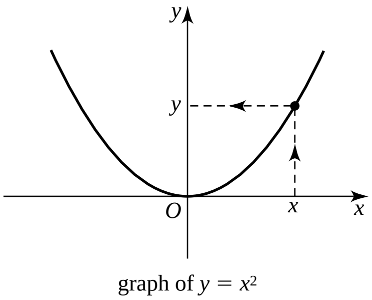

Fig. 1-28

1.6.8 Monotone functions. A function whose graph only rises (as in §1.5.6), or only falls, over some interval, is called **monotone** or **monotonic** on the interval; it is **increasing** if the graph rises, **decreasing** if it falls. In the former case an increase in x always results in an increase in $f(x)$,

$$\text{if } x_1 < x_2 \text{ then } f(x_1) < f(x_2)\,; \tag{1}$$

in the latter, an increase in x produces a decrease in $f(x)$,

$$\text{if } x_1 < x_2 \text{ then } f(x_1) > f(x_2)\,. \tag{2}$$

As Fig. 1-28 shows, the function $y = x^2$ is increasing for $x \geq 0$ (that is, (1) holds for any non-negative x_1 and x_2) and decreasing for $x \leq 0$.

A less stringent condition than (1) is:

$$\text{if } x_1 < x_2 \text{ then } f(x_1) \leq f(x_2)\,. \tag{3}$$

It is satisfied by any constant function, for example, whereas (1) is not. A function satisfying (3) can be called increasing *in the weaker sense.* Analogously we define a function decreasing in the weaker sense; both kinds are monotone in the weaker sense.

1.6.9 Functions and calculus. The laws of dependence called functions will be central to our study. In fact calculus *is* a study of functions. Once number scales have been chosen on which length, time, abscissa, mass, velocity, force, ordinate, etc., are to be measured, all of these general quantities are represented by numerical variables. *Particular* quantities—the line density, or mass distribution, or cumulative mass, of a certain wire; the force on a certain particle, or its acquisition of momentum over time intervals, or its cumulative momentum—all of them are nothing but functions, laws of the dependence of certain of the variables on others. A wire has line density λ; that is, a function $\lambda(x)$ assigns a number λ to each number x. To distinguish this function from others I can give it a name: $\lambda = g(x)$. Then the function g is the mathematical description of the line density. Again, the wire has mass distribution M; that is, a function $M([x_1, x_2])$ assigns a number M to each number-interval $[x_1, x_2]$. And the wire has cumulative mass m; that is, a function $m(x)$ assigns a number m to each number x. It is in virtue of the numerical-valued functions which describe it that the wire becomes a subject of our mathematics.

Consequently the problem of calculus can be put as follows: to determine the nature of that relation between *functions* which expresses the relation common to all pairs of total (or cumulative) and specific quantities, as we have seen it in many examples; to express one function of a related pair in terms of the other; and to find one when the other is known.

1.6.10 QUESTIONS

Q1. (a) As applied to all human beings x, is "y is the mother of x" a function? (b) The same question for "daughter of," as applied to all women.

Q2. What is the difference between "the value $y = 2$" and "the function $y = 2$"?

Q3. Is "square root" a function? Explain.

Q4. A closed curve, such as a circle, is not the graph of a function. How can such a curve be treated by calculus, if calculus studies functions?

Q5. Give an example of a function, defined for all numbers, whose graph is not a continuous curve.